

# Biological and Structure Characterisation of Eukaryotic Prefoldin

A thesis submitted for the  
Degree of Doctor of Philosophy

by

Denise Phuong Tran B. Sc. (Hons.)

Department of Chemistry



THE UNIVERSITY  
*of* ADELAIDE

January 2018

# Contents

---

<b>Contents</b> .....	<b>i</b>
<b>Acknowledgments</b> .....	<b>v</b>
<b>Declaration</b> .....	<b>vi</b>
<b>Abstract</b> .....	<b>vii</b>

## **Chapter 1: Introduction**

1.1: Overview .....	1
1.2: Proteins .....	1
1.2.1: Protein Structure.....	1
1.2.2: Protein Misfolding and Aggregation .....	5
1.2.3: Molecular Chaperones .....	10
1.2.3.1: Prefoldin .....	12
1.3: Aims and Outcomes .....	15

## **Chapter 2: Biophysical Techniques to Determine Protein Structure and Interactions**

2.1: Introduction.....	17
2.2: Biophysical Techniques .....	17
2.2.1: ThT Assay .....	17
2.2.2: NMR Spectroscopy .....	19
2.2.3: CD Spectroscopy .....	27
2.2.4: Mass Spectrometry .....	30
2.2.4.1: Ionisation .....	30
2.2.4.2: Mass Analysers.....	33
2.2.4.3: Detectors .....	34
2.2.5: IM-MS.....	35
2.2.5.1: Drift Tube Ion Mobility .....	35
2.2.5.2: Travelling Wave Ion Mobility .....	36
2.2.5.3: Computational Approaches to CCS Calculations.....	37
2.2.5.4: Ion Mobility-Mass Spectrometers .....	38
2.2.5.5: Tandem MS Analysis of Peptides .....	40
2.2.6: Protein Chemical Cross-linking .....	43

## Chapter 3: Mammalian Prefoldin, Human Prefoldin Subunits 5 and 6 and Their Chaperone Activity

3.1: Introduction.....	48
3.1.1: Aims.....	49
3.2: Results and Discussion .....	51
3.2.1: Sequence Comparison of PFD Homologues .....	51
3.2.2: Isolation of <i>b</i> PFD .....	57
3.2.3: Expression and Isolation of <i>h</i> PFD Subunits .....	67
3.2.4: The Biological Activity of Eukaryotic PFD .....	71
3.2.4.1: Chaperone Properties of <i>b</i> PFD and <i>h</i> PFD Subunits 5 and 6 .....	71
3.2.4.3: TEM analysis of the incubated A $\beta$ (1-40), <i>b</i> PFD and <i>h</i> PFD Subunits 5 and 6.....	81
3.2.4.4: Cytotoxicity assays of A $\beta$ (1-40) and <i>h</i> PFD Subunits 5 and 6.....	83
3.3: Experimental .....	91
3.3.1: BLAST Needleman-Wunsch Global Alignment.....	91
3.3.2: Extraction of <i>b</i> PFD.....	91
3.3.3: <i>h</i> PFD Subunit Expression .....	93
3.3.4: Shotgun Proteomics Analysis of Protein Fractions .....	94
3.3.5: Electrophoresis and Staining .....	95
3.3.6: Protein Concentration Determination .....	97
3.3.7: A $\beta$ (1-40) Aggregation and Inhibition ThT Assays.....	97
3.3.8: Cell Cytotoxicity Assay.....	97

## Chapter 4: Far-UV Circular Dichroism Spectroscopy Analysis of the Human Prefoldin Subunits 5 and 6

4.1: Introduction.....	99
4.1.1: Aim .....	100
4.2: Results and Discussion .....	101
4.2.1: Far-UV CD Spectroscopy of <i>h</i> PFD Subunits 5 and 6 .....	101
4.2.1.1: DichroWeb Deconvolution of CD Spectroscopy of <i>h</i> PFD Subunit 5 and 6 .....	102
4.2.2: Far-UV CD Spectroscopy of the <i>h</i> PFD Subunits 5 and 6 at Varying pH .....	105
4.2.3: Thermostability of <i>h</i> PFD Subunits 5 and 6 as monitored by Far-UV CD Spectroscopy.....	110
4.2.4: The Products of A $\beta$ (1-40) Following Incubation with <i>h</i> PFD Subunits 5 and 6 Monitored by Far-UV Circular Dichroism.....	118
4.3: Experimental .....	123
4.3.1: CD Assays.....	123
4.3.2: <i>pl</i> Computational Calculation of <i>h</i> PFD Subunits.....	123
4.3.3: Dichroweb Deconvolution of CD Spectra .....	123
5.1: Introduction.....	124

## Chapter 5: Nuclear Magnetic Resonance Analysis of the Human Prefoldin Subunits 5 and 6

5.1.1: Aim .....	126
5.2: Results and Discussion .....	127
5.2.1: 3D NMR Spectroscopy.....	127
5.2.2: 2D NMR Spectroscopy.....	133
5.2.2.1: 2D NMR Spectroscopy of <i>h</i> PFD Subunit 5 .....	133
5.2.2.1: 2D NMR Spectroscopy of <i>h</i> PFD Subunit 6 .....	138
5.3: Experimental .....	143
5.3.1: <sup>15</sup> N and <sup>13</sup> C Labelled <i>h</i> PFD Subunit Expression.....	143
5.3.2: NMR Sample Preparation .....	143
5.3.3: NMR Parameters.....	143

## Chapter 6: Ion Mobility-Mass Spectrometry of Mammalian Prefoldin and the Human Prefoldin Subunits 5 and 6

6.1: Introduction.....	145
6.1.1: Aim .....	149
6.2: Results and Discussion .....	150
6.2.1: Development of CTC .....	150
6.2.1.1: Thermal Denaturation Experiment of IgG .....	151
6.2.2: IM-MS Analysis of <i>b</i> PFD.....	153
6.2.2.1: ‘Native’ IM-MS Analysis of <i>b</i> PFD .....	153
6.2.2.2: CID IM-MS Analysis of <i>b</i> PFD.....	160
6.2.2.3: Thermal Denaturation IM-MS Analysis of <i>b</i> PFD .....	162
6.2.3: IM-MS Analysis of <i>h</i> PFD Subunits 5 and 6 .....	169
6.2.3.1: ‘Native’ IM-MS Analysis of <i>h</i> PFD Subunits 5 and 6 .....	169
6.2.3.2: Thermal Denaturation IM-MS Analysis of <i>h</i> PFD Subunits 5 and 6 .....	171
6.3: Experimental .....	179
6.3.1: Sample Preparation.....	179
6.3.2: Ion-mobility Mass Spectrometry .....	179
6.3.3: Data Analysis.....	179
6.3.4: Thermal Denaturation Experiments.....	180

## Chapter 7: Chemical Cross-linking of the Human Prefoldin Subunits 5 and 6

7.1: Introduction.....	181
7.1.1: Aim .....	185
7.2: Results and Discussion .....	186
7.2.1: Synthesis of Cross-linker S55’THS .....	186



7.2.2: Cross-linking of Ac-AAKA .....	187
7.2.3: Cross-linking of Ubiquitin .....	190
7.2.4: Cross-linking of Caerin 1.10 .....	201
7.2.5: Cross-linking of <i>h</i> PFD Subunits 5 and 6 .....	207
7.2.5.1: <i>h</i> PFD Subunit 5 Cross-linked with S55'THS .....	207
7.2.5.2: <i>h</i> PFD Subunit 6 Cross-linked with S55'THS .....	211
7.2.6: Cross-linking of <i>h</i> PFD Subunits 5 and 6 to A $\beta$ (1-40) .....	215
7.2.6.1: <i>h</i> PFD Subunit 5 Cross-linked to A $\beta$ (1-40) with S55'THS .....	216
7.2.6.2: <i>h</i> PFD Subunit 6 Cross-linked to A $\beta$ (1-40) with S55'THS .....	220
7.3: Experimental .....	227
7.3.1: Synthesis of S-methyl 5,5'-thiodipentanoylhydroxysuccinimide .....	227
7.3.2: Cross-linking of Ac-AAKA and Caerin 1.10 .....	228
7.3.3: Cross-linking of Ubiquitin .....	229
7.3.4: Cross-linking of <i>h</i> PFD Subunits 5 and 6 to A $\beta$ (1-40) .....	229
7.3.5: Waters Synapt IM-MS .....	229
7.3.6: Waters Synapt LC-IM-MS .....	230
7.3.6: Agilent 6560 LC-MS/MS .....	230
7.3.7: Agilent 6560 LC-IM-MS/MS .....	231

## Chapter 8: Summary, Conclusion and Future Directions

8.1: Summary of Results .....	232
8.2: Overall Conclusion .....	234
8.3: Future Directions .....	236
References: .....	237

## Appendix

Appendix A: The 20 Common Amino Acid Residues .....	276
Appendix B: Amino Acid Sequences of Different Species of PFD .....	278
Appendix C: Ions Identified by Proteomic Analysis Correlating to Endoplasmin .....	280
Appendix D: Genes Encoding the <i>h</i> PFD subunits 5 and 6 .....	281
Appendix E: Computationally Calculated Properties of <i>b</i> PFD and <i>h</i> PFD Subunits 5 and 6 .....	282
Appendix F: Post-hoc Test Following 1 Way ANOVA Summary of Cell Viability Assays for A $\beta$ (1-40) Co-Incubated with <i>h</i> PFD Subunit 5 .....	283
Appendix G: Post-hoc Test Following 1 Way ANOVA Summary of Cell Viability Assays for A $\beta$ (1-40) Co-Incubated with <i>h</i> PFD Subunit 6 .....	284

# Acknowledgements

---

I would like to express my gratitude to all of the people who have helped, supported and encouraged me during my time as a post graduate student. This also includes the extra period of volunteer required for my PhD.

First and foremost I would like to thank my supervisor and mentor, Prof. Tara Pukala for her guidance and encouragement during my time as a PhD student. This project has presented me with many interesting challenges and opportunities for which I am extremely appreciative. Your enthusiasm and passion for research is contagious, and I look forward to carrying that passion into the next phase of my scientific career. In addition, I would like to thank my external supervisor Prof. Grant Booker for helping with everything biochemistry and for taking me in as a student.

To the other members of the Pukala group, past and present, it has been a pleasure to work alongside you. Everyone has been extremely friendly and always willing to lend a hand or offer advice on a technical problem. Thank you for being such a wonderful bunch of people not only to work with, but also to enjoy a beer with. In addition, I must acknowledge the help and advice given by Yanqin Liu and Yin Ying Ho on all manner of issues related to my research – their problem solving skills have been invaluable and without which I may not have completed my PhD.

My family and friends have been a constant source of support and motivation during my studies. In particular, I would like to thank my best friends; Sam, Tu, Cat, Amy, Mel, my family; Stef, Jessica and Danny. They have always encouraged me to do what makes me happy, and their belief in me has kept me going during some of the most stressful times.

I would also like to thank my partner in crime, Max. For pushing me to be my absolute best and encouraging me to take leaps and bounds. I cannot imagine continuing my passion for science without you by my side.

Last, but certainly not least, I would like to thank my beautiful mum, who has been completely unwavering in support, no matter how difficult the situation. Mum, you have always been there for me during the best and worst of times. Thank you for everything that you have given me and the sacrifices that you have made. I wouldn't be the person I am today without you.

# 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 Australian Government Research Training Program Scholarship.

# Abstract

---

Prefoldin is a hexameric protein complex ubiquitously expressed and found to influence the conformation of amyloidogenic peptides. Relatively high degrees of sequence identity and conservation across evolutionary lineages are observed, however differences in binding abilities have been noted between the homologs. This thesis describes work examining the structure of eukaryotic prefoldin and its biological activities with respect to interaction with amyloid  $\beta$ . The structure and biological activities of prefoldin's individual subunits are also explored.

Although many studies have investigated the structure of prokaryotic prefoldin, there is limited information available for eukaryotic prefoldin. Two-dimensional  $^1\text{H}$ - $^1\text{H}$  and  $^1\text{H}$ - $^{13}\text{C}$  nuclear magnetic resonance (NMR) spectroscopy was utilised to probe the structure of both  $\alpha$  and  $\beta$  human prefoldin subunits. The data revealed the highly alpha helical secondary structure of the subunits, which was further verified through far-UV circular dichroism. Further thermal aggregation assays utilising this technique have demonstrated the stability of the prefoldin subunits.

The biological effect of prefoldin on the amyloid fibril formation of the Alzheimer's disease related amyloid  $\beta$  peptide was investigated using a combination of dye-binding assays and cytotoxicity assays. The presence and absence of fibrils was confirmed by transmission electron microscopy. In terms of fibril formation, prefoldin and its subunits prevented *in vitro* conversion of the amyloid  $\beta$  peptide to amyloid fibrils. In some cases, total inhibition of fibril formation occurred and a 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) assay was conducted on the resultant products. The product was incubated with healthy PC-12 cells and induced cellular death, therefore establishing the cytotoxicity of the resultant oligomeric amyloid  $\beta$  form.

Previous investigations into the binding capabilities of prokaryotic prefoldin identified the distal tips as an important structural aspect, interacting with the amyloidogenic peptide. The binding interface of prefoldin subunits 5 and 6 with amyloid  $\beta$  was probed using chemical cross-linking (CXL) experiments. Traditional methods to identify cross-linked peptides are challenging and the results are often ambiguous. In this study, CXL products were analysed by liquid chromatography-ion mobility-mass spectrometry (LC-IM-MS) to investigate the utility of IM in enhancing the CXL analytical workflow. The orthogonal separation of ion mobility enabled the identification of the cross-linked amino acids. The distal end of prefoldin subunit 5 was found to interact with the N-terminus of the amyloid peptide, whereas prefoldin subunit 6 was identified to interact with the peptide in the middle of its sequence.

Ion mobility-mass spectrometry (IM-MS) analysis of the eukaryotic prefoldin complex identified the collisional cross section of the intact hexamer. Solution disruption experiments of the intact complex revealed the disengaging sub-complexes, and information on the intersubunit contacts and relative interfacial strengths were obtained. A capillary temperature controller (CTC) was developed to observe the thermal dissociation of the complex using nano-electrospray IM-MS.

The combination of these results confirmed a structural aspect common to both mammalian prefoldin and prokaryotic prefoldin, despite the primary sequence differences. The biological assays revealed the ability of prefoldin to prevent the aggregation and amyloid fibril formation of amyloid  $\beta$ , and low resolution MS techniques were able to postulate the arrangement of the subunits and the possible interface interactions of the hexameric complex with the amyloidogenic peptide. This thesis has therefore provided an in-depth investigation of the structural characteristics of eukaryotic prefoldin and its chaperoning capability, therefore implicating a potential role for prefoldin in modulating protein misfolding and aggregation.

# Chapter 1:

## Introduction

---

### 1.1: Overview

In 1956, Francis Crick proposed ‘once information has got into a protein it can’t get out again’ as the central dogma of molecular biology [1]. This was the idea that the flow of genetic information in cells was in one direction, from DNA to RNA to protein. It is this fundamental thought that framed the characterisation of biopolymers and the transfer of biological information. The final products of gene expression are proteins. Responsible for nearly all biochemical processes in the cell, proteins are synthesised *in vivo* by the ribosome in a process known as translation [2]. Following translation the newly synthesised protein folds into its highly defined three-dimensional native conformation to allow it to perform its biological function. The exception to this rule is protein that exists natively disordered. Nevertheless, the correct structure is imperative for structured native proteins that rely on their conformation for activity and crowded intracellular conditions are arduous and can lead to misfolding of the protein. Misfolding results in the loss of original function and the resultant incorrectly folded protein may be inactive or have toxic effects that can lead to debilitating pathologies. To aid in the correct folding, proteins known as molecular chaperones interact with the misfolded molecules and redirect the conformation. A recently described molecular chaperone known as prefoldin (PFD) has been found to prevent the aggregation of toxic misfolded proteins. Understanding the mechanism by which PFD interacts with misfolded proteins could potentially lead to therapeutics for pathologies based on the incorrect folding of a particular protein. This thesis uses complementary low and high resolution structural biology techniques to investigate the structure of PFD and its interaction with amyloid beta ( $A\beta$ ), a peptide known for its misfolding propensity to form fibrils in Alzheimer’s disease.

### 1.2: Proteins

#### 1.2.1: Protein Structure

Proteins are directly involved in all biological processes. Most chemical reactions within the cell are performed, regulated or catalysed by proteins. This includes the transport of ions and electrons, the

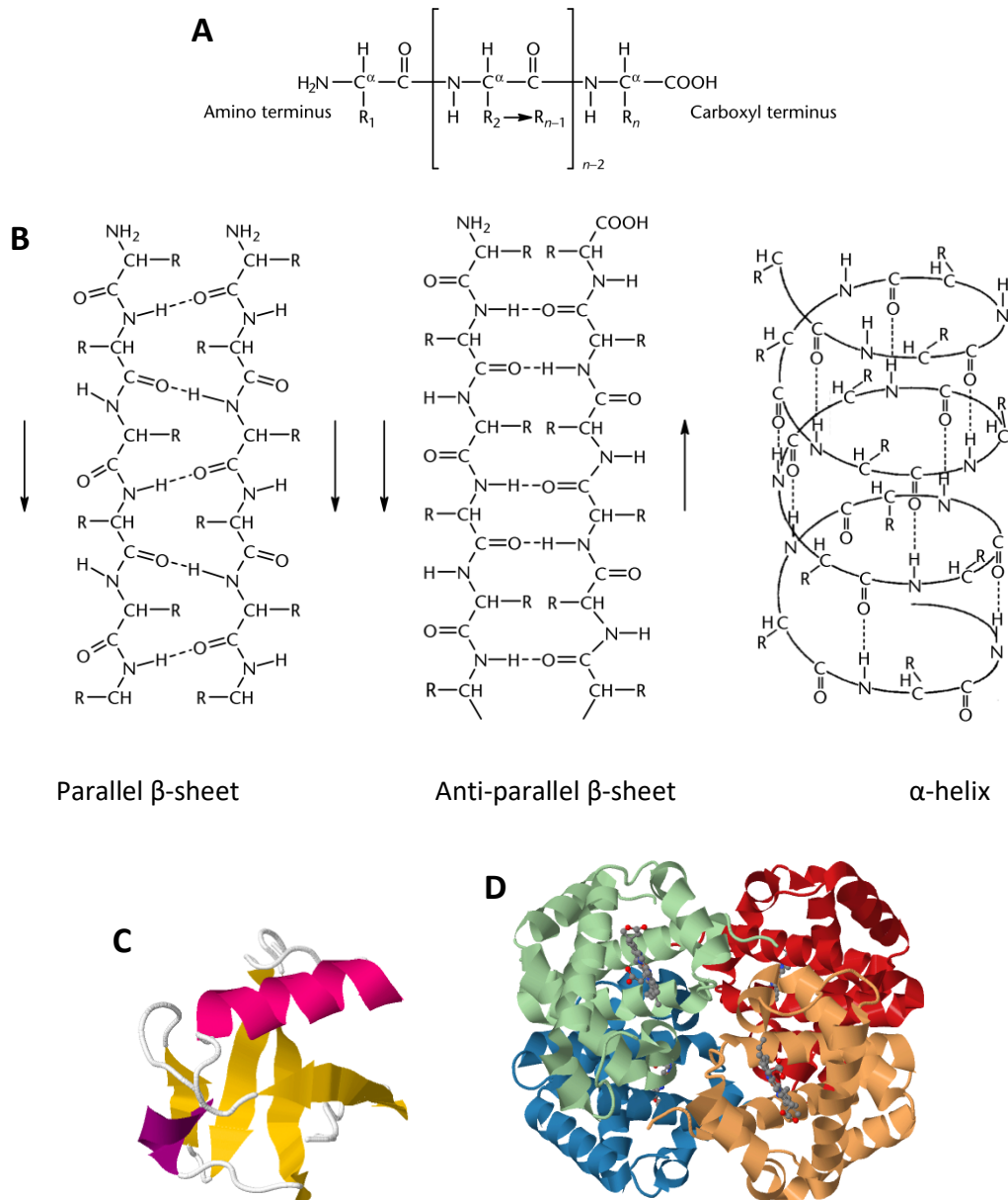
reception and transmission of stimuli, the transcription and translation of genetic code, and the mechanical contraction and coordinated motion within and between cells [3]. In 1962, John Kendrew and Max Perutz were awarded the Nobel Prize for determining the first atomic structure of a protein using X-ray crystallography [4], [5]. They postulated that the native conformation of proteins enabled the biological activity. Whilst this has been subsequently shown to be inherently untrue, as some proteins are able to function despite being intrinsically disordered [6]–[8], the vast majority of proteins rely on their highly defined native 3D structure for bioactivity. Molecular interactions are also a vital facet for biological activity [9] and flexibility of the native conformation is needed for local or global interactions to occur [10]. Bioactivity is therefore influenced by a number of structural factors and some of these include the helicity, amphipathicity, charge state and the nature of the side chains and terminal groups of the protein. These characteristics contribute to the four levels of protein structure; primary, secondary, tertiary and quaternary. For an ordered protein to exhibit its correct biological function it must have correct structures, sometimes at all four levels.

Protein synthesis *in vivo* starts at the ribosome, where the messenger RNA delivers the genetic information to the ribosome to allow the unique sequence to be made. The sequence begins with the amino or N-terminus followed by the linear arrangement of amino acids and ending with the carboxyl or C-terminus (refer to Figure 1.1A). There are over 500 naturally existing amino acids [11] and 20 of these common amino acids (refer to Appendix A) are principally utilised in the human protein code. These are the building blocks of protein with all other amino acids needed for cell functionality in humans obtained from diet. The basic structure of an amino acid can be seen in Figure 1.1A. The substituent on the R group defines the type of amino acid. This single polypeptide sequence made up of amino acids forms the primary structure of the protein. The primary structure is composed of any combination of the common amino acids to a certain degree of length. Therefore, each individual protein has its unique combination of amino acids to a distinctive extent.

As the polypeptide chain is synthesised, the partially formed sequence will begin to fold upon itself to optimise favourable interactions and protect its hydrophobic regions from macromolecules and the aqueous environment in the highly crowded cytosol [12]. This formation of secondary structure is highly dependent on non-covalent, hydrogen bonding within the protein backbone [2], [3]. This electrostatic attractive force between a polar group and a hydrogen bound to a highly electronegative atom is a weak interactive force essential for secondary structure of proteins, as seen in Figure 1.1B. There are two principle types of secondary structures resulting from extensive hydrogen bonding:  $\beta$  sheet and  $\alpha$  helix. Other characteristic features formed within this level of protein structure include turns and loops [13]. Protein chains have a directionality conferred by their

N-terminus and C-terminus and this is generally represented by an arrow towards the C-terminus. Subsequently,  $\beta$  strands are also directional and adjacent strands are able to interact through hydrogen bonding. As seen in Figure 1.1B, parallel  $\beta$ -sheet refers to successive strands arranged in the same direction. This confers a slightly unstable hydrogen bonding pattern as the interaction is non-planar. The antiparallel  $\beta$  sheet arrangement is a more stable form as the successive  $\beta$  strands are orientated in alternate directions, which allows the hydrogen bonds between backbone carbonyl and amide groups to be in their preferred planar position. The  $\alpha$  helix is the most prevalent and predictable secondary structure. Hydrogen bonds are formed between interacting amide hydrogens with the backbone carbonyl group 4 amino acids previous in sequence. This induces a right handed  $100^\circ$  turn every 3.6 residues to give a spiral conformation [14], [15].





**Figure 1.1: Levels of protein structure. (A) The primary structure of a protein [3]. (B) The secondary structure of a protein [3]. (C) The tertiary structure of ubiquitin as represented in ribbons & chains [16]. (D) The quaternary structure of haemoglobin as represented in ribbons and chains [17].**

The structure obtained when the single polypeptide chain has completely folded into all of its secondary structures, in which its hydrophobic amino acid residues are largely buried and its charged hydrophilic residues largely exposed on the surface is referred to as the tertiary structure. Specifically, there are 4 different types of bonding interactions of the side chain which confer stability to the tertiary structure. Hydrogen bonding, salt bridges, disulphide bonds and non-polar hydrophobic interactions cause the folds, bends and loops in the conformation. Hydrogen bonding at the tertiary level occurs between side chains, rather than the backbone as in secondary structure.

Salt bridges can occur with the neutralisation of an acid and amine on side chains within the vicinity of each other causing an electrostatic attraction [18], [19]. Strong covalent disulphide bonds are only formed on the amino acid cysteine through the oxidation of sulfhydryl groups with other nearby cysteines [20], [21]. Finally, non-polar groups on the side chain repel water and polar groups and attract other non-polar groups. This usually results in non-polar side chains arranged in the interior of globular proteins while the outside mainly consists of polar side chains. It is believed that hydrophobic interactions contribute most to the stability of tertiary protein structures [22], [23].

Finally, the last level of protein conformation is the quaternary structure, which is the combination of two or more folded subunits bound together through the same molecular interactions as described. However, due to some necessary unfolding of the protein for functionality, the crowding of organisms in the cell and occasional physiological imbalances to homeostasis, sometimes the native conformation of the protein is compromised and the polypeptide chain misfolds [12].

### 1.2.2: Protein Misfolding and Aggregation

Protein refolding experiments demonstrate that the native conformation of many proteins can be replicated as the highly defined protein structure information is contained in the amino acid sequence, which enables folding and functionality of the protein [24]. The stability of a folded protein is largely based on the burial of its hydrophobic residues, and is a result of the favourable free energy change [25] between unfolded states and native conformations, given by the Gibbs free energy (Equation 1.1);

$$\Delta G = \Delta H - T\Delta S$$

Equation 1.1

where  $\Delta G$  is the change in Gibbs free energy,  $\Delta H$  is the change in enthalpy,  $T$  is the temperature and  $\Delta S$  is the change in entropy. The enthalpic difference is a result of atomic interactions such as electrostatic interactions, van der Waals forces, and hydrogen bonding. The entropy difference is a result of the hydrophobic effect and conformational freedom [26].

According to Anfinsen's dogma, the thermodynamic hypothesis states that the native conformation of a protein is the lowest free energy state [25]. It was therefore believed that all folding processes will eventuate in the lowest energy state, assumedly the native conformation. Although there are experiments that prove Anfinsen's thermodynamic hypothesis correct, metastable conformations with lower free energy than the native state have been encountered [26]–[28]. Due to the many metastable conformations available, Levinthal therefore concluded that protein folding is controlled

kinetically rather than thermodynamically, and proteins traverse a specific folding pathway [2], [29]. The typical funnel-like energy landscapes commonly now described incorporate both theories, and are constructed to demonstrate the energy difference between the conversion of a high energy unfolded protein to the biologically functional native state or to non-functional aggregates through multiple folding pathways [30]–[33]. Figure 1.2 shows formation of the native conformation of a protein at a local minimum, following on-path way folding intermediates via partially folded states. It has been proposed that small proteins (<100 residues) with a single domain undergo a one-step folding process to achieve their native structure [34], whereas large multi-domain complexes require several partially folded intermediates to reach the native state [35]. Some studies have suggested that the rate at which the high energy unfolded proteins assemble affects the propensity for misfolding [36]. Lapses in folding at metastable intermediates expose hydrophobic regions creating opportunities for proteins to form large insoluble aggregates with lower free energy [37]. Factors including transportation across the membrane, functional biological activity and an imbalance in physiological conditions cause the protein to undergo cycles of unfolding and folding, increasing the risk of protein aggregation [38]. This is compounded by the crowded cellular environment, with potential interactions between the side chains and other cellular components increasing the chances of misfolding [39], [40].

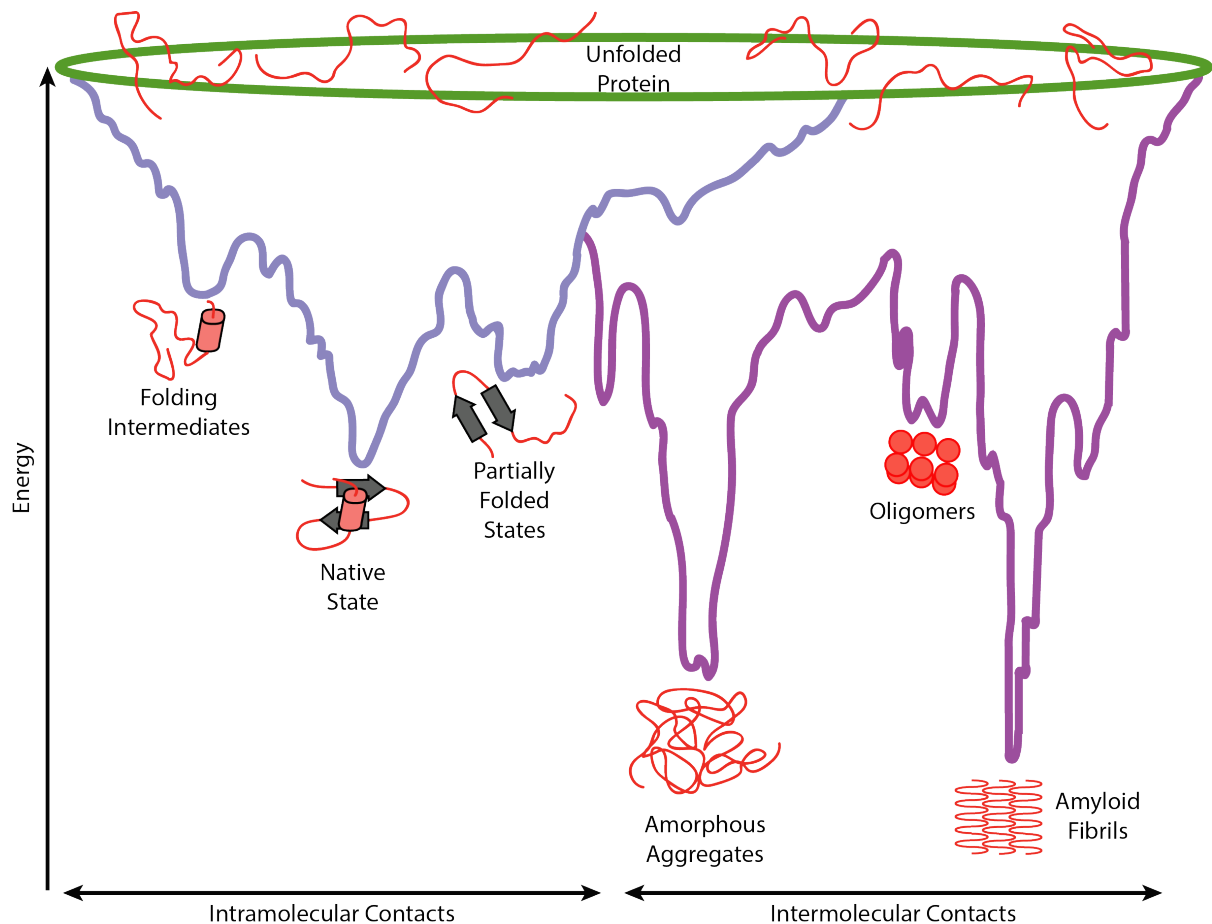


Figure 1.2: The energy landscape in the folding & aggregation of proteins. Through intramolecular contacts (highlighted light purple), intermediates converge towards a native state conformation at a local energy minimum. Intermolecular contacts (highlighted dark purple) can lead to amorphous aggregation or amyloid fibrils. Adapted from [31].

Aggregation and misfolding of a protein can occur via two distinct routes (Figure 1.3) resulting in either disordered amorphous aggregates or highly structured amyloid fibrils. Amorphous aggregation is generally non-toxic and is regulated by the cell under normal conditions. Proteostasis and degradation mechanisms are employed by proteins to clear these aggregates, whereas thermodynamically stable amyloid fibrils are resistant to degradation [38].

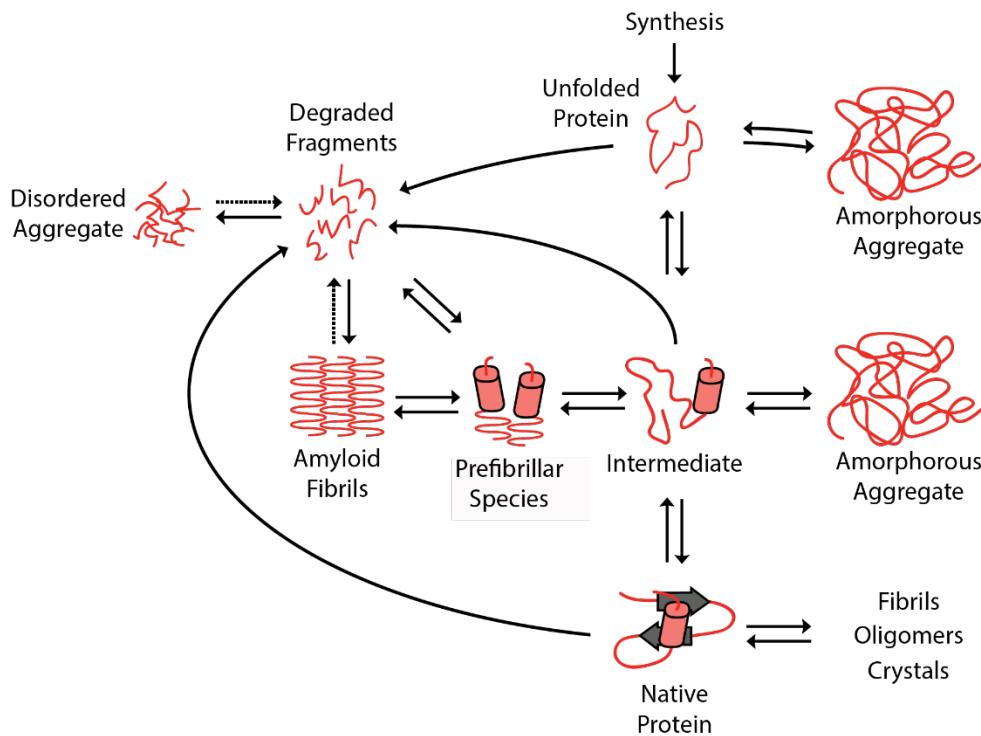
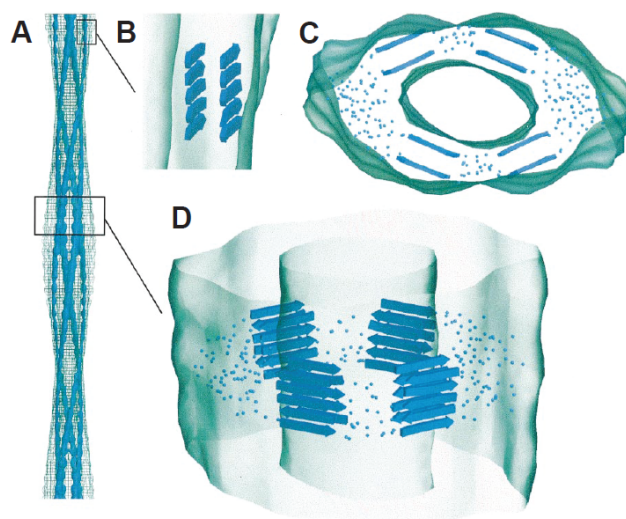


Figure 1.3: Folding of proteins into the native functional state occurs via the production of intermediately folded species on the on-folding pathway in a fast & reversible manner. Under conditions of stress, (e.g. elevated temperature, low pH, high salt concentrations), protein aggregation can occur via two off-folding pathways which can cause the formation of amorphous (disordered) aggregates or highly ordered amyloid fibrils. Adapted from [30].

The highly concentration dependent fibril forming process [41], [42] begins through a nucleation dependent mechanism from a partially folded intermediate. Other partially formed intermediates sequester around this stable nucleus extending and expanding the oligomer into a protofibril [41] (Figure 1.4). The protofibril continues to expand with the addition of more misfolded proteins, increasing in size until an insoluble precipitate is formed [38]. This highly ordered stable fibril is composed of  $\beta$ -strands running perpendicular to the fibril axis, as seen in Figure 1.4.



**Figure 1.4: The structure of amyloid fibrils. (A) The fibril is composed of intertwined protofibrils. (B) A singular protofibril showcasing the stacked  $\beta$ -strands as represented by blue arrows. (C) Top view of the fibril. (D) Side view of the fibril. Adapted from [43].**

The stability of the amyloid fibril is a result of extensive hydrogen bonding. Although all proteins are thought to be potentially capable of forming amyloid fibrils, there are over 40 known proteopathic proteins which are involved in devastating diseases [44], [45]. Table 1.1 lists the different proteins associated with the aetiology of particular diseases, with the majority of pathologies located in the brain.

**Table 1.1: Selected protein misfolding diseases and the associated proteins [46].**

Disease	Protein Deposits	Toxic Protein/Peptide
Alzheimer's disease	Extracellular plaques Intracellular tangles	Amyloid $\beta$ Tau
Cerebral amyloid angiopathy	Extracellular plaques	Amyloid $\beta$
Parkinson's disease	Lewy bodies	$\alpha$ -synuclein
Huntington's disease	Intranuclear & cytoplasmic inclusions	Huntingtin
Fatal familial insomnia	Aggregation	Prion
Amyotrophic lateral sclerosis	Bunina bodies	Superoxide dismutase I
Frontotemporal dementia	Cytoplasmic tangles	Tau
Familial British dementia	Extracellular plaques Intracellular tangles	Amyloid-Bri
Spinocerebellar ataxias	Intranuclear inclusions	Ataxias
Cataracts	Intracellular inclusion bodies	$\alpha$ -crystallin
LECT2 amyloidosis	Amyloidosis	LECT2
Familial amyloid polyneuropathies	Peripheral nerve plaques	Transthyretin

Amyloid deposits found neurologically are hallmarks of selected debilitating neurodegenerative diseases [47]. Autopsies of patients with fibril related diseases revealed a combination of mature fibril deposits as well as early stage soluble oligomers [39], [48], [49] and both have been shown to

have toxic effects on cells [50]–[52]. For a long time it was thought that the fibrils were the main cause of the disease, due to the prevalence of amyloid plaques in patients [53], [54]. Fibrils were found to disrupt the synapses and induce necrosis-like cell death with physical impairments to affected tissue [55]. Other studies suggested that fibril formation is a cellular mechanism for detoxifying the more toxic soluble oligomers by sequestering them into the insoluble and biologically inert fibrils [56]. It has since been suggested that amyloid has the highest degree of cytotoxicity in its oligomeric form [48], [57]. Amyloid oligomers are the prefibrillar multimeric states in amyloid formation that lack the guise of mature fibrils or protofibrils. Conflicting studies describe oligomers having high  $\beta$ -sheet content, while others depict random and coil-like conformations [53], [58]. Although the cytotoxic mechanisms of the amyloid related oligomers are relatively unknown, some possibilities have been formulated. Suggestions include the soluble oligomer disrupting the cell membrane to form a pore like structure resulting in unregulated membrane permeability. Studies have shown an increase in ion permeability occurring in the presence of oligomers with a cytosolic influx of  $\text{Ca}^{2+}$ , which can ultimately cause synaptic degeneration and cell death [59], [60]. Cell death can also occur from caspase-mediated apoptosis, mitochondrial dysfunctions and the production of toxic and reactive oxygen species [38], [48], [55], [60], [61].

Although amyloid fibrils are resistant to degradation, there are methods which the cell employs to prevent the formation of the toxic oligomers and the highly stable fibrils. The most effective and well-studied protective mechanism found to date is the use of molecular chaperones.

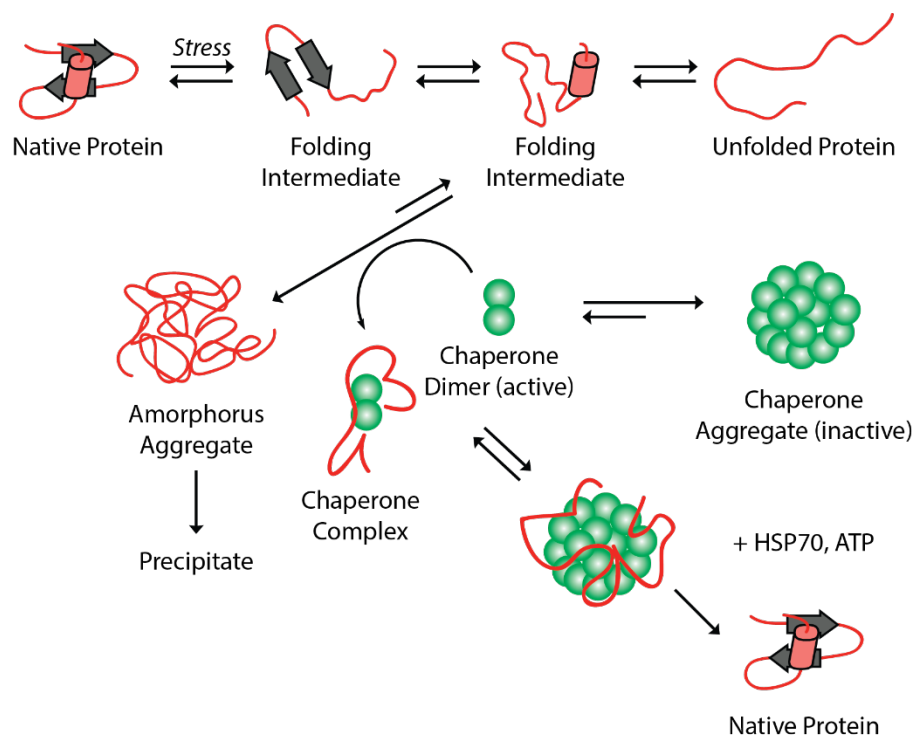
### 1.2.3: Molecular Chaperones

Molecular chaperones are defined as proteins that interact and aid in the correct folding process of highly dynamic molecules, principally other proteins. This includes interacting with partial sequences formed as the protein substrate is synthesised on the ribosome. The newly synthesised polypeptide chain undergoes several necessary folding cycles to reach the correct native form. Unfortunately the continuous folding and unravelling of the sequence exposes vulnerable hydrophobic regions and increases the propensity for misfolding and aggregation. Molecular chaperones are expressed constitutively [62] and recognise and intercept the partially folded intermediate to reinitiate the correct folding pathway.

Chaperones that are upregulated during conditions of cellular stress such as heat and changes in pH levels are referred to as heat-shock proteins (HSPs) and are classified in accordance with their molecular weight; HSP40, HSP60, HSP70, HSP90, small HSP (sHSP) [63], [64]. HSPs attempt to combat protein misfolding [65], [66] and regulate the protein folding equilibrium, ultimately

preventing the aggregation of misfolded species which could potentially lead to pathologies [67]–[69]. Although HSPs exhibit little structural or sequence similarity [64], their common feature of markedly increased expression in response to a homeostasis imbalance suggests that chaperones work together toward the correct shaping of the proteome.

Figure 1.5 provides a schematic diagram of the mechanism by which a dimeric sHSP intercepts a partially folded intermediate on the off-folding pathway. The chaperone complex then interacts with Hsp70, which forms a larger complex and the refolding process is initiated by the addition of chemical energy provided from the hydrolysis of adenosine triphosphate (ATP). The refolded native conformation is then released from the molecular chaperones. The chaperones do not contribute to the biological function of the protein and no longer have an active role after the intended protein has reached the native state [62].



**Figure 1.5: Mechanism of sHsps chaperone activity.** The dimeric sHsp intercept the intermediate folding protein on the amorphous off-folding pathway, with the aid of Hsp70 & the addition of ATP hydrolysis, the native conformation of the protein is achieved. Adapted from [70].

HSP60 molecular chaperones (also referred to as chaperonins) are classed by their intrinsic rotationally symmetrical double ring assembly. The cylindrical structure contains a central cavity that operates in two different stages to facilitate the importing of protein and redirecting of



macromolecular assemblies in the mitochondrial matrix [71], [72]. In the binding active state, one end of the cylindrical complex is open to the cytosol, exposing the hydrophobic lining that attracts the hydrophobic residues on the surface of non-native proteins [71]. A co-chaperone complex creates a lid that binds over the opening of the cavity to isolate and initiate the refolding active state through the expenditure of ATP [12], [73]. The lid disengages and the macromolecule is released following the correct refolding. Chaperonins are further classified into two separate groups based on their location. Group I chaperonins are found in bacteria and organelles of eukaryotes, and group II chaperonins in archaea and the cytoplasm of eukaryotes [74].

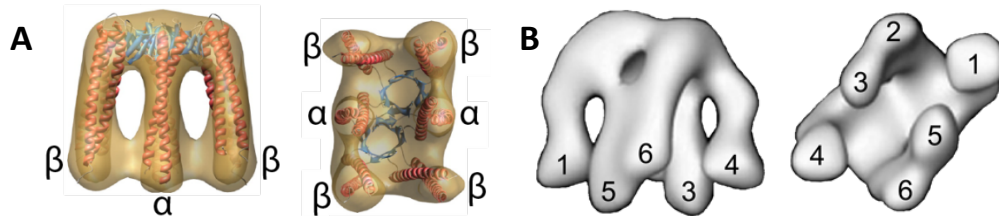
The group II chaperonin CCT/TRiC (chaperone containing TCP-1 or tailless complex polypeptide 1 ring complex) consists of a double barrel, with each ring composed of eight distinct but similar subunits (27-39 % sequence identify) [75], [76]. It is essential for cell viability as it refolds many essential proteins including actin and tubulin. The chaperone PFD/GimC (prefoldin or genes involved in microtubule bio-genesis complex) was genetically [77] and biochemically [78], [79] identified to assist CCT in the maturation of actin and members of the tubulin family [77]. An interest in the heterogeneous co-chaperone complex prefoldin has recently arisen due to its unique chaperoning abilities.

#### 1.2.3.1: Prefoldin

Prefoldin (PFD) is a ubiquitous protein associated with the folding of actin and has recently attracted considerable research interest because it was found to assist in the folding of peptides implicated in the aetiology of neurological amyloid disorders [80]–[82]. The mechanism and structure of *bovine* prefoldin (*bPFD*) are yet to be fully elucidated. Many of the previous structural studies of PFD are based on the prokaryotic complex [73], [83]–[87] however, there are a few biological studies that utilise the *human* prefoldin (*hPFD*), synthesized through recombinant protein expression [88], [89].

PFD is present in all eukaryotes [90], [91], which suggests that its ubiquitous presence in the eukaryotic kingdom is archaeal in origin [92]. The crystal structure of archaeal PFD was determined by Hartl *et al.* and the complex was found to consist of 6 subunits shaped in the form of a jelly fish with a double  $\beta$  barrel body and 6 protruding tentacles [93]. Archaeal PFD was identified to comprise of 2 identical  $\alpha$  subunits and 4 identical  $\beta$  subunits (Figure 1.6A). Native mass spectrometric analysis of the macromolecule validated the quaternary arrangement of the 2 central  $\alpha$  subunits and 4 surrounding  $\beta$  subunits [94]. In comparison, eukaryotic PFD has been found to have 6 unique subunits, with 2 different  $\alpha$  subunits and 4 different  $\beta$  subunits [89] (Figure 1.6B). Assumptions of a

similar quaternary arrangement to its prokaryotic counterpart have been made with electron microscopy images revealing a similar archetype [95].



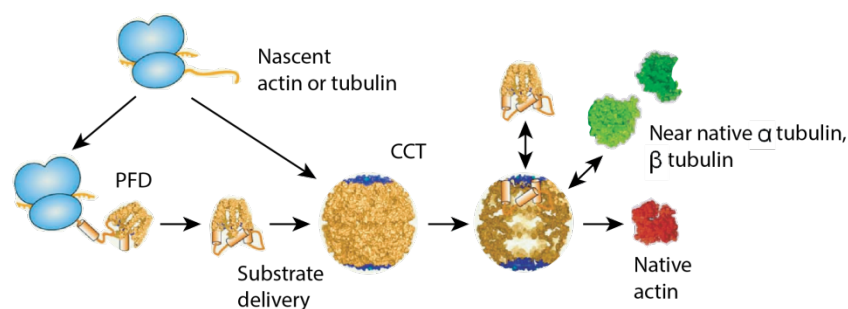
**Figure 1.6: Structure of PFD variants. (A) The 3D reconstruction of prokaryotic PFD from electron micrographs,  $\alpha$  &  $\beta$  subunits are indicated. (B) The 3D reconstruction of human PFD from electron micrographs. Adapted from [91] & [95].**

The 6 easily accessible tips of the coiled tentacles have hydrophobic regions required for chaperone activity [93]. The extending coils form a rectangular cavity that enables encapsulation of the misfolded protein. Electron microscopy images of eukaryotic PFD–actin complexes suggest a snug fit of actin within the cavity [91], [95]. It is proposed that eukaryotic PFD has evolved a specialised binding surface within its cavity for actin, as prokaryotic PFD has been found to bind substrates solely at its distal tips without encapsulation [91], [96]. The binding sites on the tentacles are therefore able to stabilise non-native proteins of different shapes and sizes reflecting the broad chaperoning interactions archaeal PFD has *in vivo*.

Although neither actin or tubulin are present in archaea, PFD, group II chaperonins and sHSP chaperones have been identified in the prokaryotic molecular chaperone system [79], [97]. Many studies replace the eukaryotic PFD with archaeal PFD due to the easy extraction and synthesis of the bacterial complex. Co-incubation of PFD from the archaeal *Methanobacterium thermoautotrophicum* (mtPFD) with non-active actin was found to stabilise the non-native proteins and release them for subsequent CCT dependent folding *in vitro*. Similarly mtPFD was able to fold chemically unfolded rhodanese with the addition of HSP60 and ATP *in vitro* [98]. PFD from archaeal *Pyrococcus horikoshii* (phPFD) was found to deliver bound unfolded proteins rapidly to both group II and group I chaperonins, suggesting that PFD recognises common characteristics present in both chaperonins for transfer of unfolded protein [73]. However, PFD exists solely in the cytosol of eukaryotes and archaea and therefore is proposed to cooperate exclusively with group II chaperonins *in vivo* [99].

Eukaryotic PFD plays an important role in motility and sensory processes within the cell through the mediation of actin and tubulin. Actin and tubulin form microfilaments and microtubules of the

cytoskeleton, a complex network involved in numerous cellular processes. These include cellular signalling pathways, endocytosis, cytokinesis and intracellular support [97]. Actin and tubulin therefore represent a large proportion of the total protein in most cells and are highly concentrated in the cytosol [99]. High concentration, compounded with the propensity to self-associate and the inability to fold unaided, is a problem for the biogenesis of the cytoskeleton as both proteins compete for access to the limited chaperone CCT [97], [99]. Eukaryotic PFD binds to the emerging nascent actin or tubulin chain from the ribosome and delivers the partially folded substrate to CCT by docking and directing the substrate hand-off (Figure 1.7). Some studies suggest PFD acts as the transient lid for CCT after substrate delivery [100], remaining bound to the chaperonin and encapsulating the substrate. Within the cavity of CCT, the addition of ATP hydrolysis enables actin to fold into its native conformation and upon correct folding is released [79], [89], [95], [97]–[99], [101]. The mediated folding pathway was investigated after the discovery of PFD and it was found that CCT incubated with actin bound to PFD enabled fast efficient refolding whereas incubation of CCT and actin without the co-chaperone PFD produced slow refolding pathways [77], [79], [100] further validating the co-chaperoning abilities of PFD.



**Figure 1.7: Schematic diagram of eukaryotic PFD chaperone activity on actin. PFD intercepts the nascent folding polypeptide chain & transfers the intermediate to CCT for correct folding. Natively folded actin is achieved with the addition of ATP hydrolysis. Adapted from [99].**

Attention to the efficient co-chaperoning abilities of PFD has been diverted to misfolding peptides and proteins implicated in pathologies. This has generated significant interest for the development of a potential therapy against these debilitating diseases. As mentioned previously, a plethora of diseases are associated with the misfolding of specific peptides or proteins. Recent studies exploiting the chaperoning abilities of proteins have targeted peptides implicated in Huntington's disease [102], Parkinson's disease [80] and Alzheimer's disease [81], [82].

Of particular interest is the chaperoning ability of PFD towards amyloid  $\beta$  ( $A\beta$ ) in Alzheimer's disease. There have been conflicting reports on the toxicity of the resultant amyloid  $\beta$  forms chaperoned by homologs of PFD. It has been reported that  $A\beta$  incubated in the presence of *ph*PFD forms soluble oligomers that are toxic to cultured rat pheochromocytoma cells [81]. However,  $A\beta$  in the presence of recombinant human PFD was found to form non-toxic soluble oligomers [82]. This could suggest that there is a difference in chaperoning abilities between PFD homologues, and the resulting oligomeric species of  $A\beta$  differ in cytotoxicity.

In a study by Takano *et al.*, cytotoxicity is exhibited by the aggregation of  $\alpha$ -synuclein in neuro-2a cells with the depletion of eukaryotic PFD subunits 2 and 5 in mice [80]. However, although the knockdown of eukaryotic PFD subunits 2 and 5 disrupts the complex within the cell, this does not necessarily deplete the eukaryotic PFD subunits 1, 3, 4 and 6. It is prudent to acknowledge the biological abilities of the individual PFD subunits as they have been found to exhibit their own unique functions. *h*PFD subunit 1 is required for lymphocyte development and function [103] and has been found to redistribute into cells under duress [104]. It is also a vital component of the *h*PFD complex as it binds specifically to CCT and transfers non-native proteins using its basic amino acid residues [83]. Individual *h*PFD subunit 3 is found to form complexes with a tumour suppressor gene and is translocated from perinuclear granules to the nucleus or cytoplasm [105]. Lastly, PFD subunit 5 has been found to be a tumour suppressor [106].

Due to the difficult extraction of the eukaryotic PFD and its inability to crystallise, a minimal amount of structural information has been obtained. Although a plethora of data is available for the bacterial homologues of archaeal PFD, it has been observed that binding and chaperoning interactions between the homologues differ. Furthermore, although individual PFD subunits exhibit distinct functions, the chaperoning ability of the individual subunits has not yet been explored. Therefore, the structure and biological properties of the intact eukaryotic PFD and its individual subunits will be investigated in this thesis, along with the interactions with  $A\beta$  and the resultant amyloidogenic products.

### 1.3: Aims and Outcomes

The overall aim of the work described in this thesis was to further characterise and identify the chaperone activity of PFD. This work aimed to elucidate the topology of eukaryotic PFD through a series of biophysical techniques to better understand its structure and chaperoning interactions in order to further understand the potential role of PFD in protein aggregation diseases.

As described in Chapter 3, *b*PFD complex was extracted from bovine testis and the individual *h*PFD subunits 5 and 6 synthesised through recombinant protein expression. Following optimisation of the purification protocol, the chaperoning abilities of *b*PFD and *h*PFD subunit 5 and 6 were investigated. Analysis of their chaperone function against aggregating amyloid beta 1-40 was examined through a variety of biophysical and spectroscopic techniques.

In Chapter 4, far-UV circular dichroism analyses of the *b*PFD complex and the *h*PFD subunits 5 and 6 were examined to determine the secondary structures of these chaperones. To obtain additional information on the stability of the complexes, the effect of pH and temperature were examined. The dynamic interaction of the chaperone with amyloid beta (1-40) is also described.

Chapter 5 details the high resolution structure investigation of *h*PFD subunits 5 and 6 through 2D and 3D NMR spectroscopy.

In Chapter 6, ion mobility-mass spectrometry was utilised to determine the overall topology of the *b*PFD complex. Solution disruption and investigations into the effect of pH and temperature complement the structural investigations.

As described in Chapter 7, chemical cross-linking techniques were implemented to further investigate the structure of the individual *h*PFD subunits 5 and 6 and the chaperone interface of each subunit with amyloid beta (1-40), providing an insight into the chaperoning mechanism of PFD.

Overall, these studies further our knowledge about the structure and molecular chaperone action of human PFD and its interaction with aggregating target proteins. This will enable us to further explore the therapeutic potential human PFD has against protein aggregation diseases.

# Chapter 2:

## Biophysical Techniques to Determine Protein Structure and Interactions

---

### 2.1: Introduction

It is well known that a protein *in vivo* will generally have its biological activity and molecular interactions determined by its highly defined 3D structure. It is therefore necessary that we study the interactions at play and the complex structures that govern these activities. Atomic level resolution structure determination techniques of X-ray crystallography, cryo-EM and Nuclear Magnetic Resonance (NMR) spectroscopy are often used to determine protein structure [5], [107], [108]. Unfortunately, acquiring spectra of large complexes can be difficult and often crystals cannot be formed for crystallography. As a result, combinations of complementary low resolution methods are regularly used instead. These techniques include electron microscopy [109], [110], small-angle X-ray scattering (SAXS) [111], circular dichroism (CD) spectroscopy [112] and mass spectrometry (MS) based methods [113], [114]. Often, a combination of both high and low resolution techniques allows the elucidation of highly complex structures and interactions. This chapter describes the biophysical techniques used in this thesis to determine protein structure and protein-protein interactions.

### 2.2: Biophysical Techniques

#### 2.2.1: ThT Assay

Thioflavin T (ThT) assays are commonly used to determine the formation of  $\beta$ -sheet rich structures through the fluorescence of the ThT dye. This cationic benzothiazole dye shows enhanced fluorescence upon binding to the secondary structure synonymous with amyloid fibrils. ThT dye binds to cross  $\beta$ -sheets in a non-specific fashion and exhibits an intrinsic shift in its excitation maximum (from 385 to 440 nm) and emission maximum (from 445 nm to 480 nm) [115], [116] by several orders of magnitude. This unique characteristic allows the lag, elongation and plateau phases of amyloid fibril formation *in vitro* to be monitored as the change in fluorescence intensity is analysed by fluorescence spectroscopy, as seen in Figure 2.1 [116], [117].

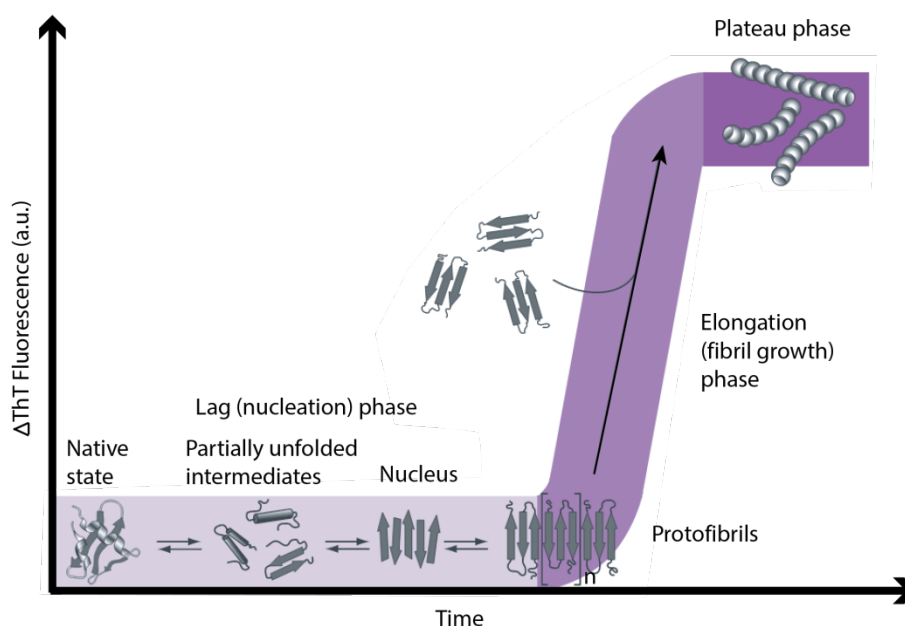


Figure 2.1: The formation of amyloid fibrils monitored by ThT fluorescence. Fibril formation begins with nucleation in which soluble oligomers are formed during the lag phase from misfolded intermediates. Intermediates begin to sequester to form the early assemblies of insoluble mature protofibrils that lead to rapid growth in the elongation phase. The plateau phase occurs at complete fibril formation.  $\Delta$ ThT fluorescence increases with the increase in  $\beta$ -sheet content. Adapted from [118].

To study amyloid fibril formation, the fluorescence is monitored over time and as the intermediates aggregate following the lag phase, the protofibrils precipitate out of solution and become insoluble. The ThT dye readily binds to the  $\beta$ -sheet secondary structure and this is observed by an increase in fluorescence intensity, until the plateau phase signals the end of fibril formation [115]–[117]. The typical sigmoidal curve is the result of the fluorescence intensity change during the lag, elongation and plateau phases [118]. ThT assays are an indirect method of determining fibril formation and are able to give insight into the secondary structure of amyloid fibrils.

Cell viability studies are sometimes conducted in conjunction with ThT assays to determine the biological effects of species present at different stages on the amyloid forming pathway. Cell viability assays predominantly use tetrazolium salts for colorimetric analysis to determine the metabolic activity of a cell [119], [120]. For example, the dye 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) can be reduced to the insoluble purple formazan by the NAD(P)H-dependant cellular oxidoreductase enzyme and dimethyl sulfoxide (DMSO) is added to dissolve the reduced dye [119]. The oxidoreductase enzyme accumulates in the cytosolic compartment of the cell and MTT is reduced to formazan. The measurement of absorbance at the wavelength 560 nm allows for the quantification of the metabolic activity of cells. Although many published papers and this thesis assumes the reduction of dye is accompanied by cellular death, it

must be noted that these assays are not direct representations of cell death [120]. However, for clear interpretations, cell viability is typically assessed as the percentage absorbance relative to the vehicle control as a mean of three independent experiments.

### 2.2.2: NMR Spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy is commonly used to characterise structures ranging from small molecules to large complex proteins. The notion of elucidating molecular skeletons of molecules through NMR was first introduced in the 1960's [121]. Since then, Fourier transform (FT) approaches have been incorporated in order to enhance sensitivity [122]. The development of 2-dimensional (2D) NMR spectroscopy enabled the elucidation of the first 3D structure of a protein [123], and to combat ambiguous 2D NMR data of large complex macromolecules, a third dimension is also commonly incorporated into the biological NMR methodology (3D NMR).

All nuclei possessing a non-zero nuclear spin quantum number ( $I$ ) can be analysed by NMR spectroscopy. For biological systems, the nuclei observed by NMR typically include  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  [123]. In the presence of an applied magnetic field ( $B_0$ ), the spins are affected, and the nuclei can be oriented in  $2I+1$  ways within the field. Given the nuclear spin quantum number of  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  equals  $\frac{1}{2}$ , there are two orientations the spin may adopt; antiparallel to  $B_0$  with high energy or aligning parallel to  $B_0$  with low energy (Figure 2.2).

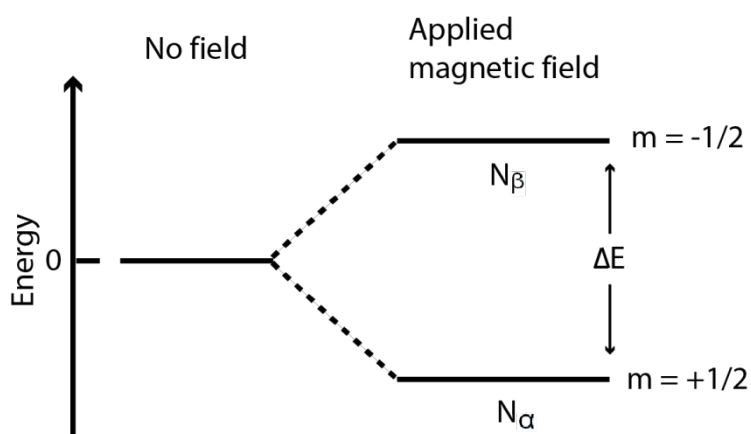


Figure 2.2: Energy levels for a nucleus with spin quantum number  $\frac{1}{2}$  & the effect of an applied magnetic field on the nuclei.  $N_\beta$  &  $N_\alpha$  refers to the population of the energy orientations &  $m$  refers to the magnetic quantum number.

The energy difference between the spins is given by Equation 2.1:



$$\Delta E = \frac{h\gamma B_0}{2\pi} \quad \text{Equation 2.1}$$

where  $h$  is Planck's constant,  $\gamma$  is the gyromagnetic ratio (a proportionality constant, differing for each nucleus) and  $B_0$  is the strength of the applied magnetic field. Nuclei can move from  $N_\alpha$  to  $N_\beta$  through the irradiation of the nucleus with an electromagnetic radiation corresponding to  $\Delta E$ . The two energy levels have different populations and this difference determines the sensitivity of the technique. To increase sensitivity and therefore  $\Delta E$ , the application of higher field strengths is used. The initial ratio of populated energy orientations is determined by the Boltzmann distribution, given by Equation 2.2:

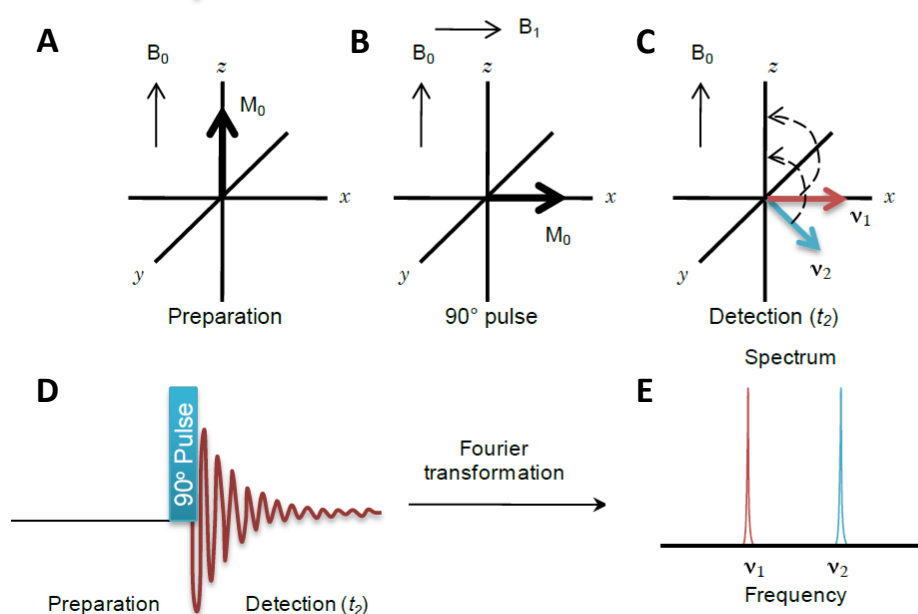
$$\frac{N_\alpha}{N_\beta} = e^{-\Delta E/kT} \quad \text{Equation 2.2}$$

where  $k$  is the Boltzmann constant and  $T$  (in K) is the temperature. The population difference results in the formation of a net macroscopic magnetisation  $M$  (i.e. the sum of the magnetisation of all the individual spins).  $M$  can be altered with the application of a radiofrequency (RF) pulse that corresponds to  $\Delta E$ . The RF pulse allows the transition of nuclei from the  $N_\alpha$  state to the  $N_\beta$  state and the change in population results in a change in  $M$ . The frequency at which the nuclei naturally precess in a magnetic field is referred to as the resonance or Larmor precession frequency ( $\nu$ ) and is defined by Equation 2.3:

$$\nu = \frac{\gamma B_0}{2\pi} \quad \text{Equation 2.3}$$

Application of  $B_0$  along the  $z$  axis will generate a magnetisation in the same direction, as depicted in Figure 2.3A. RF pulses of short duration are then applied, generating an oscillating magnetic field with the magnetic component of the pulse in the  $x$ -axis ( $B_1$ ), projected at  $90^\circ$  to  $B_0$  [124]. The generated field exerts a force on  $M_0$ , tipping the magnetisation towards the  $x$ -axis, and as long as it continues the nuclei will preferentially precess about this direction [125]. The length of application of  $B_1$  determines the angle by which the magnetisation is rotated. Irradiation with  $B_1$  is commonly halted at a magnetisation rotation of  $90^\circ$ , called a  $90^\circ$  pulse (Figure 2.3B) [125]. Following the pulse application,  $M$  relaxes back to equilibrium and the magnetisation component along the  $y$ -axis decays exponentially with time (Figure 2.3C). The decaying cosine wave produced from this relaxation is referred to as free induction decay (FID) and occurs by two relaxation processes, namely spin-spin relaxation and spin-lattice relaxation [125]. This is monitored along the  $y$ -axis through the detection phase of the experiment ( $t_2$ ). The cosine wave generated in the time domain is converted into frequency signal through Fourier transform to produce a 1D NMR spectrum as seen in Figure 2.3D.

Each chemically distinct nucleus has a different resonant frequency caused by the electrons surrounding the nucleus. The electrons create a secondary magnetic field that opposes  $B_0$  and in regions of high electron density the applied field experienced by the nucleus is weaker, and therefore lower field strength is required to bring the nucleus into resonance. Nuclei within distinct chemical environments therefore experience resonance at different frequencies in the NMR experiment [121], [126], [127]. This chemical shift ( $\delta$ ) is also therefore affected by conformation. A nucleus whose chemical shift has decreased due to a high electron density is termed shielded and a chemical shift that has increased due to the removal of electron density is termed de-shielding. As each chemically distinct nucleus has a different resonant frequency, the RF pulse contains radiation which spans a large frequency range in order to excite all the nuclei in the sample at once. The FID obtained in  $t_2$  contains multiple frequency components due to the different Larmor precession frequencies of chemically distinct nuclei. Therefore upon Fourier transformation, the NMR spectrum will show a unique signal for each of the chemically distinct nuclei of interest in the molecule.



**Figure 2.3:** (A) Schematic diagram of the initial orientation of  $M_0$ . (B) Application of a  $90^\circ$  RF pulse rotates  $M_0$  into the  $xy$ -plane. (C) Precession of spins in the  $xy$ -plane as  $M_0$  relaxes back to the equilibrium position is recorded as a free induction decay. (D) The cosine wave generated in the time domain is converted into frequency signals through Fourier transform to produce a 1D NMR spectrum. (E) The cosine wave transformed into a 1D NMR spectrum via Fourier transformation. Reproduced with permission from [128].

Although the standard 1D NMR experiment is particularly useful for characterising the structure of small molecules, it is challenging to interpret for biological samples which are large and complex,

containing many chemically distinct nuclei. The development of multi frequency dimensions has significantly increased the information content available from NMR data. 2D and 3D NMR experiments have a second and third frequency dimension, respectively, enabling the overlapped resonances to be separated and therefore yielding spectra of increased resolution. The development of multinuclear experiments has also allowed structural problems of significantly increased complexity to be addressed using NMR spectroscopy [129]–[131]. Standard 1D NMR experiments occur with only  $90^\circ$  pulses before the  $t_2$  phase, separating the spins (Figure 2.3B). The addition of evolution ( $t_1$ ) and mixing ( $\tau_m$ ) periods between the preparation and  $t_2$  stages give rise to 2D NMR experiments [132], [133]. Delays and RF pulses construct the different phases of  $t_1$  and  $\tau_m$ , ultimately generating the pulse sequence. 2D data sets require several experiments recorded with incremented  $t_1$  periods, during which magnetisation transfer occurs. Different pulse sequences relay different information depending on the transfer of magnetisation, which can occur between nuclei covalently bound (spin-spin coupled) or between nuclei close in space (dipolar coupled spins). A contour plot allows the intensity of the peaks to be quantified [133]. Information from the primary structure to the quaternary structure can be deduced from this information. A schematic representation of the 2D NMR experiment is seen in Figure 2.4.

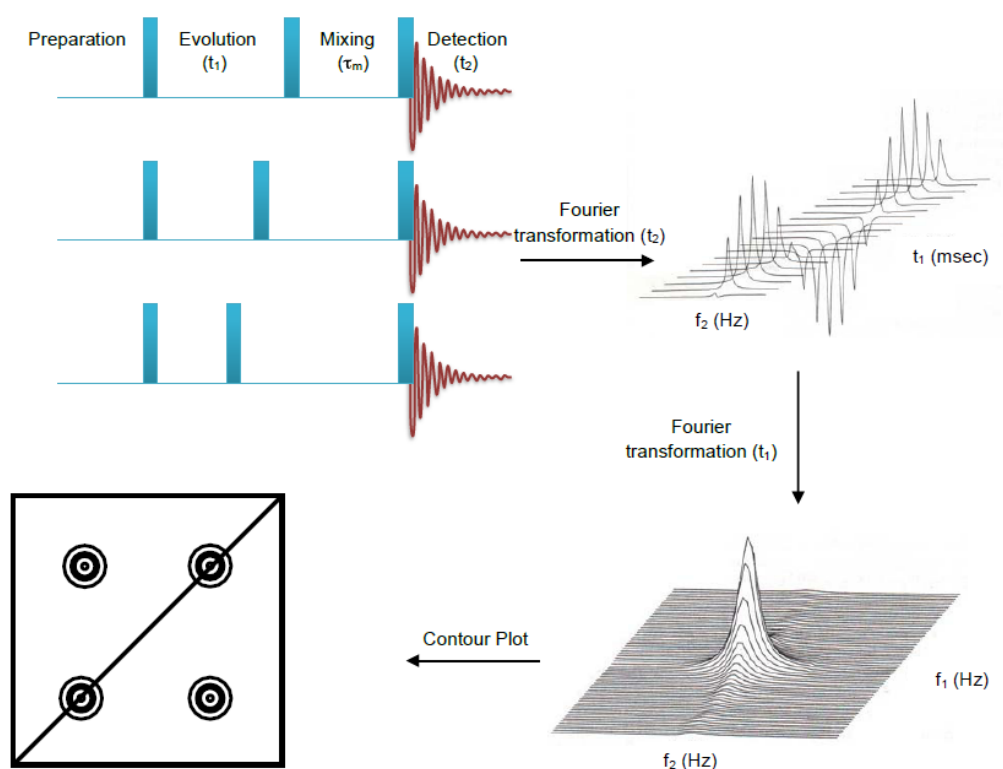


Figure 2.4: Schematic representation of the 2D NMR experiment. Different sequences of the  $t_1$  &  $\tau_m$  are detected in  $t_2$  & Fourier transformation converts these signals into a contour plot on the frequency scale. Adapted from [128]–[130].

A number of 2D NMR experiments are required to determine the 3D structure of proteins and peptides utilising different pulse sequences, due to the different structural information obtained from these experiments. Total correlation spectroscopy (TOCSY) experiments show cross magnetisation between scalar coupled (through-bond) spins, therefore ultimately correlating these protons [134]–[137]. A set of nuclei that share no coupling with the nuclei outside of that particular system is referred to as a spin system. Scalar coupling does not occur across double bonds, and due to the partial double bond characteristic across the peptide bond, TOCSY experiments of proteins will show all proton resonance correlations of each separate amino acid residue as a single spin system unless there are aromatic groups present. For example, Phe contains 2 spin systems and Trp contains 3 spin systems as their aromatic proton signals exist as separate spin systems [10], [24]. Each amino acid residue displays a characteristic pattern of connectivities and therefore analysis of the coupled spins in the TOCSY spectrum allows the amino acid identity to be determined [126]. For a relatively small protein, almost all of the amino acid residues can be identified through the analysis of the TOCSY spectrum, however, ambiguity arising from overlapping signals in the spectrum requires further 2D NMR experiments, namely correlated spectroscopy (COSY) [133]. COSY experiments show spin-spin coupling of protons that are only two or three bonds apart and can therefore assist to unambiguously assign resonances [126], [132], [133]. Nuclear Overhauser effect spectroscopy (NOESY) experiments are then used to connect the spin systems together.

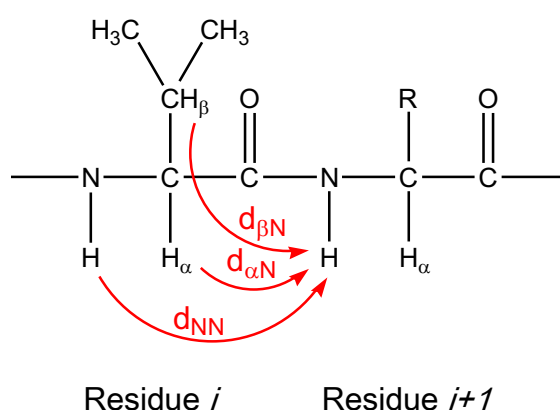
NOESY experiments exploit the Nuclear Overhauser Effect (NOE) to assign connecting amino acid residues in a polypeptide and provide distance restraints for structure calculations. NOE signals occur when a selected resonance is irradiated and cross-relaxation (dipolar coupling) changes the intensity of NMR resonance from a nearby nucleus [126], [138]–[140]. The correlation between the relative intensity ( $I$ ) of a NOESY cross-peak and the inter-atomic distance ( $r$ ) of the protons is shown in Equation 2.4:

$$I \propto r^{-6} \quad \text{Equation 2.4}$$

As the distance between the dipolar coupled spins increases, the NOE signal intensity dramatically decreases. Therefore, the maximum observable spatial separation between spins is approximately 5 Å [126], [139].

To assign the NMR resonances of a biomolecule using 2D NMR experiments, a combination of sequential [139] and main chain [141] directed strategies are employed. Sequential assignment commences by determining the identity of each spin system by comparing the chemical shifts with published cross-peak patterns and random coil shifts [126], [142]. 2D TOCSY experiments show correlations of all protons in the amino acid residue with the amide proton, e.g. all resonances

within a spin system. However, TOSCY data of large peptides and proteins may have ambiguous data and signal overlaps. This is counteracted through the acquisition of COSY spectra, as only connectivities of nuclei up to three bonds apart are shown in COSY data. Following the identification of each spin system, the position of each spin system in the protein sequence is determined using correlations in NOESY experiments. As NOESY experiments present cross-peaks of spatially proximal protons, the presence of signals corresponding to protons in successive residues allow the determination of sequential spin systems. This is particularly useful for the determination of protons proximal to the amide protons as shown in Figure 2.5 [130].



**Figure 2.5: NOE connectivities used to sequentially assign spin systems. Inter-residue NOE connectivities are denoted using the nomenclature  $d_{\alpha\beta(i, i+1)}$ , which represents the distance between two protons,  $\alpha$  (on the backbone carbon) &  $\beta$  (on the sidechain carbon) to the proton on the backbone nitrogen on residue *i* & residue *i*+1. NOE connectivities of protons on backbone nitrogens are denoted  $d_{NN(i, i+1)}$ . Adapted from [139].**

The main chain directed assignment was developed following the sequential assignment strategy for simpler interpretation of the 2D spectra. In this strategy, the primary focus is the ordering of the sequential spin systems through the recognition of the distinct predefined NOE pattern of  $\alpha$ -helices,  $\beta$ -sheets, turns, and extended chains before the identification of the amino acid side-chains [141], [143]. Once the spin systems are ordered, this is aligned and assigned to the known amino acid sequence. This approach broaches the ambiguity present when there are many residues with similar chemical shift signatures.

Following assignment of signals in the 2D NMR spectra to distinct nuclei, further analysis of the data through the investigation of secondary chemical shifts, NOE connectivities and coupling constants can be performed to draw structural conclusions. This data can also be combined with computational studies to determine the overall 3D structure.

As the chemical shift is highly sensitive to conformation, NMR can be used to determine regions of secondary structure.  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  NMR experiments for the 20 common amino acids in peptides with no defined secondary structure and prepared in denaturing conditions serve as the bench mark to determine the secondary structures of proteins and peptides [142]. The literature random coil chemical shift value is compared to the observed value and the difference between them (Equation 2.5), termed the secondary shift ( $\Delta\delta$ ), allows the secondary structure elucidation of the polypeptide chains.

$$\Delta\delta = \delta_{\text{observed}} - \delta_{\text{random coil}} \quad \text{Equation 2.5}$$

A positive  $\Delta\delta$  value specifies a downfield shift from random coil values whilst a negative  $\Delta\delta$  indicates an upfield shift [127], [142]. Previous studies of  $\alpha$ -helices have reported negative  $\Delta\delta$  values of  $\alpha^1\text{H}$  and NH with  $^1\text{H}_\alpha$  secondary shifts in the order of 0.39 ppm, whilst  $^{13}\text{C}_\alpha$   $\Delta\delta$  values are positive. The opposite is known for  $\beta$ -sheet structures with  $\alpha^1\text{H}$  and N-H displaying downfield resonances ( $^1\text{H}_\alpha$  shifts in the order of 0.37 ppm,  $\Delta\delta$  values are positive) and  $^{13}\text{C}_\alpha$  revealing upfield resonances ( $\Delta\delta$  values are negative). When these  $\Delta\delta$  values are plotted against the amino acid sequence, areas of secondary structure can be revealed [142], [144], [145].

Resonance shift values that remain unchanged or exhibit minor deviations from the random coil values are regions found to be flexible or unstructured [142]. Local electronic effects can influence the resonance shift and ultimately  $^1\text{H}_\alpha$  and  $^{13}\text{C}_\alpha$   $\Delta\delta$  are averaged over a window of 2 residues. This smooths out any local variance to enable a clearer interpretation of secondary shifts [146]. The exception to these guidelines for secondary shifts is amphipathic  $\alpha$ -helices. The hydrophobic face of the helix has been found to have NH  $\Delta\delta > 0$ , and the hydrophilic face has NH  $\Delta\delta < 0.42$ , with this periodic variation seen over 3 to 4 residues [145].

The NOESY spectrum of a polypeptide chain is also able to give a good indication of secondary structure through observation of NOE connectivities. Typically, protons within the same amino acid and of adjacent residues will display NOE connectivities, however, some medium range inter-proton distances of 2-4 residues apart are characteristically short in regions of regular secondary structure and can therefore be observed in the NOESY spectrum [138], [140]. Figure 2.6 shows the NOE patterns observed in the spectrum that can be indicative of regular secondary structure [138], [140], [147].

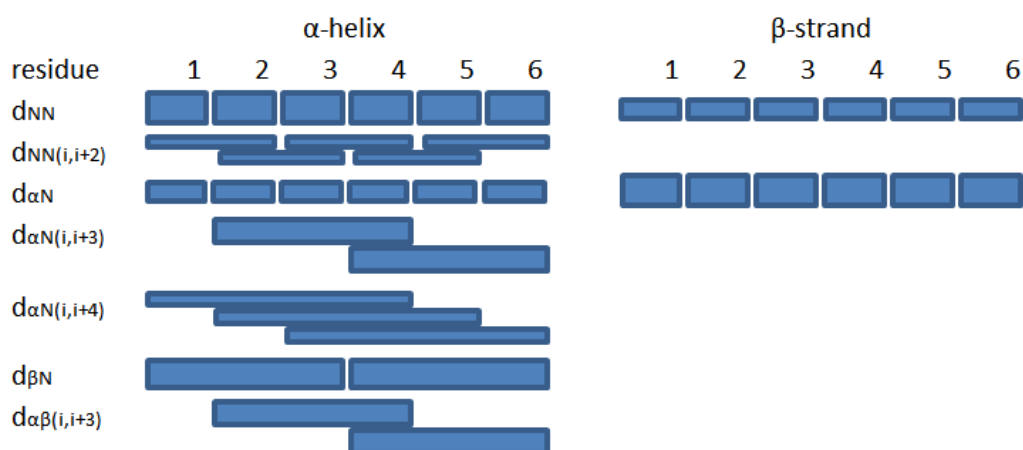


Figure 2.6: Standard pattern of NOE signals for ideal  $\alpha$ -helices &  $\beta$ -strands. The thickness of the bands indicates the relative intensity of the NOE signal. Adapted from [148].

The isotropic coupling between nuclei separated by a small number of covalent bonds can also be used to determine regions of regular secondary structure. The peptide backbone dihedral angles  $\varphi$  and  $\psi$  are defined by the C and N atoms shown in Figure 2.7 and the secondary structures of polypeptides have well-defined backbone dihedral angles [126]. The size of the vicinal coupling constant of  $^3J_{\text{NH}\alpha\text{H}}$  is related to the  $\varphi$  angle by the Karplus relationship shown in Equation 2.6:

$$^3J_{\text{NH}\alpha\text{H}} = 6.4\cos^2(\varphi - 60^\circ) - 1.4\cos(\varphi - 60^\circ) + 1.9 \quad \text{Equation 2.6}$$

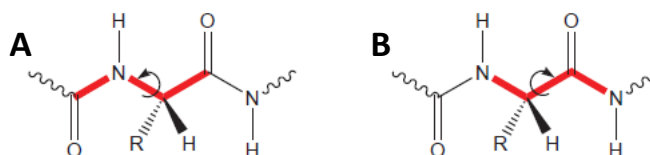


Figure 2.7: The (A) phi ( $\varphi$ ) & (B) psi ( $\psi$ ) backbone dihedral angles (shown in red).

The relationship between dihedral angles and secondary structure ultimately means that the magnitude of  $^3J_{\text{NH}\alpha\text{H}}$  can be correlated with the presence of local structural elements [126].  $\alpha$ -helices exhibit  $^3J_{\text{NH}\alpha\text{H}}$  values of less than 6 Hz,  $\beta$ -sheet structures exhibit values greater than 8 Hz and the values between 6 - 8 Hz are affiliated with random structures [139]. It is noted that the presence of Pro residues can perturb these values. Although coupling constants can easily be measured from high-resolution 1D  $^1\text{H}$ -NMR spectra, when the  $^3J_{\text{NH}\alpha\text{H}}$  value is less than the signal's line width, the calculation is unreliable [126], [139]. Therefore, for large peptides or proteins, the  $^3J_{\text{NH}\alpha\text{H}}$  value is

obtained by the separation between the peak and trough of the antiphase multiplets of the COSY spectra [149].

### 2.2.3: CD Spectroscopy

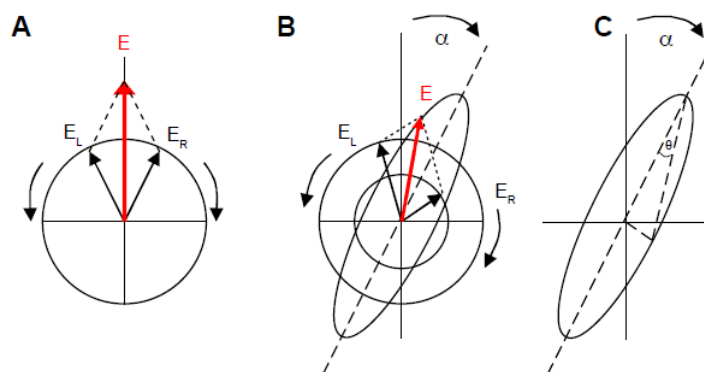
Since its discovery in the mid-19<sup>th</sup> century, circular dichroism (CD) spectroscopy has been extensively used to determine the chirality of optically active molecules [112]. CD spectroscopy has become an efficient rapid low resolution technique to study protein structure, interactions, folding and thermostability through the analysis of the chromophores exhibited [150]–[154]. Although CD is a low resolution technique there are several advantages to this method. Results are acquired fast and efficiently and samples may be measured in physiological conditions. Unfortunately, specific residues cannot be resolved and consequently CD spectroscopy is typically used in conjunction with other techniques to elucidate protein structure [151].

The setup for a CD experiment consists of a beam of light that passes through prisms and filters to become circularly polarised [152]. The resultant light has two vectors of equal length that rotate in opposite directions and are 90 degrees out of phase to each other. The polarised waves have a physical existence with one rotating clockwise (right-handed  $E_R$ ) and one anticlockwise (left-handed  $E_L$ ) and the resultant sinusoidal wave of each can be visualised as a moving spring [112]. Optically active molecules that interact with this light will absorb the right or left-handed waves differentially, and the resulting plane polarised light is rotated by an angle ( $\alpha$ ) [112]. The addition of these absorptions result in a vector that traces out an ellipse and the light is therefore known to be elliptically polarised. Circular dichroism is measured in either units of  $\Delta E$ , the difference in absorbance of  $E_R$  and  $E_L$ , or in degrees of ellipticity, defined as the angle whose tangent is the ratio of the minor to the major axis of the ellipse. The molar ellipticity ( $\theta$ ) is measured in millidegrees as given by Equation 2.7:

$$\theta = 32.98\Delta E$$
Equation 2.7

Alternatively,  $\theta$  can be measured directly as the arctangent of the ratio of minor to major axes of the ellipse as seen in Figure 2.8.





**Figure 2.8:** (A) The electric field vector  $E$  of linearly polarised light represented by left ( $E_L$ ) & right ( $E_R$ ) circularly polarised components. (B) An optically active sample absorbs  $E_L$  &  $E_R$  differently & the remaining components result in the electric field vector  $E$ , which traces out an ellipse that has been rotated through an angle  $\alpha$ . (C) The ellipticity  $\theta$  is the arctangent of the ratio of the major & minor axes of the ellipse. Reproduced with permission from [128].

The determination of protein secondary structure typically requires the CD measurement over a range of UV wavelengths. Although near-UV (320-260 nm) and far-UV (240-180 nm) ranges are measurable for optically active samples, the chromophores of protein chains have characteristic molar ellipticity in the far-UV region. CD data acquired at each  $\lambda$  is converted to the mean residue ellipticity (MRE;  $\text{deg.cm}^2.\text{dmol}^{-1}$ ) as given by Equation 2.8:

$$\text{MRE}_{\lambda \text{ nm}} = \frac{(100 * \Psi_{\lambda})}{lC} \quad \text{Equation 2.8}$$

where  $\text{MRE}_{\lambda \text{ nm}}$  is the mean residue ellipticity at wavelength  $\lambda$ ,  $\Psi_{\lambda}$  is the ellipticity at wavelength  $\lambda$  in millidegrees,  $l$  is the optical path length (cm) and  $C$  denotes the concentration per residue, calculated by dividing the concentration of protein (g/L) by the mean residue molecular weight.

Protein spectra are dominated by the amide absorbance of the peptide backbone as the optical transitions are shifted or split depending on the secondary structural conformation. A schematic diagram showing the different spectra for standard protein secondary structures is displayed in Figure 2.9. The spectra of  $\alpha$ -helices are found to have a positive maximum at 193 nm and two negative minima at 222 nm and 208 nm.  $\beta$ -sheet structures give rise to a maximum at 198 nm and a minimum at 215 nm, and proteins that are intrinsically disordered will give rise to a minimum at 195 nm [151], [152], [155], [156]. Protein structural changes result in shifts in absorbance by chiral chromophores and therefore, CD spectroscopy can be used to monitor the conformational changes due to temperature, mutations, heat, denaturants or binding interactions as a change in ellipticity [153], [157]. In the far-UV region, loss of  $\beta$ -sheet and  $\alpha$ -helical content is observed at  $\lambda$  of 215 and 222 nm with a smaller change of the MRE, generally caused by a loss of periodic structure in the

environment of the peptide bonds [112], [158]. This technique is also able to deduce the predominant secondary structure of an unknown protein.

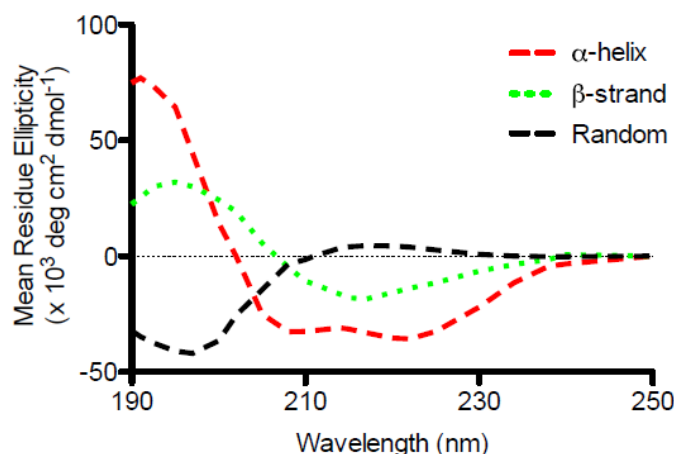


Figure 2.9: Far-UV CD spectra expected for  $\alpha$ -helices (red),  $\beta$ -sheets (green) & random coil secondary structures (black). Reproduced with permission from [128].

Proteins that contain multiple secondary structural elements give spectra that are a combination of these representative spectra seen in Figure 2.9. The convoluted spectrum requires computational deconvolution to be performed to give an estimate of the percentage of each structural component within the protein [159]. DichroWeb is an online open server dedicated to the deconvolution of CD spectroscopy data [160]–[162]. Raw data obtained from CD experiments are uploaded to the server and an algorithm and reference set is selected before deconvolution of the spectrum can begin. The pre-eminent available algorithm; CDSSTR has been found to provide the best estimate of  $\alpha$ -helical content of proteins [152]. This algorithm is categorised as a variable selection program, designed to be flexible so that its parameters can be manipulated to suit data. The deconvolution of raw data is instigated through the creation of secondary structures based on large databases of standard spectra from proteins with known spectra [152], [163]. For CD data with unique curve characteristics dissimilar to the spectra of known proteins, a basis curve is constructed, extracted from the combination of the protein spectra of known characteristics in the reference set. These reconstructed curves will therefore have a defined mixture of secondary structure fractions determined from the proteins in the data set [157]. The fit of the basis curve to the experimental data is calculated through the normalized root mean square deviation (NRMSD) [163] and this value is an indication of the accuracy of the reconstructed curve. Although CDSSTR is time consuming due to the volume of calculations, this method produces the most accurate analysis and it can produce

results where other methods fail to analyse the data [160]. A superior fit of the conformation of globular proteins is seen compared to other algorithms, particularly for estimates of  $\alpha$ -helical content. Unfortunately, it has been seen in other studies to provide inadequate estimates of  $\beta$ -sheets and turns if data are not collected to at least 184 nm [152], [162], [164].

#### 2.2.4: Mass Spectrometry

Although developed in the early 1900's, the use of mass spectrometry (MS) to analyse protein structures has only been implemented primarily over recent decades [165]. The use of MS to identify and characterize biological molecules is now considered a fundamental technique in protein biochemistry and proteomic analysis. Strategies to determine the primary structure of proteins generally involve proteolytic digestion prior to MS analysis. The information provided regarding the individual peptides is then pieced together to reveal the protein identity and/or characteristics (co- and post-translational modifications or isoforms). Secondary, tertiary and quaternary protein structures are elucidated through the use of several MS based approaches. With the advent of new ionisation methods, developments in mass analysers, the pairing of ion mobility (IM) to MS and chemical cross-linking (CXL) to MS, it is now possible to obtain a plethora of structural information enabling the modelling of protein topologies. These approaches are therefore indispensable tools for low resolution structure determination of large biomolecular complexes and are highly regarded in protein structural biology. All traditional mass spectrometers have three main components essential for sample analysis, namely the ionisation source, mass analyser, and detector.

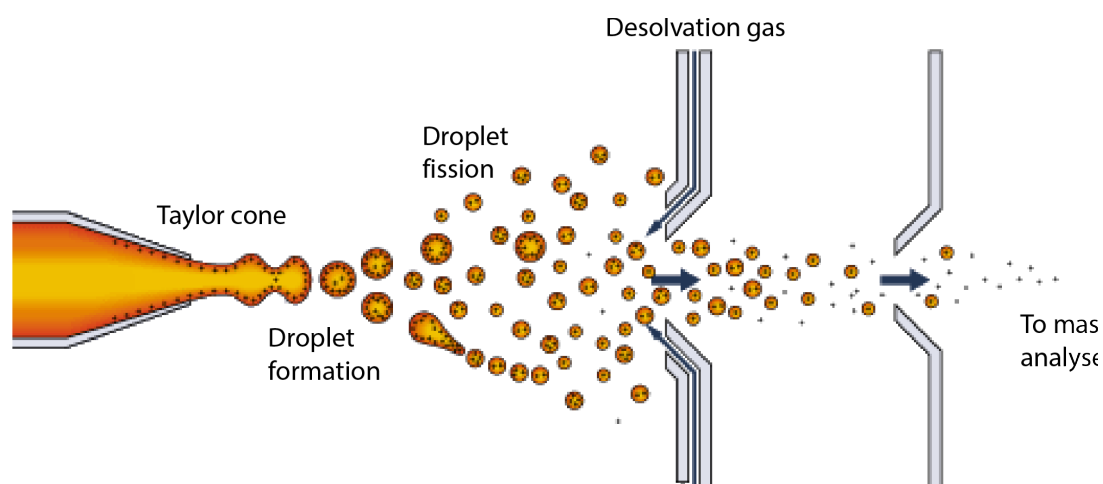
##### 2.2.4.1: Ionisation

Key developments in ionisation have underpinned biological MS. For example, electron impact (EI) [166] and the related chemical ionisation (CI) [167] are traditional methods for ionisation, mainly used for volatile molecules [165]. Field and plasma desorption ionisation were then developed for ionising non-volatile samples. However, the difficult sample preparation and radioactive source required means that they have largely been superseded [168]. A more recently developed method is the popular Matrix-Assisted Laser Desorption Ionisation (MALDI) [169]–[171]. MALDI, similar to fast atom bombardment (FAB), involves the ablation of a pre-prepared mixture of matrix and sample via a laser. The matrix efficiently absorbs the high energy laser light and dissipates, assisting in the ionisation of the embedded analytes [168]. MALDI and electrospray ionisation (ESI) are the 'soft' methods currently dominating the field for ionisation of biomolecules [172]. However, electrospray is the favoured form of ionisation for large native complexes due to the ability to incur multiple charged states, the ability to easily detect and fragment precursor ions and the overall 'softer'

ionisation of the analyte. This technique can occur to give either positive or negative ions and solution phase structures and interactions are largely retained following transition into the gas phase [173]. The techniques used in this thesis include ESI and a modified version of ESI known as nanospray ionisation, used due to the reduced amount of sample needed.

The term 'soft' ionisation was coined by Beuhler *et al.* [174] and is one of the main attractions of ESI. It was based on the idea that the transition of molecules from the liquid phase to vapour form would occur faster than the molecule would decompose [175], consequently allowing the native conformations to be transferred into the gaseous state. ESI was pioneered in the 1960's by Malcolm Doyle and then coupled to MS by John Fenn [168], [174], [176], [177]. This technique allowed the ionisation of native, non-volatile, thermally stable biomolecules and has since been increasingly used in proteomics. The concept was eventually further developed to accommodate small sample amounts, now known as nanospray ionisation. Although both operate by the same principles, ESI requires a syringe pump to obtain a continuous stream of solution at a flow rate between 1-20  $\mu\text{L}\cdot\text{min}^{-1}$  [174], whereas nanospray utilises a flow rate between 10-50  $\text{nL}\cdot\text{min}^{-1}$  [178].

Nanospray ionisation enables the vaporisation of a liquid sample through the dispersal of a fine spray and the transfer of ions into the spectrometer. An electrostatic force is created through a conductive element coated on the walls of the needle. This retains a voltage across the tip of the needle and charges the fluid flowing through the capillary [179]. As the solution passes through the capillary, an elongated droplet is formed at the tip of the needle. This is referred to as the Taylor cone as seen in Figure 2.10. The initial droplet has a large amount of solvent adhered to the analyte and these highly charged droplets have the same polarity as the applied capillary voltage [180]. The droplets are repelled from the charged capillary tip and are attracted towards the mass analyser through a pressure and potential gradient. Nebulising gases (usually  $\text{N}_2$ ) flowing in the opposite direction to the analyte assist in droplet evaporation. Neutral ions and ions of opposing polarity are removed from the system by this process until largely free ions remain.



**Figure 2.10: Schematic representation of nanospray ionisation.** Sample mixed in a volatile solution is placed in a coated needle with a voltage applied across the capillary. The fluid is dispersed from the tip as a fine spray & ionised droplets are formed, migrating towards the mass analyser. Adapted from [181].

Samples are normally dissolved or buffer exchanged into a solvent with a high evaporation index. The most common solutions for proteins in denaturing conditions are organic solvents such as acetonitrile (ACN) and methanol. To preserve proteins in the native state, low concentrations of ammonium acetate (AA) (10-500 mM) is added to the solution [182], [183]. The ammonium cations pair with the acidic functional groups on the surface of the protein and acetate anions pair with basic functional groups [183]. High performance liquid chromatography (HPLC) is often coupled to ESI and MS to increase sensitivity and resolution of complex mixtures [184].

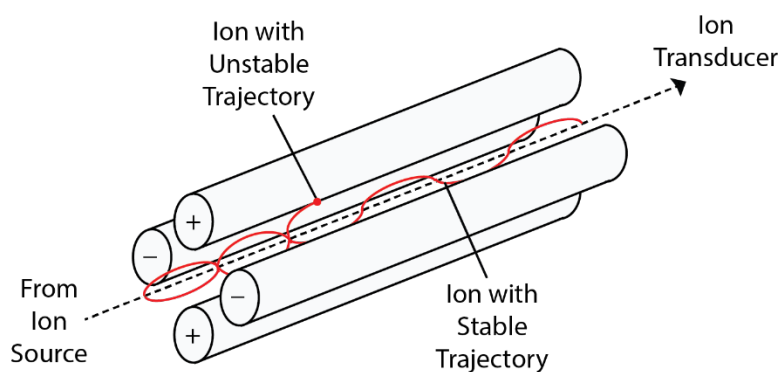
High performance liquid chromatography (HPLC) refers to the separation of proteins, peptides or small compounds by a liquid mobile phase and a solid stationary phase, operating at high pressures. Molecules are separated based on their different chemical properties such as polarity, electrical charge, hydrophobicity or molecular size [185]. This can change depending on the medium and column used. Reverse phase C18 columns are the most commonly used in peptide HPLC systems. These columns have a non-polar stationary phase, and compounds bind to the medium through non-polar and hydrophobic interactions. Typically, a HPLC run will consist of changes in the composition of the polar mobile phase to alter the phase partitioning of each compound. Compounds are then eluted from the column in an order relative to the strength of its interaction between the stationary and mobile phase [186]. In systems where HPLC is directly coupled to MS, the molecules eluting from the column undergo ionisation by ESI and subsequent MS analysis.

### 2.2.4.2: Mass Analysers

Following ionisation, the charged species undergo mass analysis to identify ions based on their mass to charge ratio ( $m/z$ ). Instrumentation made specifically for this purpose include the quadrupole [187], time of flight (ToF) [187], ion trap [188], fourier transform ion cyclotron resonance [189] and orbitrap mass analysers [190] amongst others. The mass analyser systems used in this research are the quadrupole and ToF.

#### 2.2.4.2.1: Quadrupole Mass Analyser

The quadrupole mass analyser was developed in the 1950's by Paul *et al.* to separate ions based on their  $m/z$  ratio [191], [192]. Advantages of the quadrupole include the low cost to manufacture, the compact nature and the ability to scan at a high rate. The quadrupole consists of four metal rods placed parallel and perpendicular to each other. The opposing pair of rods share identical direct current (DC) voltage and radio frequency (RF), with each pair having opposite DC and RF fields 180° out of phase [191], [193]. Ions are accelerated into the rods from the ion source (Figure 2.11) in the  $z$  axis (parallel to the rods) and, as they experience forces in the  $x$  and  $y$  axis from the electric field, the ions oscillate along these axes. Ions with unstable trajectories will oscillate erratically and collide into the rods and are unable to travel the entire length of the quadrupole [193], [194].



**Figure 2.11: Schematic representation of a quadrupole mass analyser. Ions with a stable trajectory are able to travel the full length of the quadrupole, whereas ions with unstable trajectories collide with the metal rods of the quadrupole. Adapted from [195].**

The applied quadrupole field force changes depending on the MS mode [196]. In traditional MS the quadrupole operates as an ion guide and transfers all ions to subsequent sectors of the spectrometer with only an RF voltage application. Alternatively, a range of  $m/z$  ions can be selected for mass analysis and the quadrupole will apply a varied field force to allow only these ions through

the instrument. Lastly, the instrument can operate with a constant electric field to allow a particular  $m/z$  ion of interest to migrate through the quadrupole [197]. This selected ion is passed into the collision cell which has the ability to increase the gas pressure, enabling collisions with the ion. This is known as collision induced dissociation (CID) and the process is referred to as tandem mass spectrometry or MS/MS, as the ions must undergo a subsequent mass analysis [198], [199].

#### 2.2.4.2.2: Time of Flight

In 1946, Stephens described the ToF mass analyser [200] which has since developed to become a fundamental MS analyser, particularly for biomolecules. As the name implies, the ToF sector separates ions based on the time they take to 'fly' from the beginning of the tube to the detector. The tube is normally 1-2 m in length and is a field-free region. The ions are pushed through the tube by a high accelerating potential (V) known as the pusher, and each ion is supposedly provided with an identical translational energy. Flight tube distance (d) and accelerating voltages (V) are fixed and therefore the time (t) taken for the ions to migrate the length of the tube depends only on the mass (m) and charge (z), as given in Equation 2.9. Common accelerating voltages are generally 10-30 kV.

$$t = \left( \frac{m}{2Vz} \right)^{\frac{1}{2}} d \quad \text{Equation 2.9}$$

The time taken for two ions with known mass to charge ratios to reach the detector is measured and the time scale is able to be correlated with the  $m/z$  values. A singly charged ion with the molecular weight of 10,000 Da typically takes approximately 100  $\mu\text{sec}$  to reach the detector [168].

#### 2.2.4.3: Detectors

The final process in MS is the detection of the  $m/z$  separated ions. In the early 1900's the first mass spectrometers used photographic paper onto which line images were visualised. Since then, detectors have developed to include electron multipliers [201], Faraday cup [202] and the non-scanning array detector [203]. This thesis focuses on the microchannel plate detector as it is used in the Waters HDMS Synapt and the Agilent 6560 mass spectrometers.

Microchannel plate (MCP) detectors are able to analyse ions through the secondary electrons emitted from the detector's surface. A plate comprised of an array of micro channels is angled towards the ion beam and each channel on the plate functions as an individual electron multiplier. Ions striking the internal walls of the channels generate secondary ions and these continuously travel through the channel, further striking the interior and subsequently amplifying the current

[204], [205]. This amplified electron signal is sent to the time to digital converter and the arrival time is digitally recorded. The summation of the arrival time over the acquisition period produces a mass spectrum and the ions can now be analysed based on their  $m/z$  ratio.

### 2.2.5: IM-MS

The combination of IM with complementary techniques is useful in elucidating a more detailed structural description of large biomolecular complexes. Specifically, IM coupled to MS enables the measurement of both the mobility and the mass of the ion, and structural biology has greatly benefitted from the introduction of commercial IM-MS spectrometers [7], [206]–[211]. In the 1990's, 'soft' ionisation sources coupled to IM allowed the mobility of native, large biomolecules to be measured, from which tertiary and quaternary structures could be proposed [212], [213]. This is especially useful for native samples that are unable to be analysed through other spectroscopic techniques. As a consequence, the orthogonal separation of IM has therefore become an integral aspect in structural proteomics.

#### 2.2.5.1: Drift Tube Ion Mobility

Traditional IM, also known as drift tube IM (DTIM), involves the separation of ions based on their charge and size, and the time it takes for the ions to migrate through the drift tube is converted to a collisional cross section ( $\Omega$  or CCS) [214].

In DTIM, an electrical field is applied across the drift tube and the tube is filled with a stationary buffer gas [215]. The size and charge of the ion influences the time it takes to travel the length of the tube and subsequently, these factors determine mobility and CCS [215]. Ions with a higher charge state experience a greater force from the constant electrical field and will therefore move faster through the buffer gas than singly charged ions. Concurrently, ions that are topologically bigger will collide more frequently with the stationary gas due to their larger surface areas, retarding the ions' movement. The time it takes for ions to drift the length of the tube is measured as the arrival time distribution (ATD) [214], [216], [217]. The ion mobility ( $K$ ) of a molecule is calculated by the drift velocities ( $V_d$ ) in the presence of an applied electric field ( $E$ ), shown in Equation 2.10.

$$K = \frac{V_d}{E} \quad \text{Equation 2.10}$$

The ion mobility ( $K$ ) is converted to CCS through measurement of the time required for the ions to drift through the distance of the drift tube ( $t_D$ ). A direct relationship exists between  $t_D$  and  $\Omega$  by integration of the Mason-Schamp equation as given in Equation 2.11 [218].



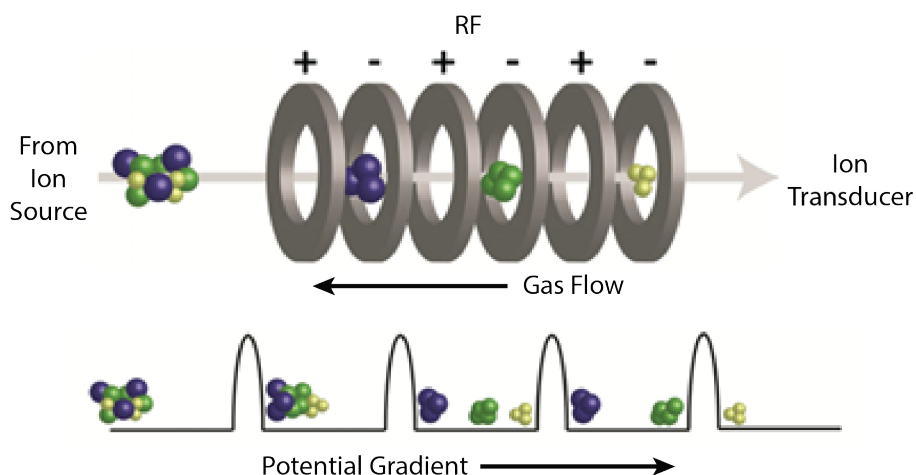
$$\Omega = \frac{(18\pi)^{\frac{1}{2}}}{16} \frac{q}{(k_b T)^{\frac{1}{2}}} \left[ \frac{1}{m_i} \frac{1}{m_b} \right]^{\frac{1}{2}} t_D \frac{E}{L} \frac{760}{P} \frac{T}{273.2} \frac{1}{N}$$

In order to calculate  $\Omega$ , the charge ( $q$ ) of the ion, the absolute gas temperature ( $T$ ), the buffer gas pressure ( $P$ ), the length ( $L$ ) of the drift tube, the Boltzmann constant ( $k_b$ ) and the mass of the ion ( $m_i$ ) and buffer gas ( $m_b$ ) are required. This direct relationship allows experimentally determined drift time data to be converted to CCS, indicating the topology and size of the ions [212], [218]–[220]. Automated algorithms to calculate  $\Omega$  using the ATD derived from DTIM are available online and algorithms have further developed to predict the ion mobilities of peptides [220], [221].

In addition to traditional DTIM, the most commonly used types of IM for protein analysis are Travelling Wave Ion Mobility (TWIM) and Field-Asymmetric Ion Mobility Spectrometry (FAIMS) also known as Differential Mobility (DM). FAIM separates ions through a combination of gas flows and non-linear electrical fields [222], whereas TWIM as employed by the Waters Synapt instrument separates ions using a potential traveling ‘wave’ method [206].

#### 2.2.5.2: Travelling Wave Ion Mobility

TWIM uses a non-linear technique based on a potential electrical wave to determine the mobility of ions. A TWIM sector is comprised of stacked planar electrode rings constructed parallel and equally spaced. Alternating phases of a RF voltage are applied to adjacent electrodes to generate a continuous potential wave enabling changes in the ion’s trajectory as it ‘surfs’ the wave towards the detector, as seen in Figure 2.12. Simultaneously, an inert buffer gas is applied in the opposite direction to the propagating waves, colliding with the advancing ions [114], [206], [223]. The concept of TWIM is similar to that of DTIM. Small and multiply charged ions collide less frequently with the advancing buffer gas and move through the medium quickly. The drift time for these molecules will therefore be smaller than ions that are larger and less charged [213], [216], [218], [224].



**Figure 2.12: Schematic representation of Travelling Wave Ion Mobility (TWIM).** Stacked ring electrodes with alternating radio frequencies that pulse out of phase to the adjacent electrode create a potential gradient for ions to ‘surf’ on. An inert buffer gas flows in the opposite direction. Adapted from [223].

The data obtained from TWIM have non-linear functions and therefore must be calibrated to known mobility values of resolved ions extracted from empirical multipoint calibrations, generally measured by DTIMS [225]–[227]. Measurements for unknown ions diverging significantly from the calibrated range increase the standard deviation uncertainty and can be associated with incorrect CCS calculations. Many computational approaches to calculate CCS have been optimised to minimise these uncertainties [227].

### 2.2.5.3: Computational Approaches to CCS Calculations

The ability to correlate experimental IM data with theoretical CCS calculations of conformational candidate structures from modelling provides a powerful analytical and structural tool for the characterisation of biomolecules [228]. CCS calculations are often determined from molecular structures obtained from the Protein Data Bank (PDB) or theoretically derived models, using software such as the Leeds method [229], [230] or MOBCAL [224]. Although MOBCAL is able to incorporate three traditional algorithms; the exact hard sphere scattering (EHSS) model, the hard sphere projection approximation (PA) and the trajectory method, it requires long computational times to run and the use of a super computer is generally needed, particularly for the trajectory method. In comparison, the Leeds method incorporates the PA model and projected CCSs are able to be generated quickly.

In the PA model, the theoretical CCS is calculated by replicating and modelling the ion collisions with the buffer gas *in silico* by replacing the cross section with a selection of the average projection of the

ion's possible orientations [231]. The average projection assumes that spherical ions possess a uniform charge distribution, therefore dismissing the gas scattering and the long-range interactions that would occur in real situations between the ions and the buffer gas, therefore potentially underestimating the experimental CCS [232]. PA is therefore a good initial indicator of CSS calculated *in silico*. The EHHS [231], [233], trajectory method [218] and some recently suggested algorithms [234]–[237] aim to minimise this underestimation, however these algorithms are not time efficient and high levels of computing performance is required.

#### 2.2.5.4: Ion Mobility-Mass Spectrometers

The combination of IM and MS (IM-MS) has advanced low resolution structure determination of large biomolecular complexes [208], [209], [223], [226], [238]–[240]. This section will detail the IM-MS instrumentation used in this thesis, namely the Waters Synapt HDMS and the Agilent 6560 Ion Mobility Q-ToF. As seen in Figure 2.13, the Waters Synapt HDMS utilises a nanospray ionisation source to ionise sample molecules, a quadrupole to filter ions, the TWIM to separate ions based on their size to charge ratio, a ToF sector to determine the  $m/z$  and a MCP detector to digitally convert the ion signal. The sample ions are therefore analysed by size to charge, mass to charge and given quantitatively. Also, increasing the voltages in the trap and transfer region of the T-wave can enable fragmentation of the molecule either before or after IM analysis, respectively [241], [242]. It is noted that the nanospray ionisation is easily substituted with a Waters Acquity LC system coupled to electrospray ionisation which is also used in this thesis.

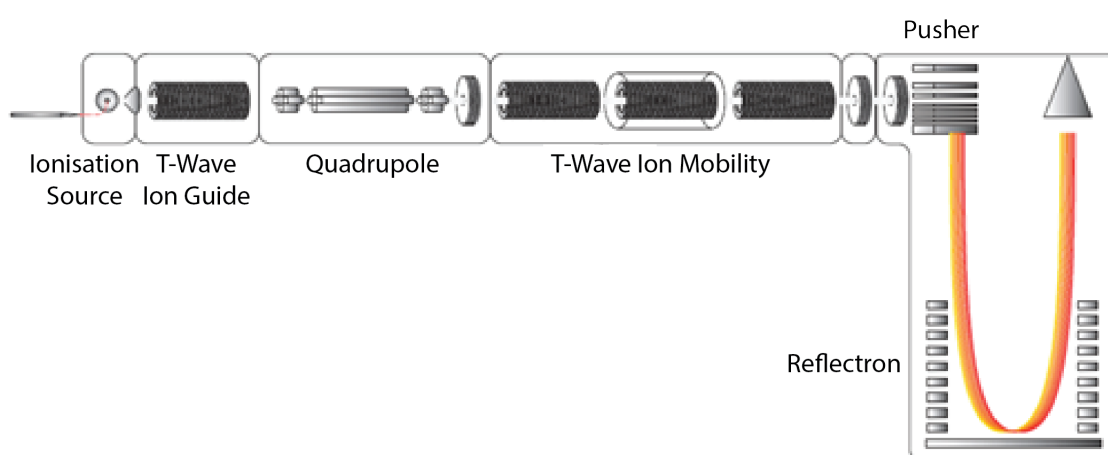


Figure 2.13: Schematic representation of the Waters Synapt HDMS Mass Spectrometer. Adapted from [114].

In comparison, the Agilent 6560 IM-Q-ToF incorporates the same main components as the Waters Synapt, but has a different configuration of sectors, as seen in Figure 2.14. Here, the ionisation source employed to ionise the sample molecules is electrospray ionisation. Similar to the Waters Synapt, the Agilent 6560 can be coupled to the Agilent Infinity BioHPLC system. The instrument then has a trapping funnel and trapping gate to hold the packet of ions before IM analysis. A linear drift tube is employed for IM separation and then ions are transferred to the rear funnel for refocussing. The quadrupole mass analyser then filters ions and allows selected ions into the collision cell. Collision induced dissociation is optional at this stage and ions are pushed through to the ToF sector with a reflectron. In the same fashion as the Waters Synapt, the ions are then received by the MCP detector and converted into an electronic form [243], [244].

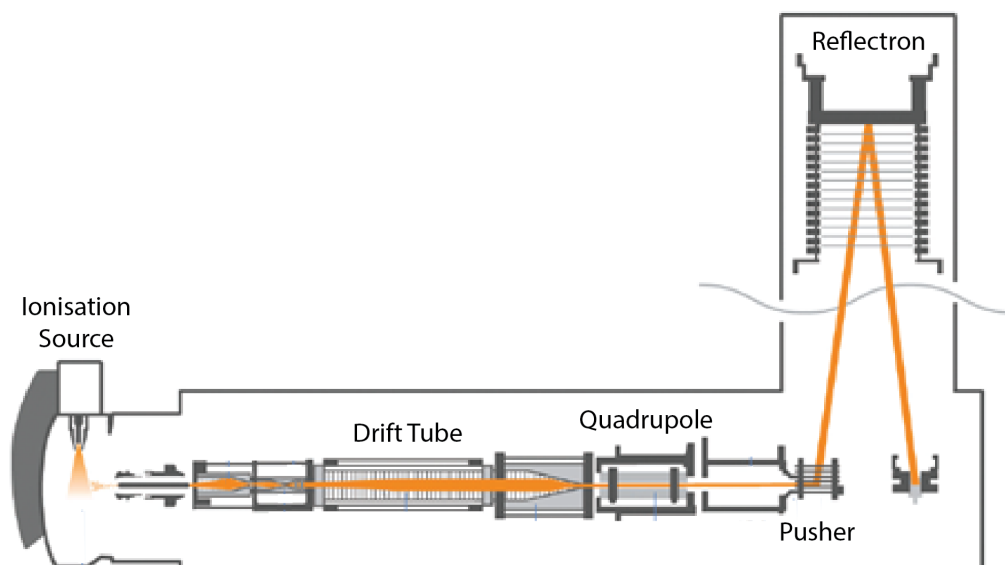


Figure 2.14: Schematic representation of the Agilent 6560 IM-Q-ToF instrument. Adapted from [244].

#### 2.2.5.5: Tandem MS Analysis of Peptides

The activation technique available in the Waters Synapt and Agilent 6560, and commonly used to induce fragmentation of molecular ions, is collision induced dissociation (CID). CID is effected by increasing voltages in the collision cell to kinetically charge the ions which collide with a neutral gas (normally helium or argon). This kinetic energy is transferred and partially converted into internal energy, enabling the breaking of bonds within the molecule [245]–[247]. Bond breakage and ion degradation can provide structural information; in particular, peptides have characteristic fragmentation patterns that allow the primary structure of a protein or peptide chain to be determined as seen in Figure 2.15 [198], [248]. CID is able to fragment peptides principally between the carbonyl and nitrogen of the amino acid. The fragment ions correlating to the N-terminus fragment are referred to as b ions and the fragment ions correlating to the C-terminus are referred to as y ions. Other activation techniques such as electron capture dissociation (ECD) and electron transfer dissociation (ETD) are able to cleave peptides between the nitrogen and  $\alpha$ -carbon, therefore generating c and z type fragment ions and 193 nm ultraviolet photo dissociation (UVPD) is able to fragment peptides at the  $\alpha$ -carbon and carbonyl bond to observe a and x ions in addition to b, y, c, and z ions [249] (Figure 2.15).

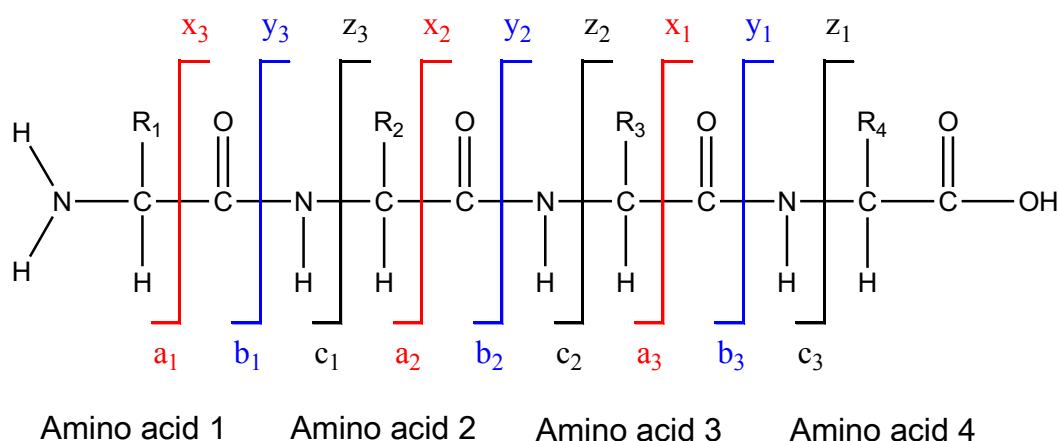


Figure 2.15: Representation of the fragmentation pattern observed for a 4 amino acid long sequence.

Mass spectrometers with quadrupole mass analysers are able to select specific precursor ions for fragmentation and this process is called tandem MS or MS/MS. Automated systems of LC-MS/MS enable high throughput analysis of samples through the selection and fragmentation of abundant ions eluting from the column. This is the fundamental technique in proteomics. ‘Shotgun’ proteomics, termed by Yates due to the resemblance to shotgun genomic sequencing [250], refers to the indirect measurement and identification of proteins by analysis of the proteolytic peptides derived from proteolysis of intact proteins. Typically, samples containing mixtures of protein are denatured, proteolytically digested and subjected to LC-MS/MS. Comparisons of the tandem mass spectra obtained with theoretical tandem mass spectra generated *in silico* from a protein database allows peptide identification, and therefore protein identification, to be achieved. Occasionally, peptide sequences can be assigned to more than one protein, and in these situations the identified proteins are scored and grouped based on their observed peptides [250]–[252].

As an example, Figure 2.16A shows the MS/MS spectrum of the selected precursor ion 628.8 ( $[M+2H]^{2+}$ ) of FLLADNLYCK, a unique peptide identified from *b*PFD subunit 3. The fragmented ions correspond to either b or y backbone cleavages are labelled accordingly. Figure 2.16B reveals the *in silico* generated ions corresponding to all possible b and y ions of peptide FLLADNLYCK by MASCOT using the database Swiss-Prot [253], [254]. Of the 80 possible fragment ions, 13 are identified in the spectrum and 8 are singly charged y backbone ions. Assignment and identification of this peptide to *b*PFD subunit 3 is with high confidence due to the unique amino acid sequence.

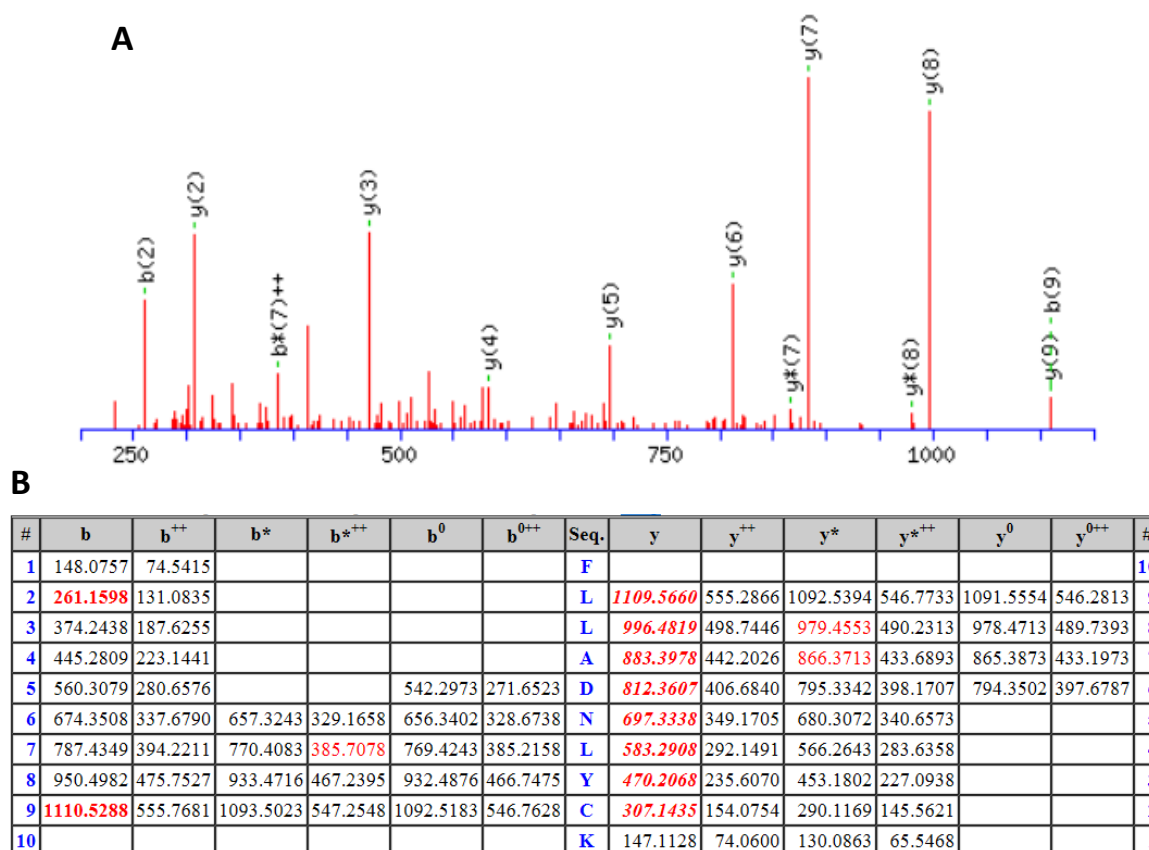


Figure 2.16: (A) MS/MS (CID) of 628.8 ( $[M+2H]^{2+}$ ) precursor cation of FLLADNLYCK (segment from PFD subunit 3). Backbone cleavage processes are indicated according to the b or y ions. (B) A table showing all the possible fragments of the peptide 13/80 fragment ion matches. The observed fragmentations are highlighted in red. Doubly charged ions are indicated by <sup>++</sup>, loss of OH are indicated by \*, & loss of H<sub>2</sub>O are indicated by <sup>0</sup>.

In circumstances of identification of novel peptides, sequencing of unknown peptides through tandem MS is referred to as de novo peptide sequencing. This technique typically constructs sequences through algorithm that assembles individual sequence reads into longer continuous sequences in the absence of a reference sequence. This method enables the construction of novel peptides from data obtained through MS/MS and can also be used to detect any modifications to the amino acid. Modifications are in general calculated from the mass difference of the fragment ions. In particular, the algorithms available for de novo sequencing are also used for the identification of post translational modifications and chemical cross-linking bio-conjugations [198], [248], [255].

Although LC-MS/MS is fundamental to the typical proteomic workflow, the instruments used in this thesis also have an IM sector and are able to induce fragmentation before or after IM separation, depending on the order of placement of the CID cell/s and IM sectors. In the Waters HDMS Synapt, CID can be applied in the trap T-wave cell or in the transfer T-wave cell (Figure 2.13) and therefore,

users have the option of fragmenting the molecular ion before or after IM analysis. In comparison, the Agilent 6560 has a CID cell located between the IM sector and ToF (Figure 2.14) and subsequently, fragmentation of the molecular ions must occur after IM analysis and mass selection. Importantly, inducing CID after IM separation gives the capacity to acquire time aligned fragmentation, a term that has been used to describe the alignment of fragment ions with the precursor in drift time [211], [256].

Time aligned parallel fragmentation is a term used to describe the parallel acquisition of ions and its fragment ions, achieved through CID following IM analysis. The powerful orthogonal separation technique of IM and the parallel match of fragmentation ions to precursor ions allow comprehensive analysis of samples. This technique is a recently developed phenomenon that several recent published papers have exploited to simultaneously separate and identify proteins, isomers, sites of modifications and structural conformations [207], [256]–[258]. These studies were able to obtain time aligned fragmentations through a number of different ways. Typically, sample ions are analysed by IM and a selected constant CID voltage allowed both precursor and fragment ions to be observed was chosen. In this mode, selection of a drift time in the drift plot generates a spectrum with peaks seen of both precursor ions and fragment ions. Alternatively, the mobility-separated precursor ions are subjected to fragmentation at parallel low and high energy cycles for achieving exact mass analyses and fragment ion generation. In this mode, the fragment ions are aligned with their precursors with respect to their drift times. In this thesis, time aligned parallel fragmentation was applied to identify cross-links in protein chemical cross-linking.

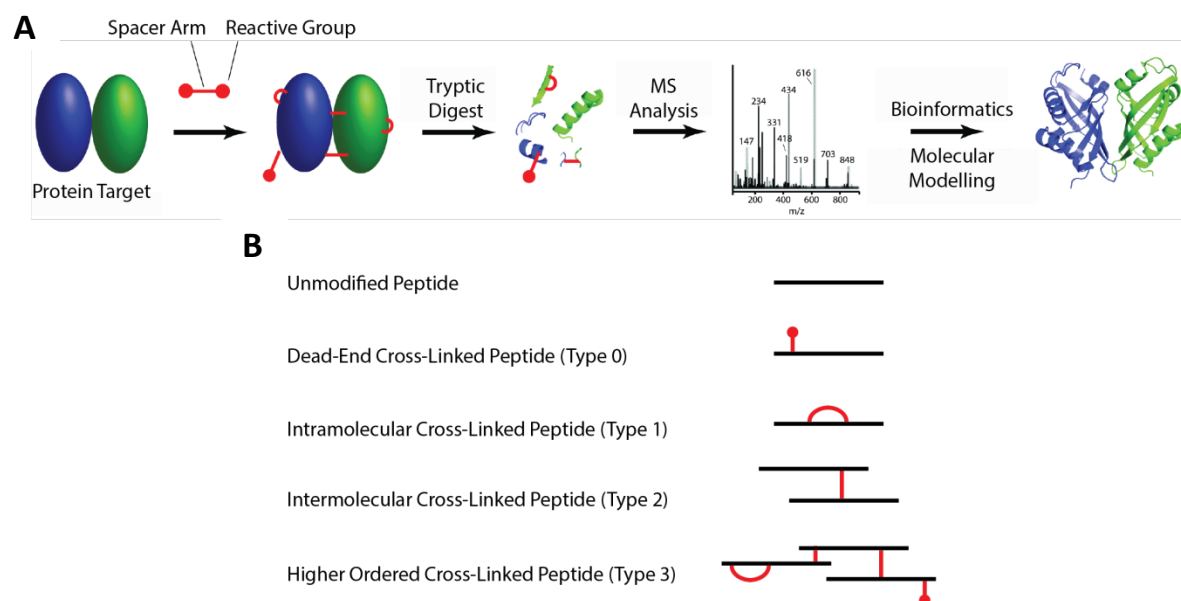
### 2.2.6: Protein Chemical Cross-linking

When transient conformations and protein-protein interactions are at play, chemical cross-linking (CXL) is able to trap these structures so that they may be analysed and subsequently elucidated [259]. Furthermore as CXL reagents have a fixed length, the distance at which intermolecular bio-conjugations occur must therefore be equal to or smaller than the length of the reagent. This is referred to as a distance constraint and hence invaluable structural information is derived from CXL by identification of cross-linking sites.

The CXL workflow often employed is seen in Figure 2.17A. A native protein complex is cross-linked in solution with a CXL reagent, a chemical with two or more reactive groups able to covalently bond with specific residues on the protein. The cross-linked proteins are then enzymatically digested and the peptides are analysed by MS. The data obtained from this analysis is correlated to the protein

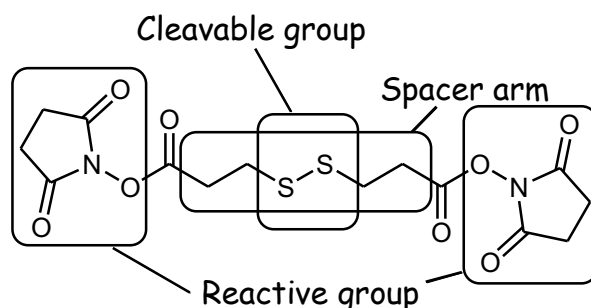


sequence, and cross-link locations are identified by peptide mass and/or characteristic fragmentation patterns [259]–[263].



**Figure 2.17: (A) Schematic representation of the typical bottom up chemical cross-linking mass spectrometry workflow. (B) Products of chemical cross-linking include dead-end cross-links, inter-molecular cross-links & intra-molecular cross-links. Adapted from [259].**

A number of products can arise from chemical cross-linking and these include inter-molecular, intra-molecular, dead-end cross-links and unmodified peptide (Figure 2.17B). To aid in the cross-linking process, the appropriate CXL reagent is chosen based on their chemical specificity, spacer arm length, water-solubility and cell membrane permeability. Figure 2.18 shows the molecular structure of dithiobis (succinimidyl propionate) (DSP), a typical commercially available CXL reagent.



**Figure 2.18: The molecular structure of the common homobifunctional CXL reagent dithiobis (succinimidyl propionate) (DSP). It has 2 NHS ester reactive groups, a spacer arm length of 12 Å & a cleavable disulphide group integrated into the spacer arm.**

Cross-linkers are defined by a spacer arm of pre-determined length, usually a carbon chain, and have two or more reactive groups attached so that they are able to covalently 'link' molecules together or 'fix' transient interactions. Primary amines, carboxyls, sulfhydryls and carbonyls are the 4 major chemical protein functional groups that are able to bio-conjugate and are therefore targets for CXL. Different reactive groups are able to modify these functional groups and covalently bind to them. The most commonly used reactive group is the N-hydroxysuccinimide (NHS) ester due to its quick coupling to amines (principally Lys sidechains which are abundant on the protein surface). Figure 2.19 shows the mechanism of the NHS ester reacting with primary amines.

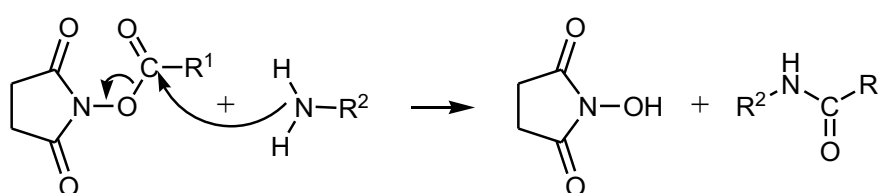


Figure 2.19: Coupling mechanism of the NHS ester. The lone pair of electrons on the amine attacks the carbon on the cross-linker. Amide bond formation results in the bio-conjugation of R<sup>1</sup> & R<sup>2</sup>, where R<sup>1</sup> is generally a spacer arm & R<sup>2</sup> is typically a lysine side chain.

Cross-linkers with identical reactive groups are classified as homobifunctional, and cross-linkers possessing at least two different reactive groups are referred to as heterobifunctional due to their ability to react to different functional groups. This process involves the conjugation of one reactive group to the target protein and, once removing all uncross-linked proteins, the second protein is added so that the second reactive group may be conjugated. This minimises the polymerisation and self-conjugation known to occur in homobifunctional reactions [264]. Although these reagents are good for capturing transient interactions, some reactive groups are able to hydrolyse with the solvent, producing dead end cross-links. However, this gives a good indication of solvent accessibility [265].

The spacer arm of the CXL reagent is also a major component of deliberation as the molecular span and rigidity of a cross-linker will affect the accessibility to potential cross-linking sites [263]. The solubility of the cross-linker is also taken into account, especially for proteins that require certain reaction conditions to maintain their native conformation [264]. CXL reactions are generally performed in near-physiological conditions and occasionally require the cross-linker to be water soluble without affecting the reactivity of the reactive group. The addition of a sulfonate group to NHS esters (-SO<sub>3</sub>) enables this solubility. However, the charge on the sulfonate group inhibits the

permeability of cell membranes and therefore must be reconsidered for intracellular reactions. Other considerations include the ratio of cross-linker to protein for optimum cross-linking reactions to occur. Therefore, the number of functional groups on the protein's surface and the protein concentration is considered.

Even following optimisation of the CXL reagent and reaction conditions to give optimal cross-linking, a plethora of products is obtained from CXL and site specific localisation of linked residues is challenging. Modifications to the CXL reagent to aid in the identification of the cross-linked sites include affinity tags, often based on biotinylation. These added functional groups are able to bind to a specific resin (such as avidin), thus reducing the complexity of the data following purification and enrichment of modified peptides [266]–[270]. More recently, the addition of Staudinger ligation enables selective enrichment of azide tagged cross-linked peptides [267]. Other studies have also incorporated isotopic labelling to allow easy identification of the inter-molecular cross-links [113], [266], fluorescently labelled reagents [271], and mass tag-labelled reagents [263], [272]. However, the addition of tags and labels are expensive and time consuming and purifications of the cross-linked peptides do not remove dead end cross-links. Another common modification to cross-linkers is the incorporation of a cleavage site in the spacer arm between the reactive groups [263], [273]–[276]. Cleavable spacer arms facilitate the easy identification of the cross-linked sites through their recognisable fragmentation pattern following CID.

In this work, IM-MS analysis of the cleavable cross-linking reactions enables an extra orthogonal dimension of separation. Cross-linked peptides are separated based on their size and charge in IM. The precursor ions undergo CID and the cleavable bond within the spacer arm breaks, separating the covalently linked peptides. These product ions are analysed by  $m/z$  and the resulting spectrum/spectra reveal time aligned parallel fragmentations as described in more detail in Section 2.2.5.5. The proposed advantage of IM-MS analysis and in particular time aligned fragmentation, is the ability to simplify the spectra and filter out unwanted ions, enabling easy match and identification of the bio-conjugation that has occurred in the CXL experiment. Selection of a drift time from the drift plot will generate spectrum/spectra of ions with the same mobility. Although overlaps of ions with similar mobilities can occur and consequently be observed in the simplified spectra, this is likely to be minimal. Clear parallel correlations between the ions of the cross-linked peptide to its unique fragment ions will undoubtedly permit easy identification. Time aligned fragmentation is especially useful for the separation and determination of cross-links with identical masses, as the powerful separation of IM distinguishes isomers by conformation and parallel matching of the fragment ions to each precursor ion enables cross-linker identification. Overall, this

unprecedented phenomenon enables the validation of the cross-linked peptides, identification of inter/intra molecular products from dead end products, and separation and determination of structural isomers.

# Chapter 3:

## Mammalian Prefoldin,

### Human Prefoldin Subunits 5 and 6

### and Their Chaperone Activity

---

#### 3.1: Introduction

Eukaryotic PFD and its chaperoning capabilities were first demonstrated by Cowan *et. al* in 1998 [79]. As introduced in Chapter 1, more recent studies have focused on PFD's chaperoning abilities to prevent misfolding and aggregation of peptides involved in pathological diseases. These studies detail PFD as having a passive role in the chaperone mechanism by preventing aggregation, rather than actively enabling the refolding of misfolding peptides. PFD has been found to inhibit the formation of amyloid fibrils, specifically those of amyloid  $\beta$  ( $A\beta$ ) peptides associated with Alzheimer's disease [81], [82].

There have been extensive studies on the peptides  $A\beta(1-42)$  and  $A\beta(1-40)$  due to their implication in the aetiology of Alzheimer's disease [51], [56], [277]. The pathology of Alzheimer's disease is characterised by senile plaques and neurofibrillary tangles present in the brain, resulting in the loss of neurons and synapses [278]. Although Alzheimer's disease was first described in 1906, it was not until 1984 that advanced research and instrumental development allowed George Glenner and Caine Wong to implicate  $A\beta$  aggregation as one of the underlying causes of the neurodegenerative disease [279]. High levels of  $A\beta(1-42)$  and  $A\beta(1-40)$  in the cerebrospinal fluid found in patients with Alzheimer's disease linked the misfolding of the peptides to the formation of fibrils and consequently plaques, a suggested indicator of Alzheimer's disease [280]. These findings prompted research aimed at the development of potential therapeutics with the ability to inhibit or dissociate the fibrils formed from  $A\beta$ .

The strategies that are typically employed to inhibit  $A\beta$  aggregation aim to disrupt the amyloid aggregation pathway [281]. Small planar aromatic compounds such as congo red (CR), chrysamine G (CG) and curcumin have shown promising results in counteracting the formation of fibrils, and have

the potential to be good protein misfolding inhibitors [281]–[283]. Derivatives of these compounds have been synthesised as a bifunctional compound so that one end binds to the fibrillar core region of A $\beta$  through hydrophobic interactions and the other end incorporates a disrupting or large steric element to hinder A $\beta$  aggregation. The presence of excess metals has also been suggested to contribute to the formation of fibrils, and therefore, chelating molecules have also been investigated in the search for a potential therapeutic [283]. Other molecules developed include inhibitors of beta-secretase 1 (BACE1), the enzyme known to cleave amyloid precursor protein (APP) to form A $\beta$  (1-42) and A $\beta$  (1-40) [284]. Although many of these small molecules have been inducted into clinical trials, they have shown many adverse side effects including reduced nerve myelination, neurodegeneration, altered glucose homeostasis, and hepatotoxicity [282], [284], [285]. Therefore attention has been directed to larger molecules with the ability to mimic biological mechanisms for protein homeostasis.

Development of the antibody aducanumab enabled the selective interaction with A $\beta$  aggregates, including soluble oligomers and insoluble fibrils. After a year of receiving regular monthly treatments of the antibody, a significantly reduced amount of A $\beta$  plaque was recorded in test subjects. Unfortunately, the results also included an increased risk of brain haemorrhage and potential dangerous fluid shifts in the brain [286]. These side effects are due to the unprecedented introduction of a synthetically developed therapeutic. It is for this reason that the investigation of endogenous and ubiquitous proteins to combat the A $\beta$  fibril forming process is commendable. Several studies have investigated the addition of ubiquitous molecular chaperones to the fibril forming peptide with promising results [287]–[291]. This chapter focuses on the chaperoning abilities of eukaryotic PFD and its subunits to the amyloidogenic A $\beta$  peptide found in Alzheimer's disease.

Although there is extensive structural and biological data available for several homologues of PFD, only a few scientific papers have focussed on eukaryotic PFD. Some studies have suggested different binding and biological capacities between homologues of PFD despite having regions of conserved sequences [91], [95]. This may signify that large sequence deviations exist between species of PFD and conclusions based on homologues of PFD do not necessarily extend to eukaryotic PFD.

### 3.1.1: Aims

The work described in this chapter aims to compare PFD sequences from different species using a global alignment, to isolate mammalian PFD (*bPFD*) from bovine testis using a method adapted from Cowan *et al.* [292] and to synthesise two subunits of human PFD (*hPFD* subunit 5 and *hPFD* subunit

6) using recombinant protein expression. The isolated *bPFD*, *hPFD* subunit 5 and *hPFD* subunit 6 form the subject of a series of analytical measurements to determine the biological activity of the hexameric chaperone and its selected subunits against the prefibrillar peptide A $\beta$ (1-40), specifically to determine the inhibitory effects of prefoldin against A $\beta$  aggregation and the biological effect of the resulting products formed from the chaperoning activity.

## 3.2: Results and Discussion

### 3.2.1: Sequence Comparison of PFD Homologues

Although alignment of PFD homologues from different organisms using Multiple Sequence Alignment (MSA) has previously been explored [93], [98], the output for these analyses are normally only insightful for evolutionary relationships and binding similarities, and are not generally used to conclude tertiary or quaternary structural information [293]–[295]. However, due to the limited structural data available for eukaryotic *human* PFD (*hPFD*), pre-existing structural models of PFD from other origins have been used to suggest the structure of *hPFD*. A pairwise global alignment will allow comparison of the PFD sequences to ascertain whether or not utilising different species of PFD as a guideline for structural models is creditable. To determine the similarities, PFD subunits from different species were individually aligned to the corresponding *hPFD* subunit. Using the global alignment software BLAST, implemented with the global Needleman-Wunsch (NW) algorithm for scoring, available on the National Centre of Biotechnology (NCIB) webpage [296], the species subjected to this comparison were bovine *mammalian* PFD (*bPFD*), yeast PFD (*yPFD*), *methanobacterium thermoautotrophicum* PFD (*mtPFD*), *pyrococcus horikoshii* PFD (*phPFD*) and *methanocaldococcus jannaschi* PFD (*mjPFD*). The sequences of *bPFD* are compared to *hPFD* to determine the feasibility of substituting and correlating *hPFD* data with *bPFD* data obtained. The species *yPFD*, *mtPFD*, *phPFD* and *mjPFD* are chosen due to their available structural data present in the PDB. The sequences of the different PFD species are listed in Appendix B.

The software BLAST employs a heuristic method and begins the process of aligning sequences through the location of a short amino acid residue match between protein sequences. This initial process is termed ‘seeding’ and is followed by the identification of local alignments and similarities in sequences through sets of common residues along the entire sequence [297]. The score depicted in the final output result includes the percentage of identities; the same residue at the same position in the alignment, the percentage positives; the similarity of the residue at the same position in the alignment, and lastly, the percentage of gaps; the gaps needed to align the sequences. Figure 3.1 shows an example of the BLAST global alignment output, comparing *hPFD* subunit 6 to *yPFD* subunit 6. It can be seen that there are extensive sections of conserved amino acid residues with 42 % of amino acids matching (identities) and, in total, 53 % matching and similar amino acids (positives). This suggests a shared ancestry and homology between these subunits. The NW score for this alignment is 188. The alignment scoring scheme for NW is residue based and it consists of residue substitution scores, where +1 is given for a match and -1 is the penalty for a mismatch. The



magnitude of the NW alignment score is governed by the number of amino acid residues in the overall sequence. This ultimately means that, the longer the sequences are, the higher the possible NW score. Subsequently, the scores are generally compared to the NW score of randomised sets of amino acids of the same length to determine the significance of the alignment [298]. For this particular comparison, a BLAST global alignment of the sequence itself and sets of computer generated randomised sequences will be compared. The NW alignment scores of PFD that were found to be higher than the score of the averaged randomised sequences are accepted to have global sequence similarity. Scores that are close to the NW global alignment of the sequence itself, have a higher degree of similarity.

NW Score		Identities	Positives	Gaps
188		54/129(42%)	69/129(53%)	15/129(11%)
Query	1	MAELIQKKLQGEVEKYQQLQKDLSSMSGRKLEAQLTENNIVKEELALLDGSNNVFKLL		60
		M+EL KYQQLQ +L + + RQKLE QL EN IV EE L+ V+KL		
Sbjct	1	MSEL-----GAKYQQLQNELEEFIVARQKLETQLQENKIVNEEFDQLEEDTPVYKLT		52
Query	61	GPVLVKQELGEARATVGKRLDYITAEIKRYESQLRDLERQSEQQRETLAQLQQEFQRAQA		120
		G VL+ E EAR V KRL++I EI R E +RD + + E+ R L +L A		
Sbjct	53	GNVLLPVEQSEARTNVDKRLEFIETETRCENIRDKQEELEKMRSELIKLNNT-----A		107
Query	121	AKAGAPGKA	129	
		A G PG+		
Sbjct	108	ASTG-PGR	114	

Figure 3.1: An excerpt of the BLAST output for comparison of *h*PFD subunit 6 (query) & *y*PFD subunit 6 (subject). The percentages of mismatched amino acids are not computed here, although it can be manually calculated (129-(69+15)=45 eg. 35 %).

The highest possible score allowed for a sequence comparison to *h*PFD subunit 6 is 624, and the average NW score of the randomised sets is -14 (Table 3.1). This ultimately means that the NW score of *y*PFD subunit 6 compared to *h*PFD subunit 6 of 188 reveals an overall high global sequence similarity. Further comparisons of the PFD subunits conducted are summarised in Table 3.1. As archaeal PFD (*mt*PFD, *ph*PFD) is comprised of 2 copies of an  $\alpha$  subunit and 4 copies of a  $\beta$  subunit to make up the hexameric complex [98], whilst *h*PFD has been characterised as having 6 individual subunits (2 separate  $\alpha$ -like structures and 4 different subunits with  $\beta$ -like structures [79], [91]), in this BLAST global pairwise sequence alignment, the archaeal  $\alpha$  subunits were compared to both *h*PFD subunits 3 and 5 and the archaeal  $\beta$  subunits were compared to *h*PFD subunits 1, 2, 4 and 6. As *mj*PFD has 2  $\alpha$  subunits, both were compared to *h*PFD subunits 3 and 5.

Table 3.1 shows the scores and percentages of the BLAST sequence alignments. The alignment of the *h*PFD subunits to *b*PFD subunits stand out, with high NW scores ranging between 574-943 for each

subunit. The identities and positives are seen to be higher than 94% and gaps are less than 1% for all the subunits. The high degree of sequence similarity for *hPFD* and *bPFD* suggests the inferences of existing *bPFD* studies can be reasonably applied to *hPFD*.

In comparison, the global alignment of *hPFD* to *yPFD* reveal NW scores that range from 67-241, with identities ranging between 22-42 %, positives ranging between 42-62 % and gaps between 3-27 %. It is observed that the alignment of the *hPFD* and *yPFD* species within the  $\alpha$  and  $\beta$  subunit types reveal *hPFD* subunit 5 shows the highest degree of global similarity to the *yPFD*  $\alpha$  subunit, while *hPFD* subunit 4 shows the highest global alignment score for *yPFD*  $\beta$  subunit. The identities between these subunits equate to more than 35 %, with positives over 56 % and gaps of less than 5 %. As all the NW scores for *yPFD* are significantly greater than the averaged random sequences, it is established that *yPFD* has a reasonable degree of sequence similarity to *hPFD* and existing structural models of *yPFD* may be used as a guide to elucidate the quaternary structure of eukaryotic PFD.

**Table 3.1: Global sequence comparison of PFD subunits of different species to the human sequence using BLAST, formulated with the Needleman-Wansch algorithm.**

	NW score	Identities	Positives	Gaps
<b>hPFD subunit 1 (<math>\beta</math>)</b>				
Human	591	122/122 (100%)	122/122 (100%)	0/122 (0%)
Bovine	574	119/122 (98%)	121/122 (99%)	1/122 (0%)
Yeast	67	27/121 (22%)	58/121 (47%)	12/121 (9%)
PH	60	33/125 (26%)	51/125 (40%)	12/125 (9%)
MT	38	26/128 (20%)	60/128 (46%)	14/128 (10%)
MJ	74	30/122 (25%)	61/122 (50%)	10/122 (8%)
Random	6 (average of 5: -7)	26/131 (20%)	45/131 (34%)	20/131 (15%)
<b>hPFD subunit 2 (<math>\beta</math>)</b>				
Human	755	154/154 (100%)	154/154 (100%)	0/154 (0%)
Bovine	747	152/154 (99%)	152/154 (98%)	0/154 (0%)
Yeast	85	41/154 (27%)	65/154 (42%)	43/154 (27%)
PH	-17	20/154 (13%)	59/154 (38%)	37/154 (24%)
MT	16	33/154 (21%)	64/154 (41%)	33/154 (21%)
MJ	29	28/154 (18%)	57/154 (37%)	41/154 (26%)
Random	-33 (average of 5: -17)	27/178 (15%)	57/178 (32%)	36/178 (20%)
<b>hPFD subunit 3 (<math>\alpha</math>)</b>				
Human	1017	197/197 (100%)	197/197 (100%)	0/197 (0%)
Bovine	943	185/197 (94%)	189/197 (95%)	1/197 (0%)
Yeast	267	81/210 (39%)	118/210 (56%)	25/210 (11%)
PH	21	40/200 (20%)	70/200 (35%)	55/200 (27%)
MT	-75	28/197 (14%)	59/197 (29%)	56/197 (28%)
MJ1	-24	27/197 (14%)	69/197 (35%)	55/197 (27%)
MJ2	-8	32/198 (16%)	62/198 (31%)	51/198 (25%)
Random	-29 (average of 5: -49)	42/228 (18%)	77/228 (33%)	59/228 (25%)
<b>hPFD subunit 4 (<math>\beta</math>)</b>				
Human	659	133/133 (100%)	133/133 (100%)	0/133 (0%)
Bovine	642	132/134 (99%)	133/134 (99%)	0/134 (0%)
Yeast	184	47/133 (35%)	75/133 (56%)	4/133 (3%)
PH	-1	19/133 (14%)	53/133 (39%)	16/133 (12%)
MT	33	26/134 (19%)	57/134 (42%)	14/134 (10%)
MJ	42	29/133 (22%)	55/133 (41%)	20/133 (15%)
Random	-13 (average of 5: -13)	29/144 (20%)	55/144 (38%)	22/144 (15%)
<b>hPFD subunit 5 (<math>\alpha</math>)</b>				
Human	765	154/154 (100%)	154/154 (100%)	0/154 (0%)
Bovine	749	150/154 (97%)	152/154 (98%)	0/154 (0%)
Yeast	241	58/163 (36%)	102/163 (62%)	9/163 (5%)
PH	103	37/155 (24%)	77/155 (49%)	8/155 (5%)
MT	67	33/157 (21%)	74/157 (47%)	19/157 (12%)
MJ1	60	39/161 (24%)	73/161 (45%)	26/161 (16%)
MJ2	59	37/161 (23%)	66/161 (40%)	21/161 (13%)
Random	-52 (average of 5: -25)	27/182 (15%)	53/182 (29%)	56/182 (30%)
<b>hPFD subunit 6 (<math>\beta</math>)</b>				
Human	624	129/129 (100%)	129/129 (100%)	0/129 (0%)
Bovine	620	128/129 (99%)	129/129 (100%)	0/129 (0%)
Yeast	188	54/129 (42%)	69/129 (53%)	15/129 (11%)
PH	64	31/129 (24%)	56/129 (43%)	12/129 (9%)
MT	67	30/130 (23%)	59/130 (45%)	9/130 (6%)
MJ	90	34/129 (26%)	66/129 (51%)	16/129 (12%)
Random	-5 (average of 5: -14)	27/153 (18%)	49/153 (32%)	50/153 (32%)

Identities: the same acid residue at the same position in the alignment. Positives: the similarity of the acid residue at the same position in the alignment (including identities). Gaps: insertion or deletion of a residue. Evolutionary substitution matrices are not included in the table. Subunits were also aligned to themselves to show the highest possible NW score and a computer generated random sequence. An average score of 5 sets of randomised sequences is also included.

In contrast, the alignment NW scores of *ph*PFD and *mt*PFD subunits with *h*PFD subunits exhibit low global sequence similarities. The identities and positives percentages are low, however the % of gaps are not as high as the randomised set. Although these results suggest a low correlation between the sequences, only 2 alignments are rejected on the basis of no correlation, (*ph*PFD subunit  $\beta$  to *h*PFD subunit 2 and *mt*PFD  $\alpha$  to *h*PFD subunit 3 have NW scores below the averaged randomised sets). It is interesting to see that *mj*PFD  $\beta$  subunit has a highest NW score of the prokaryotic PFD when compared against *h*PFD subunit 1 whilst having the lowest NW score observed for *h*PFD subunits 3 and 5. The sequence similarities observed between *ph*PFD and *mt*PFD to *h*PFD may be due to regions of conserved function between the PFD species as previous MSA studies have indicated [79]. Structure prediction algorithms strongly suggest that the *h*PFD complex contains secondary structures of a central region of  $\beta$ -strands and several extended hydrophobic coils of highly helical structure content [98]. These core hydrophobic coil residues were simultaneously aligned to the optimum-score obtained from the MSA analysis, revealing shared homology between the different species of PFD. Although these regions give a high degree of amino acid sequence similarity among archaeal and eukaryotic subunits [79], [93], [98], the global BLAST alignment reveals an overall low sequence similarity. This therefore means that existing structural information for *ph*PFD, *mt*PFD and *mj*PFD should be interpreted with caution when aiding in the quaternary complex structure determination of *h*PFD.

Global BLAST NW alignments to compare the sequences of the 6 *h*PFD subunits according to their respective classes were undertaken, as detailed in Table 3.2.

**Table 3.2: Global sequence comparison of *h*PFD subunits ( $\alpha/\beta$ ) using BLAST, formulated with the Needleman-Wansch algorithm.**

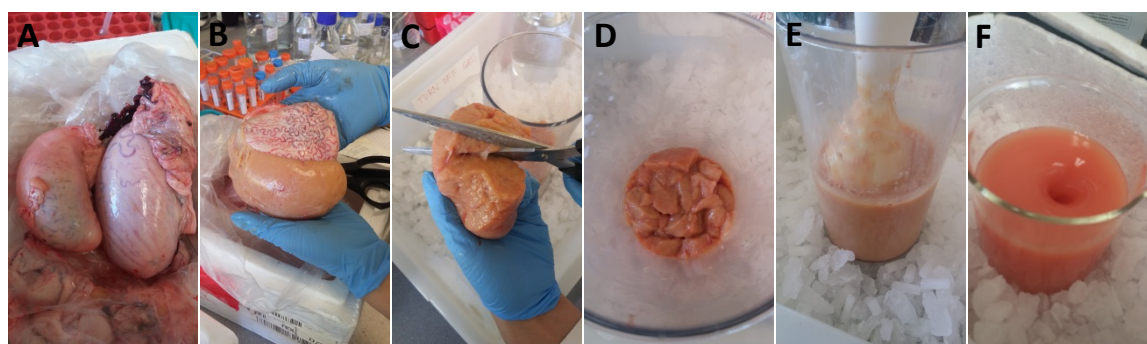
	NW score	Identities	Positives	Gaps
<b><i>h</i>PFD subunit 1 (<math>\beta</math>)</b>				
<i>h</i> PFD subunit 1	591	122/122 (100%)	122/122 (100%)	0/122 (0%)
<i>h</i> PFD subunit 2	-24	32/160 (20%)	54/160 (33%)	39/160 (24%)
<i>h</i> PFD subunit 4	47	32/137 (23%)	60/137 (43%)	20/137 (14%)
<i>h</i> PFD subunit 6	18	32/138 (23%)	55/138 (39%)	16/138 (11%)
Random	6 (average of 5: -7)	26/131 (20%)	45/131 (34%)	20/131 (15%)
<b><i>h</i>PFD subunit 2 (<math>\beta</math>)</b>				
<i>h</i> PFD subunit 2	755	154/154 (100%)	154/154 (100%)	0/154 (0%)
<i>h</i> PFD subunit 1	-24	32/160 (20%)	54/160 (33%)	39/160 (24%)
<i>h</i> PFD subunit 4	5	22/154 (14%)	53/154 (34%)	21/154 (13%)
<i>h</i> PFD subunit 6	35	32/154 (21%)	66/154 (42%)	25/154 (16%)
Random	-33 (average of 5: -17)	27/178 (15%)	57/178 (32%)	36/178 (20%)
<b><i>h</i>PFD subunit 3 (<math>\alpha</math>)</b>				
<i>h</i> PFD subunit3	1017	197/197 (100%)	197/197 (100%)	0/197 (0%)
<i>h</i> PFD subunit 5	-9	31/197 (16%)	69/197 (35%)	43/197 (21%)
Random	-29 (average of 5: -49)	42/228 (18%)	77/228 (33%)	59/228 (25%)
<b><i>h</i>PFD subunit 4 (<math>\beta</math>)</b>				
<i>h</i> PFD subunit 4	659	133/133 (100%)	133/133 (100%)	0/133 (0%)
<i>h</i> PFD subunit 1	47	32/137 (23%)	60/137 (43%)	20/137 (14%)
<i>h</i> PFD subunit 2	5	22/154 (14%)	53/154 (34%)	21/154 (13%)
<i>h</i> PFD subunit 6	20	22/140 (16%)	49/140 (35%)	18/140 (12%)
Random	-13 (average of 5: -13)	29/144 (20%)	55/144 (38%)	22/144 (15%)
<b><i>h</i>PFD subunit 5 (<math>\alpha</math>)</b>				
<i>h</i> PFD subunit 5	765	154/154 (100%)	154/154 (100%)	0/154 (0%)
<i>h</i> PFD subunit 3	-9	31/197 (16%)	69/197 (35%)	43/197 (21%)
Random	-52 (average of 5: -25)	27/182 (15%)	53/182 (29%)	56/182 (30%)
<b><i>h</i>PFD subunit 6 (<math>\beta</math>)</b>				
<i>h</i> PFD subunit 6	624	129/129 (100%)	129/129 (100%)	0/129 (0%)
<i>h</i> PFD subunit 1	18	32/138 (23%)	55/138 (39%)	16/138 (11%)
<i>h</i> PFD subunit 2	35	32/154 (21%)	66/154 (42%)	25/154 (16%)
<i>h</i> PFD subunit 4	20	22/140 (16%)	49/140 (35%)	18/140 (12%)
Random	-5 (average of 5: -14)	27/153 (18%)	49/153 (32%)	50/153 (32%)

Identities: the same acid residue at the same position in the alignment. Positives: the similarity of the acid residue at the same position in the alignment (including identities). Gaps: insertion or deletion of a residue. Evolutionary substitution matrices are not included in the table.

Interestingly, a BLAST global alignment of the individual subunits according to their type, reveals that they have low sequence similarities to each other. The alignment that is rejected as having any degree of sequence similarity is *h*PFD subunit 1 to *h*PFD subunit 2. Therefore, it is proposed that each subunit has evolved independently to assist in the biological chaperone process of the PFD complex. As homology between the species of PFD has been established through MSA [98], it is proposed that over time, the subunits of PFD have specifically evolved to exhibit their own individual functionality. They have evolved to the extent where they no longer resemble each other in terms of sequence. However, it has been proposed in previous studies that structure is more likely to be conserved than sequence in evolutionary homologs [299].

### 3.2.2: Isolation of *b*PFD

In order to further evaluate the structure and activity of *b*PFD, attempts were made to purify the complex from bovine tissue. While a protocol exists for purification of *b*PFD from bovine testes [292], in our hands this did not yield the complex and therefore extensive effort was aimed at optimising purification. Isolation of *b*PFD from approximately 500-600 g of crude bovine testis began with removal of connective tissue and the pulverisation of the remaining fatty tissue as seen in Figure 3.2A-E. Following homogenisation of tissue, protease inhibitors and phenylmethylsulfonyl fluoride (PMSF) were added to inhibit the enzymatic digestion of *b*PFD to allow the maximum amount of pure native protein to be extracted. The supernatant was subjected to a  $(\text{NH}_4)_2\text{SO}_4$  fractionation, so that *b*PFD may be precipitated out of solution. Due to the markedly different solubilities of proteins, this process removes many unwanted proteins present in the solution [300].



**Figure 3.2: Prefoldin extraction from bovine testes. (A) Unprocessed bulls' testicles. (B) Connective tissue removal from sample. (C) Cross section of the tissue. (D) Manageable chunks of tissue. (E) Manual lysis of tissue in homogenisation buffer. (F) Ammonium sulphur fractionation of the extracted supernatant.**

A previously reported purification of *b*PFD found it to precipitate between 35-55 % saturation of  $(\text{NH}_4)_2\text{SO}_4$  [292]. Here the  $(\text{NH}_4)_2\text{SO}_4$  precipitation process was replicated in the ranges 20-40 %, 40-60 % and 60-80 %. To determine which  $(\text{NH}_4)_2\text{SO}_4$  precipitation pellet contained *b*PFD, a sample from each pellet was dissolved, proteolytically digested and subjected to shotgun proteomic data analysis as described in section 2.2.5.5. The MS data was searched using Mascot and the Swiss-Prot database and revealed the  $(\text{NH}_4)_2\text{SO}_4$  20-40 % pellet contained *b*PFD subunits 4 and 6. Pelleted fraction 40-60 %  $(\text{NH}_4)_2\text{SO}_4$  had ions corresponding to *b*PFD subunits 2-6. The ions found to correlate to *b*PFD subunits are summarised Table 3.3.

**Table 3.3: Ions identified by proteomic analysis correlating to *b*PFD subunits in the ammonium sulphate precipitation fractions. *b*PFD subunits observed are ordered by decreasing ion scores.**

	Observed <i>m/z</i>	Mass expected	Mass calculated	Error (ppm)	Miss	Sequence	Ion Score
<b>20-40 % Ammonium Sulphate Pellet</b>							
<i>b</i> PFD subunit 4	832.8930	1663.7715	1663.7689	2	0	K.AAAEDVNVTFEDQQK.I	42
<i>b</i> PFD subunit 6	412.8816	1235.6229	1235.6292	-5	2	K.DLSKSMGRQK.L	1
	629.8411	1257.667	1257.6677	-0	1	K.QELGEARATVGK.R	5
	472.2679	1413.7818	1413.7688	9	2	K.QELGEARATVGKR.R	6
	611.3469	1220.6793	1220.6765	2	1	K.RLDYITAEIK.R	24
<b>40-60 % Ammonium Sulphate Pellet</b>							
<i>b</i> PFD subunit 4	832.8930	1663.7715	1663.7689	4	0	K.AAAEDVNVTFEDQQK.I	38
	708.8555	1415.6964	1415.6892	5	0	K.NLQEEIDALESR.V	51
<i>b</i> PFD subunit 5	550.6351	1648.8836	1648.8746	5	0	K.ELLVPLTSSMYVPGK.L (oxidation M)	3
	547.2852	2185.1115	2185.1055	3	0	K.LHDVEHVLIDVGTGYVEK.T	1
	729.3782	2185.1129	2185.1055	3	0	K.LHDVEHVLIDVGTGYVEK.T	49
<i>b</i> PFD subunit 3	789.8892	1577.7638	1577.7573	4	0	K.NLDSLEEDLDFLR.D	49
<i>b</i> PFD subunit 2	719.0395	2154.0966	2154.0878	4	0	K.AAELEMELNEHSLVIDTLK.E	35
<i>b</i> PFD subunit 6	892.9594	1783.9043	1783.8887	9	2	K.YQQLQKDLKSMGR.Q	1
	543.3072	1626.8997	1626.8940	4	1	R.QKLEAQLTENNIVK.E	1
	686.3803	1370.7461	1370.7405	4	0	K.LEAQLTENNIVK.E	8
	745.8867	1489.7588	1489.7525	4	0	R.ETLAQLQQEFQR.A	32
	814.4268	2440.2586	2440.2822	-10	2	R.ETLAQLQQEFQRAAKAGAPGK.A	4

Error (ppm): refers to the error window of the mass in parts per million. Miss: Refers to the b or y ions of a particular amino acid that were not identified in the MS/MS frame. Ion score: refers to the significance threshold, based on the calculated probability P,  $-10\log(P)$ . N.O.= not observed

*b*PFD subunits 4 and 6 were confidently identified in the 20-40 %  $(\text{NH}_4)_2\text{SO}_4$  pellet analysis due to their unique amino acid sequence and the low number of misses in the MS/MS peptide sequencing frame. However, the total ion score is low and ions correlating to *b*PFD subunits 1, 2, 3 and 5 were not identified in the analysis (Table 3.3). Consequently, it is likely that the extraction process may have denatured small amounts of *b*PFD complexes, resulting in disruption of the chaperone assembly into its smaller components. The individual *b*PFD subunits have different solubilities, and as a consequence, it is proposed that *b*PFD subunits 4 and 6 precipitated at lower concentrations of  $(\text{NH}_4)_2\text{SO}_4$  compared to the intact complex.

The MS analysis suggest with high confidence that the intact *b*PFD complex precipitated in the 40-60 % ammonium sulphate pellet. *b*PFD subunits 2-6 were observed to be present in the sample, with ion scores higher than 32 for each subunit and either no or low misses in the MS/MS peptide sequence analysis. *b*PFD subunit 1 is not observed in the MS analysis, however this does not necessarily mean that the subunit is not present in the sample. A common challenge frequently encountered in shotgun proteomics is the absence of peptides in the spectra analysis due to the limitations of automated LC-MS/MS systems [251], [301], [302]. Typically, automated LC-MS/MS systems select the most abundant ions in a scan for fragmentation. As a consequence, low abundance or poorly ionised proteins and peptides eluting from the LC can be overlooked in the

MS/MS process due to the overwhelming number proteolytic peptides in a crude sample. This is likely to have occurred in this situation.

The 40-60 % ammonium sulphate pellet was dissolved into an aqueous buffer. A series of chromatographic separations were then performed to further purify *b*PFD. The first column used in this purification was an anion exchange column, Q-Sepharose HP. The anion exchange column separates analytes by interaction with a medium that is positively charged. At equilibrium, anions in the aqueous mobile phase are attracted to the immobilised cation surface. When the sample is introduced the negatively charged sample ions displace the mobile phase ions [303], and leave positively charged particulates to elute. The column is thoroughly washed to remove all unwanted ions. A high salt buffer gradient is then applied to remove all proteins bound to the column [303]. The salt displaces the negatively charged sample ions from the column resin and a UV detector is able to measure the absorbance of the eluted proteins, indicating the abundance of proteins present in the sample.

Due to the high number of proteins observed in the proteomic analysis of the 40-60% (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> pellet (data not shown), a step wise salt gradient was applied in this column purification step. Following loading and a low salt wash of 20 mM NaH<sub>2</sub>PO<sub>4</sub>, proteins were eluted from the column at 100 mM NaH<sub>2</sub>PO<sub>4</sub>, 200 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM and then 400 mM NaH<sub>2</sub>PO<sub>4</sub>. A small sample of the proteins collected at these different NaH<sub>2</sub>PO<sub>4</sub> concentrations were denatured, proteolytically digested and subjected to shotgun proteomics. Table 3.4 reveals that *b*PFD subunit 2 was identified in the 200 mM NaH<sub>2</sub>PO<sub>4</sub> fraction.

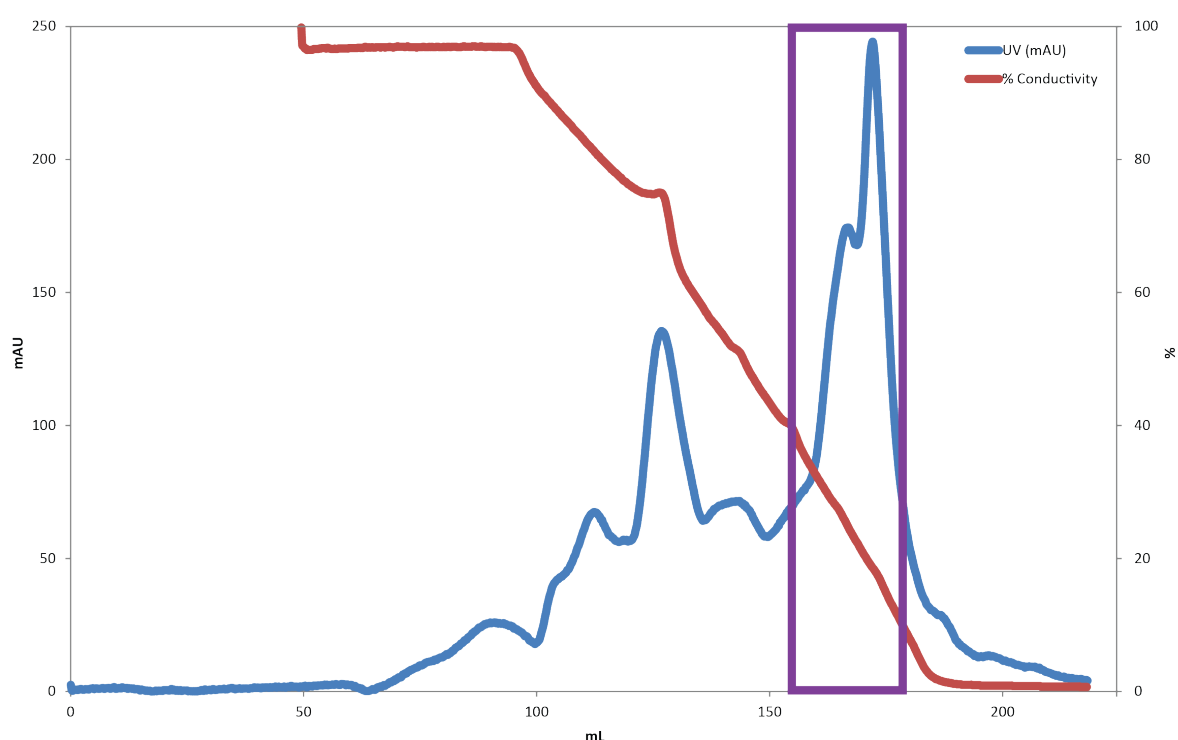
**Table 3.4: Ions identified by proteomic analysis correlating to *b*PFD subunits in the 200 mM NaH<sub>2</sub>PO<sub>4</sub> fraction from an anion exchange column.**

Fractions	Observed <i>m/z</i>	Mass expected	Mass calculated	Error (ppm)	Miss	Sequence	Ion Score
<b>100 mM NaH<sub>2</sub>PO<sub>4</sub></b>							
<i>b</i> PFD N.O.	-	-	-	-	-	-	-
<b>200 mM NaH<sub>2</sub>PO<sub>4</sub></b>							
<i>b</i> PFD subunit 2	721.8641 694.3937	2883.4275 2080.1592	2883.4171 2080.1528	4 3	1 2	K.AAELEMELNEHSLVIDTLKEVDETR.K R.TVKEVLPALENNKEQIQK.I	6 16
<b>300 mM NaH<sub>2</sub>PO<sub>4</sub></b>							
<i>b</i> PFD N.O.	-	-	-	-	-	-	-
<b>400 mM NaH<sub>2</sub>PO<sub>4</sub></b>							
<i>b</i> PFD N.O.	-	-	-	-	-	-	-

Error (ppm): refers to the error window of the mass in parts per million. Miss: Refers to the b or y ions of a particular amino acid that were not identified in the MS/MS frame. Ion score: refers to the significance threshold, based on the calculated probability P, -10Log(P). N.O.= not observed



Despite the fact that only *b*PFD subunit 2 was identified in the proteomic analysis, the fractions containing the protein were carried forward for further purification using a phenyl Superose column. This column separates proteins through their varying degrees of hydrophobicity. The column consists of hydrophobic groups (in this instance phenyl groups) attached to a stationary phase. The hydrophobic amino acid side chains on protein surfaces then interact and bind to the hydrophobic column. Buffers with a high salt content reduce solvation and further expose hydrophobic regions of a protein [304]. Therefore, a high ionic strength buffer is initially applied to allow the protein to interact with the hydrophobic groups. However, the more hydrophobic the protein is, the less salt is needed in the buffer for the protein to bind to the column. To desorb the protein from the column, the salt concentration is decreased [304]. The 200 mM NaH<sub>2</sub>PO<sub>4</sub> fraction was concentrated and buffer exchanged. Following loading and subsequent washing of the column with 1 M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, a linear decreasing salt gradient was applied, as seen in Figure 3.3.



**Figure 3.3:** The UV absorbance elution profile of *b*PFD from a phenyl Superose HP G20 column. The column was equilibrated and following loading was washed with an aqueous buffer of 20 mM Tris-HCl, 1 M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 1 mM MgCl<sub>2</sub>, 1 mM EGTA, 1 mM DTT at pH 7.2. A linear decreasing salt concentration gradient over 40 mins was applied, with the aqueous buffer 20 mM Tris-HCl, 1 mM MgCl<sub>2</sub>, 1 mM EGTA, 1 mM DTT at pH 7.2 and monitored through the % of conductivity (red). Proteins eluting from the column were detected through UV absorbance at  $\lambda$  280 nm (blue). Highlighted in purple are the fractions in which *b*PFD was identified.

A sample of the proteins eluting in fractions at 105 mL (96 % conductivity), 125 mL (72 % conductivity), 140 mL (54 % conductivity), 150 mL (41 % conductivity), 160 mL (28 % conductivity), and 170 mL (12 % conductivity) were denatured, proteolytically digested and subjected to proteomic analysis. The ions found correlating to *b*PFD subunits are summarised in Table 3.5.

**Table 3.5: Ions identified by proteomic analysis correlating to *b*PFD subunits in the fractions containing protein following elution from a phenyl Superose column. *b*PFD subunits observed are ordered by decreasing ion scores.**

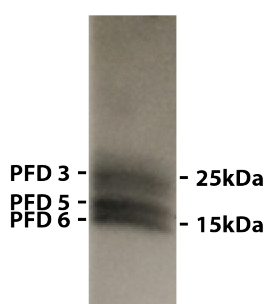
Fractions (Conductivity)	Observed <i>m/z</i>	Mass Expected	Mass Calculated	Error (ppm)	Miss	Sequence	Ion Score
<b>105 mL (96 %)</b>							
<i>b</i> PFD N.O.	-	-	-	-	-	-	-
<b>125 mL (72 %)</b>							
<i>b</i> PFD N.O.	-	-	-	-	-	-	-
<b>140 mL (54 %)</b>							
<i>b</i> PFD N.O.	-	-	-	-	-	-	-
<b>150 mL (41 %)</b>							
<i>b</i> PFD subunit 4	798.7430	2393.2071	2393.1975	4	2	K.AAAEDVNVTFEDQQKINKFAR.N	25
<b>160 mL (28 %)</b>							
<i>b</i> PFD subunit 3	628.8251	1255.6356	1255.6271	7	0	R.FLLADNLYCK.A	26
	463.5418	1850.1383	1850.1240	8	2	R.LKGQIPEIKQTLEILK.Y	9
<i>b</i> PFD subunit 4	454.2806	1359.8198	1359.8126	5	1	R.VLADLKVVQLYAK.F	20
	798.7448	2393.2126	2393.1975	6	2	K.AAAEDVNVTFEDQQKINKFAR.N	19
<i>b</i> PFD subunit 6	541.3295	1620.9668	1620.9563	6	1	K.LLGPVLVKQELGEAR.A	17
	578.1155	2885.5413	2885.5386	1	1	K.LEAQLTENNIVKEELALLDGSNNVFK.L	4
<i>b</i> PFD subunit 2	480.2740	958.5335	958.5270	7	0	R.MVGGVLVER.T	6
	694.3946	2080.1621	2080.1528	5	2	R.TVKEVLPALENNKEQIQK.I	17
<b>170 mL (12 %)</b>							
<i>b</i> PFD subunit 4	680.9186	1359.8227	1359.8126	7	1	R.VLADLKVVQLYAK.F	38
	454.2818	1359.8235	1359.8126	8	1	R.VLADLKVVQLYAK.F	23
	596.9990	1787.9751	1787.9628	7	2	R.NTSRITELKEEIEVK.K	9
	798.7448	2393.2126	2393.1975	6	2	K.AAAEDVNVTFEDQQKINKFAR.N	31
	599.3113	2393.2160	2393.1975	8	2	K.AAAEDVNVTFEDQQKINKFAR.N	18
<i>b</i> PFD subunit 1	631.3204	1890.9395	1890.9291	5	2	R.SVKEAEDNIREMLMAR.R	9
	636.6519	1906.9338	1906.9240	5	2	R.SVKEAEDNIREMLMAR.R	10
						(oxidation M)	
	636.6539	1906.9400	1906.9240	8	2	R.SVKEAEDNIREMLMAR.R	13
						(oxidation M)	
	911.4331	2731.2775	2731.2582	7	0	K.HAHLTDTEIMTLVDETNNMYEGVGR.M	15
	683.8270	2731.2787	2731.2582	8	0	K.HAHLTDTEIMTLVDETNNMYEGVGR.M	50
	687.8254	2747.2724	2747.2531	7	0	K.HAHLTDTEIMTLVDETNNMYEGVGR.M	50
						(oxidation M)	
	715.8494	2859.3686	2859.3531	5	1	K.HAHLTDTEIMTLVDETNNMYEGVGR.M	21
	719.8483	2875.3642	2875.3480	6	1	K.HAHLTDTEIMTLVDETNNMYEGVGR.M	10
						(oxidation M)	
	719.8500	2875.3708	2875.3480	8	1	K.HAHLTDTEIMTLVDETNNMYEGVGR.M	3
						(oxidation M)	
	773.1376	3088.5212	3088.4958	8	2	R.TKKHAHLTDTEIMTLVDETNNMYEGVGR.M	12
	777.1337	3104.5056	3104.4907	5	2	R.TKKHAHLTDTEIMTLVDETNNMYEGVGR.M	5
						(oxidation M)	
	777.1339	3104.5065	3104.4907	5	2	R.TKKHAHLTDTEIMTLVDETNNMYEGVGR.M	11
						(oxidation M)	
	621.9102	3104.5144	3104.4907	8	2	R.TKKHAHLTDTEIMTLVDETNNMYEGVGR.M	8
						(oxidation M)	
<i>b</i> PFD subunit 3	547.2914	1092.5683	1092.5637	4	0	K.FMELNLAQK.K	4
	628.8252	1255.6358	1255.6271	7	0	R.FLLADNLYCK.A	41
	617.7189	1850.1348	1850.1240	6	2	R.LKGQIPEIKQTLEILK.Y	29
	463.5425	1850.1408	1850.1240	9	2	R.LKGQIPEIKQTLEILK.Y	24
	991.4777	2971.4112	2971.3869	8	1	K.NLDSLEEDLDFLRDQFTTEVNMAR.V	1
<i>b</i> PFD subunit 6	541.3303	1620.9691	1620.9563	8	1	K.LLGPVLVKQELGEAR.A	31
	678.7046	2033.0919	2033.0793	6	2	K.LQGEVEKYQQLQKDLK.S	10

Error (ppm): refers to the error window of the mass in parts per million. Miss: Refers to the b or y ions of a particular amino acid that were not identified in the MS/MS frame. Ion score: refers to the significance threshold, based on the calculated probability P, -10Log(P). N.O.= not observed

The fractions between 155-175 mL (34-5 % conductivity) (identified to contain the majority of *b*PFD subunits) were next combined, concentrated, buffer exchanged and further purified over a mono S

column. The mono S column is a strong cation exchanger and works similarly to the Q-sepharose HP, however with the opposite polarity. Following loading, the column was thoroughly washed with 20 mM NaCl to remove all unwanted ions and a linear high salt buffer gradient to 250 mM NaCl was applied to remove all proteins bound to the column. The salt displaces the positively charged ions from the column resin, and the eluted sample was fractionated. A western blot analysis of the fractions containing protein revealed that *b*PFD did not bind to the column and was found in the load and wash (Figure 3.4).

It should be noted that earlier attempts to isolate *b*PFD involved probing protein fractions by Western transfer and immunoblotting utilising monoclonal antibodies (PFD subunit 3, PFD subunit 5 and PFD subunit 6) from Santa Cruz (Dallas, Texas). Unfortunately, several different batches of antibodies purchased were all found to display significant cross-reactivity with the abundant IgG in the sample, which hindered this approach.



**Figure 3.4:** A representative western blot analysis of the fraction (washings) containing *b*PFD following elution from the mono S column using antibodies probing for *b*PFD subunit 3, 5 & 6.

The washings from the mono S column were therefore combined, concentrated, buffer exchanged and purified over a ceramic hydroxylapatite (CHT) column. The fourth column used in this purification utilises hydroxylated calcium phosphate, also known as hydroxylapatite to separate proteins based on their electrostatic interaction with the matrix. Hydroxylapatite interacts with proteins through phosphate sites that bind to amino groups, whilst calcium sites attract carboxyl groups on the protein [305]. Protein desorption can be conducted by reversing phosphate site interactions or calcium site interactions using either linear or step gradients. Typically, phosphate buffers are commonly employed desorb proteins from the CHT column. Following loading of the *b*PFD containing extracts and a low salt wash of 20 mM NaH<sub>2</sub>PO<sub>4</sub>, a linear gradient to 250 mM NaH<sub>2</sub>PO<sub>4</sub> was applied and fractions of 1 mL were collected. Figure 3.5 reveals 2 major UV peaks eluting from the CHT column.

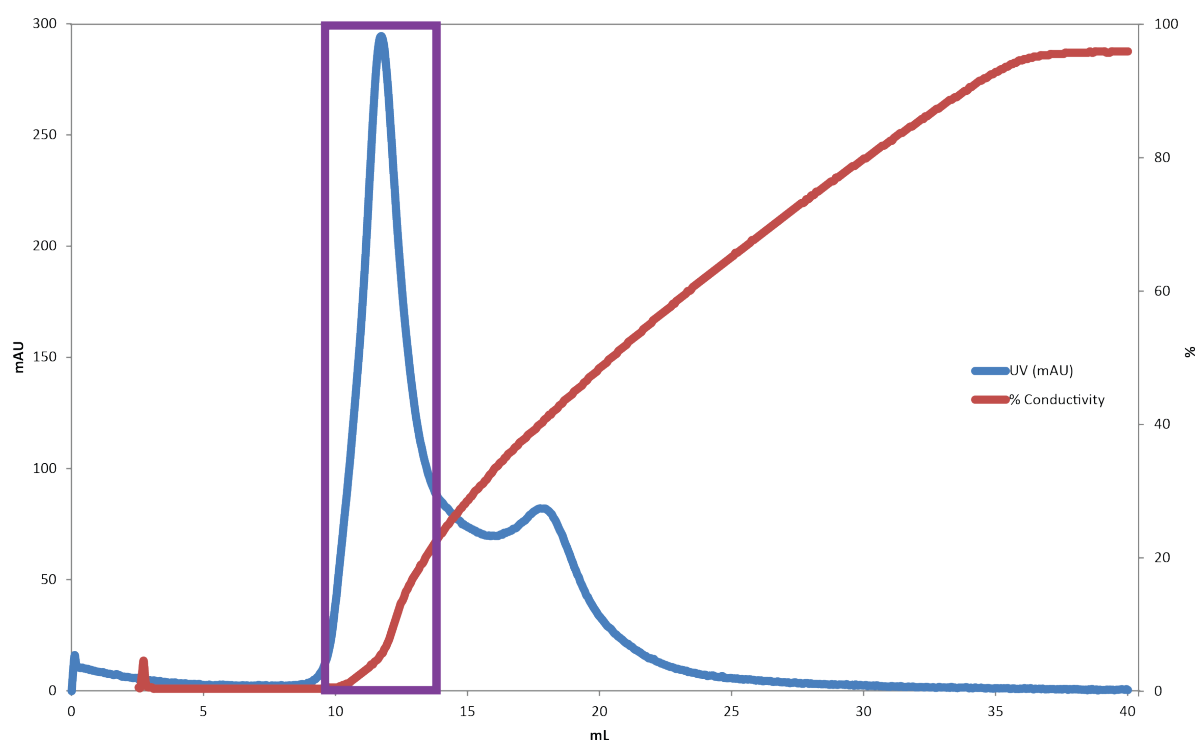


Figure 3.5: The UV absorbance elution profile of *bPFD* from a CHT type II column. The column was equilibrated and following loading was washed with an aqueous buffer of 20 mM  $\text{NaH}_2\text{PO}_4$ , 20 mM KCl, 1mM  $\text{MgCl}_2$ , 0.1 mM  $\text{CaCl}_2$ , 1 mM DTT at pH 6.8. A linear increasing salt concentration gradient over 30 mins was applied, with the aqueous buffer 250 mM  $\text{NaH}_2\text{PO}_4$ , 20 mM KCl, 1mM  $\text{MgCl}_2$ , 0.1 mM  $\text{CaCl}_2$ , 1 mM DTT at pH 6.8 and monitored through the % of conductivity (red). Proteins eluting from the column were detected through UV absorbance at  $\lambda$  280 nm (blue). Highlighted in purple are the fractions in which *bPFD* was identified.

Fractions at 12 mL (conductivity of 18 %) and 18 mL (conductivity of 31 %) were analysed by SDS PAGE. This was then subjected to a Western transfer and probed by PFD subunit 3, 5 and 6 antibodies (Figure 3.6). *bPFD* was seen in the first peak eluting from the column. Fractions 9-15 mL were combined and concentrated before loading onto a Superdex 200 (S200).

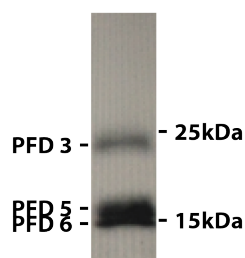


Figure 3.6: A representative western blot analysis of the fraction (at 12 mL) containing *b*PFD following elution from the CHT column using antibodies probing *b*PFD subunit 3, 5 & 6.

The last column used in this series of chromatographic purification steps was the S200, a size exclusion column separating proteins based on their size. Also known as gel-filtration chromatography, the medium of a size exclusion column enables separation by trapping smaller molecules in the absorbent material and allowing larger molecules to pass through [306], [307]. This method is also useful for exchanging large proteins into a different buffer. *b*PFD was found to elute between 11-16 mL as seen in Figure 3.7.

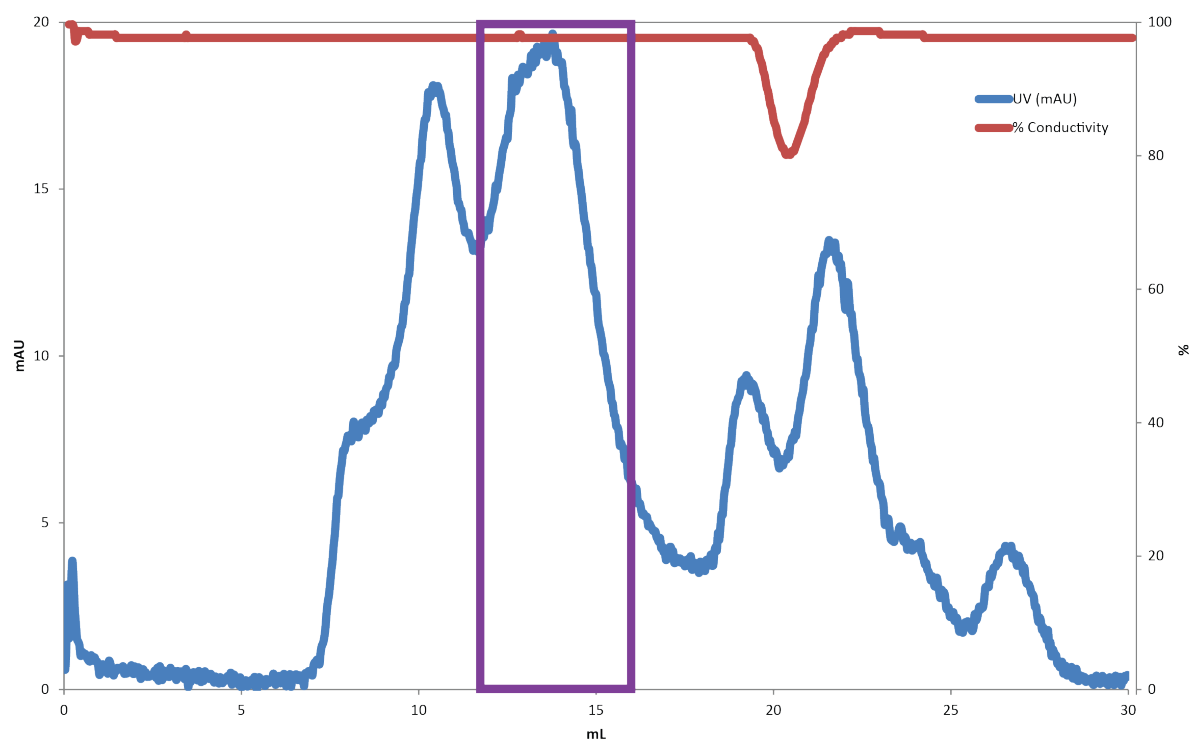


Figure 3.7: The UV absorbance elution profile of *b*PFD from a size exclusion S200 column. The aqueous used was 100 mM AA buffer at pH 7.4. and monitored through the % of conductivity (red). Proteins eluting from the column were detected through UV absorbance at  $\lambda$  280 nm (blue). Highlighted in purple are the fractions in which *b*PFD was identified.

Fractions containing protein were analysed by SDS PAGE. The resulting gel was subjected to a Western transfer and fractions containing *b*PFD were probed using antibodies probing for PFD subunits 3, 5, and 6 (Figure 3.8).

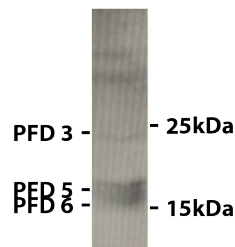


Figure 3.8: A representative western blot analysis of the fraction (at 13 mL) containing *b*PFD following elution from the S200 column using antibodies *b*PFD subunit 3, 5 & 6.

The SDS PAGE gel analysis was replicated using the same protein fractions and subjected to silver stain (data not shown). Unfortunately, no protein bands were observed on the gel, despite bands seen in the Western blot analysis. This suggests the protein concentrated to be between 10-0.1 ng as the sensitivity of silver staining is marked at 10 ng and Western blots are able to image protein concentrations of 0.1 ng [308]. Purity of the sample cannot be determined through these means. Therefore, a sample of the fraction found to contain the predominant amount of *b*PFD as probed by antibodies (fraction collected at 13 mL) was subjected to proteolysis and analysed by shotgun proteomics to confirm the presence of *b*PFD. Table 3.7 reveals *b*PFD subunits 1-4 and 6 were identified in the spectra.

**Table 3.6: Ions identified by proteomic analysis correlating to *b*PFD subunits in the fractions containing protein following elution from Superdex 200. *b*PFD subunits observed are ordered by decreasing ion scores.**

Fraction	Observed <i>m/z</i>	Mass Expected	Mass Calculated	Error (ppm)	Miss	Sequence	Ion Score
<b>13 mL</b>							
<i>b</i> PFD subunit 6	419.7919	837.5692	837.5688	0	0	K.LLGPVLVK.Q	29
	407.8958	1220.6655	1220.6765	-9	1	R.LDYITAEIKR.Y	10
	459.9311	1376.7714	1376.7776	-4	2	K.RLDYITAEIKR.Y	44
	530.9514	1589.8324	1589.8413	-6	1	K.LQGEVEKYQQLQK.D	14
	538.3094	1611.9065	1611.9195	-8	2	M.AELIQKKLQGEVEK.Y	17
	811.4819	1620.9492	1620.9563	-4	1	K.LLGPVLVKQELGEAR.A	35
	541.3245	1620.9516	1620.9563	-3	1	K.LLGPVLVKQELGEAR.A	70
	573.6508	1717.9304	1717.9362	-3	2	K.KLQGEVEKYQQLQK.D	16
	509.2742	2033.0675	2033.0793	-6	2	K.LQGEVEKYQQLQKDLK.S	6
	678.6966	2033.0681	2033.0793	-6	2	K.LQGEVEKYQQLQKDLK.S	17
<i>b</i> PFD subunit 4	648.2917	1294.5689	1294.5677	1	0	K.FGSNINLEADES.-	16
	454.2780	1359.8121	1359.8126	0	1	R.VLADLKVVQLYAK.F	73
	680.9135	1359.8124	1359.8126	0	1	R.VLADLKVVQLYAK.F	51
	599.3040	2393.1868	2393.1975	-4	2	K.AAAEDVNVTFEDQQKINKFAR.N	70
<i>b</i> PFD subunit 2	480.2682	958.5218	958.5270	-5	0	R.MVGGVLVER.T	63
	488.2669	974.5192	974.5219	-3	0	R.MVGGVLVER.T (Oxidation M)	48
	693.4026	1384.7906	1384.7926	-1	0	K.IIETLTQQLQAK.G	70
	442.4901	1765.9312	1765.9331	-1	2	K.CYRMVGGVLVERTVK.E	1
	521.0409	2080.1345	2080.1528	-9	2	R.TVKEVLPALENNKEQIQK.I	18
	694.3857	2080.1352	2080.1528	-8	2	R.TVKEVLPALENNKEQIQK.I	15
<i>b</i> PFD subunit 3	547.2867	1092.5589	1092.5637	-4	0	K.FMELNLAQK.K	55
	555.2836	1108.5526		-5	0	K.FMELNLAQK.K (Oxidation M)	9
	407.8958	1220.6655		6	1	K.FMELNLAQKK.R	7
	628.2492	1255.6292	1255.6271	2	0	R.FLLADNLYCK.A	68
<i>b</i> PFD subunit 1	462.2492	1383.7258		3	1	R.YKFMELNLAQK.K	19
	433.7423	865.4701	865.4731	-4	0	R.MFILQSK.E	38
	729.4209	1456.8273	1456.8136	9	2	K.IAEKIKELEQK.K	6

Error (ppm): refers to the error window of the mass in parts per million. Miss: Refers to the b or y ions of a particular amino acid that were not identified in the MS/MS frame. Ion score: refers to the significance threshold, based on the calculated probability  $P$ ,  $-10\log(P)$ . N.O.= not observed

Unfortunately the proteomic and native MS analysis of the fraction containing *b*PFD revealed that there was contamination by the protein endoplasmin (Appendix C). Approximately 1/3 of the sample is estimated to be endoplasmin (Section 7.2.1.1). The sample was concentrated using an Amicon membrane (30 KDa) fitted concentrator and the concentration was calculated using a Bradford coomassie assay (ThermoFisher Scientific Pierce, Massachusetts, United States). The proteins were frozen and stored at -20 °C until subjected to analysis or required for further assays. Concentrations were adjusted accordingly to include the contamination of endoplasmin. Overall yield was found to be approximately 500 µg.

### 3.2.3: Expression and Isolation of *h*PFD Subunits

Individual *h*PFD subunits were obtained through recombinant protein expression. To facilitate the purification of the subunits, each subunit was expressed as a fusion protein with a His-tag using the pPROX HTb bacterial expression vector (Aldevron) allowing cleavage of the fusion partner post-expression [309].

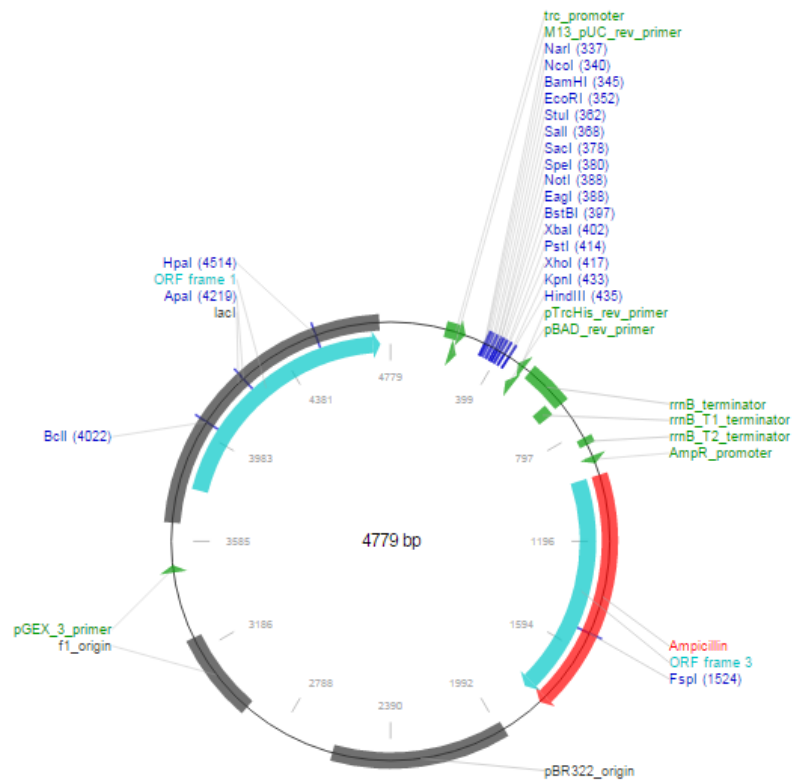


Each subunit was adapted to accommodate a 6 x His tag, and a TEV protease site was included between the His-tag and prefoldin subunit sequences. The ideal recognition site for TEV protease is ENLYFQG, with cleavage occurring between the glutamine and glycine residues, leaving an N-terminal glycine residue on the passenger protein [310], [311]. Therefore, an N-terminal sequence of MSYYHHHHHDYDIPTTENLYFQGAMGS and a C-terminal fusion sequence of AAAFESRACSLACGTLGCFGG was added. The oligonucleotides encoding the PFD subunit sequences and the translated amino acid sequences are shown in Appendix D.

The cDNA sequence corresponding to *h*PFD subunits 1-6 were cloned into the restriction sites BamHI (345bp) and NotI (388bp) (Figure 3.9). Upon successful sequencing and cloning of the inserted vector, the plasmid was transformed into *Escherichia coli* BL21(DE3) cells for protein expression tests. After transformation, the cells were grown on an agar and ampicillin (0.1%) plate. A colony was inoculated into an overnight culture and then scaled up and expressed the next day using IPTG for induction. Cellular growth was monitored by optical density at a wavelength of 600 nm. Growth to induction density ( $OD_{600}$  approximately 0.4-0.6) was rapid, taking approximately 2 hours, and growth to completion ( $OD_{600}$  approximately 2.5-2.8) was equally as rapid, taking approximately 7 hours. The cells were harvested and lysed with the addition of lysozyme and freeze thaw cycles. The supernatant was dialysed to remove detergent and this was loaded onto a pre-equilibrated nickel affinity column. The nickel column is able to form a metal chelate with the poly His-tag on the N-terminus of the subunits [312]. Washing the column with low amounts of imidazole aids in selectivity as the histidine analogue removes clusters of histidine exclusive of the tag bound to the stationary phase. Increasing the amount of imidazole competitively elutes the bound polyhistidine residues [313].

All 6 *h*PFD subunits were cloned, expressed and analysed by SDS PAGE utilising a solution of 50mM  $NaH_2PO_4$  and 0.1% SDS to extract and solubilise the subunits. All the subunits were seen to be highly expressed when analysed through SDS PAGE (data not shown), however, SDS was difficult to remove from the samples and refolding of the denatured proteins could not be achieved. Therefore, a solution of 50mM  $NaH_2PO_4$  and 0.1% TWEEN 100 was used instead to extract and solubilise the expressed subunits. SDS PAGE analysis (of the same batch of expressed proteins as previously extracted by 0.1 % SDS) following extraction with this solution revealed that many of the subunits could not be solubilised with TWEEN. Comparisons between the SDS PAGE analyses of the samples extracted with the different detergent solutions showed inclusion bodies forming when utilising TWEEN 100. This ultimately led to difficulty in the extraction process. Fortunately, it was found that *h*PFD subunits 5 and 6 were soluble with the aid of TWEEN 100 and although there is a lack of global

sequence similarity between the subunits (Table 3.2), there are regions of conserved sequence similarities [79], [98]. It is for these reasons that this thesis focuses on *h*PFD subunit 5 to represent  $\alpha$ -type subunits and subunit 6 to represent  $\beta$ -type subunits of the complex PFD.



**Figure 3.9: Schematic representation of the bacterial expression vector pPROEX HTb. Features include the trc promoter, M13 pUC rev primer, pTrcHis rev primer, pBAD rev primer, rrnB T1 terminator, rrnB T2 terminator, AmpR promoter, Ampicillin, pBR322 origin, pGEX 3 primer, & lacI. Adapted from [314].**

Fractions containing protein from the nickel affinity column were combined, concentrated and purified over an Superdex 200 size exclusion column as seen in Figure 3.10 for *h*PFD subunit 6, which was found to elute from the column at approximately 15 mL. Proteases and additional proteins related to *E. coli*. were eluted from the Superdex 200 column before and after the elution of *h*PFD subunit 6 as seen in Figure 3.10. Unfortunately, the S200 column was not calibrated. The protein was analysed by MS following isolation and hence an approximation of the molecular weight was not necessary.

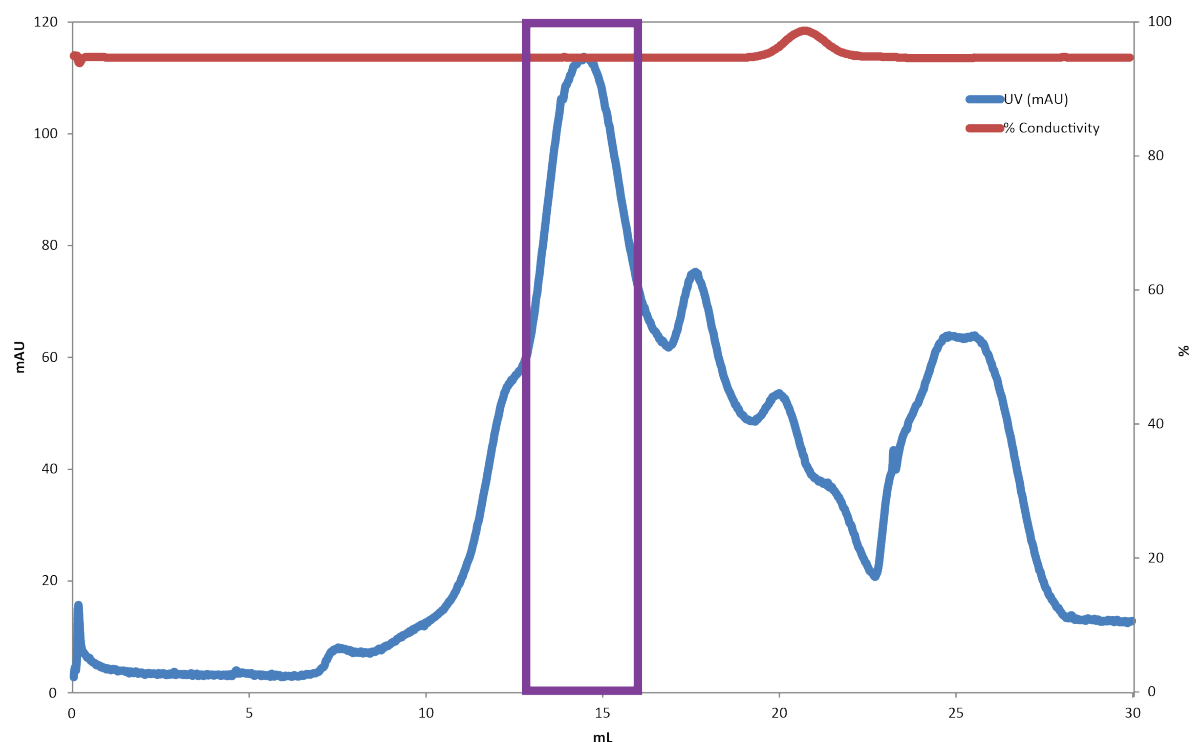
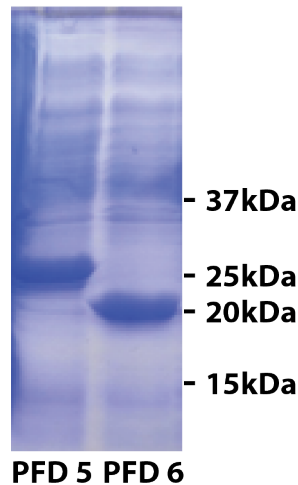


Figure 3.10: The UV absorbance elution profile of *hPFD* subunit 6 from a size exclusion S200 column. The aqueous used was 100 mM AA buffer at pH 7.4. and monitored through the % of conductivity (red). Proteins eluting from the column were detected through UV absorbance at  $\lambda$  280 nm (blue). Highlighted in purple are the fractions in which *hPFD* was identified.

Purified subunits were incubated with TEV, however less than half of the protein was found in the truncated form (analysed by SDS PAGE, data not shown). There is some precedent to suggest, for certain proteins, there is little effect on structure or function with the fusion tags located on both termini [315]. Ideally, comparisons between tagged and non-tagged forms of the *hPFD* subunits in structural and functional assays would give an indication to whether the fusion His-tag has an effect on the overall characteristics of the protein. Unfortunately, here it was not possible to produce the untagged protein in high quantity and with minimal steps of purification, and hence the tag was left on the sequence. Since the protein was shown to form oligomers and have chaperone activity, it was assumed the tag had minimal influence.

The subunit identities were confirmed by a combination of Western blotting, SDS gel electrophoresis (Figure 3.11) or by MS analysis (Section 6.2.2&6.2.3). Shotgun MS was not required at this stage due to the increased purity of the samples.



**Figure 3.11:** *h*PFD subunits 5 & 6 following overexpression, SDS PAGE 15% analysis stained by Coomassie. Densitometric analysis of the gel revealed approximately 35 % of purity for both expressed proteins

Densitometric analysis of the gels were conducted using an open software called ImageJ (NIH, Maryland) to measure the purity of the proteins. Analysis of the gels directly following overexpression reveals a purity of approximately 35 % for both expressed *b*PFD subunits. Samples following the purification protocol detailed in Section 3.3.3 were subjected to SDS PAGE and densitometric analysis and were found to be pure (data not shown). Concentrations of the samples were then measured with a Nano 2000 (Varian Ltd, Australia). The purified *h*PFD subunits were frozen and stored at -20 °C until required for analysis. Suitable material for subsequent structural and functional analyses was successfully obtained (Section 6.2.2, 6.2.3).

### 3.2.4: The Biological Activity of Eukaryotic PFD

#### 3.2.4.1: Chaperone Properties of *b*PFD and *h*PFD Subunits 5 and 6

The chaperoning abilities of PFD has been found to extend to the misfolding peptides of A $\beta$  [80]–[82], [102], [316], and as previously mentioned in Chapter 1, the individual subunits of PFD also exhibit unique functionality. The chaperoning abilities of the extracted *b*PFD complex and individual *h*PFD subunits 5 and 6 are therefore explored in this section.

Due to its aetiology in Alzheimer’s disease and similar biological activity to A $\beta$ (1-42), the amyloid beta isoform A $\beta$ (1-40) was chosen as the prefibrillar peptide to utilise for aggregation assays. Although A $\beta$ (1-42) has been found to be the predominant neurotoxic peptide [317], the 2 extra amino acids increases the hydrophobicity of the peptide, resulting in a faster rate of fibrillogenesis at

physiological pH and therefore manageability is difficult [280], [318], [319]. Both monomeric A $\beta$  peptides convert from a random coil structure to the highly structured  $\beta$ -sheet fibrils *in vitro* through the incubation of the peptide at neutral pH and at 37 °C.

To measure the influence of the extracted *b*PFD complex and individual *h*PFD subunits 5 and 6 on fibril inhibition, a ThT assay was first performed. ThT has the unique property of binding to cross  $\beta$ -sheet structures, upon which it becomes highly fluorescent as described in Section 2.2.1. The more cross  $\beta$ -sheet rich a sample is, the more binding occurs and a higher amount of fluorescence is observed. This characteristic enables the quantification and monitoring of cross  $\beta$ -sheet amyloid fibril formation.

A $\beta$ (1-40) was incubated with either *b*PFD, *h*PFD subunit 5 or *h*PFD subunit 6 and the dye ThT. There are 3 phases of amyloid fibril formation *in vitro*; namely the lag phase, elongation phase and plateau phase. These phases give insight into the nucleation rate, growth rate and quantity of fibril formation, respectively. The peptide A $\beta$ (1-40) in the absence of molecular chaperones was found to have a lag phase of approximately 30 minutes before a sigmoidal increase in ThT fluorescence emission intensity (490 nm) was observed. This lag phase suggests the peptides existed primarily in monomeric form before incubation [320]. The ThT fluorescence for A $\beta$ (1-40) alone was found to plateau at approximately 25 hours, reaching fluorescence at approximately 2200 a.u.. The formation of amyloid fibrils was later confirmed by TEM (Refer to Section 3.2.4.3). A sample of co-incubated A $\beta$ (1-40) and IgG at a ratio of 1:0.02, under the same conditions, was shown to affect the elongation phase when compared to the fibril formation of independent A $\beta$ (1-40). However, fibril plateau was reached at a similar time frame of approximately 25 hours. Although this co-incubation was initially used as a negative control, several papers revealed the possible chaperoning functions of the antibody [321], [322]. Consequently, the control conditions here are the samples incubated without PFD. It cannot be concluded from this data whether or not *b*PFD directly forms a stable complex with the misfolding peptide. Further biophysical characterisations are required to examine the possibility of protein complex formation.

No change in ThT fluorescence was produced by *b*PFD, *h*PFD subunit 5 or *h*PFD subunit 6 when the proteins were incubated independently under the same conditions which promoted the conversion of the A $\beta$ (1-40) monomer to amyloid fibrils, i.e. 37 °C, pH 7.4, without shaking. This result suggested the  $\alpha$ -helical proteins did not form any amyloid fibrils over the course of the experiment. To verify this, samples were viewed by TEM (Figure 3.17). As there were no aggregates observed, the proteins are inherently not amyloidogenic nor likely to interact directly with free ThT.

To conserve the low quantities of the extracted *b*PFD, it was decided that ThT assays conducted with the intact complex would be at low chaperone ratios. The addition of *b*PFD to A $\beta$ (1-40) at a ratio of 1:0.02 was observed to have a profound effect on both the elongation and plateau phase of the fibril formation. It was seen, however, that the lag phase remained similar to that observed in the absence of a chaperone. Figure 3.12 reveals the inhibition of fibril formation when prefibrillar A $\beta$ (1-40) is treated with *b*PFD. The elongation phase is observed to have a decreased rate of fibril formation and plateau is seen to be reached at approximately 35 hours, indicating that the prefibrillar A $\beta$ (1-40) has reached structural equilibrium. As the assay utilises A $\beta$ (1-40) from the same stock, the fibril formation can be quantified relative to the batch it is run in, suggesting *b*PFD is acting as a molecular chaperone towards A $\beta$ (1-40). Unfortunately, differences in independent A $\beta$ (1-40) fibril formations were observed in varying batches of the peptide, therefore representative data is presented. It should also be noted that the literature evidence supports that the contaminant endoplasmic reticulum chaperone, is able to inhibit A $\beta$  fibril formation in the presence of ATP [323]–[325], and Evans *et al.* demonstrated that the co-incubation of the prefibrillar amyloid peptide, Hsp90 and a nonhydrolysable ATP analogue was able to inhibit approximately 30 % of A $\beta$ (1-42) from forming fibrils at a ratio of 1:0.01 (A $\beta$ :Hsp90) [324]. However, all ThT assays in this Chapter were conducted in the absence of ATP or other comparable co-factors, and therefore the chaperone activity observed can likely be attributed to *b*PFD.

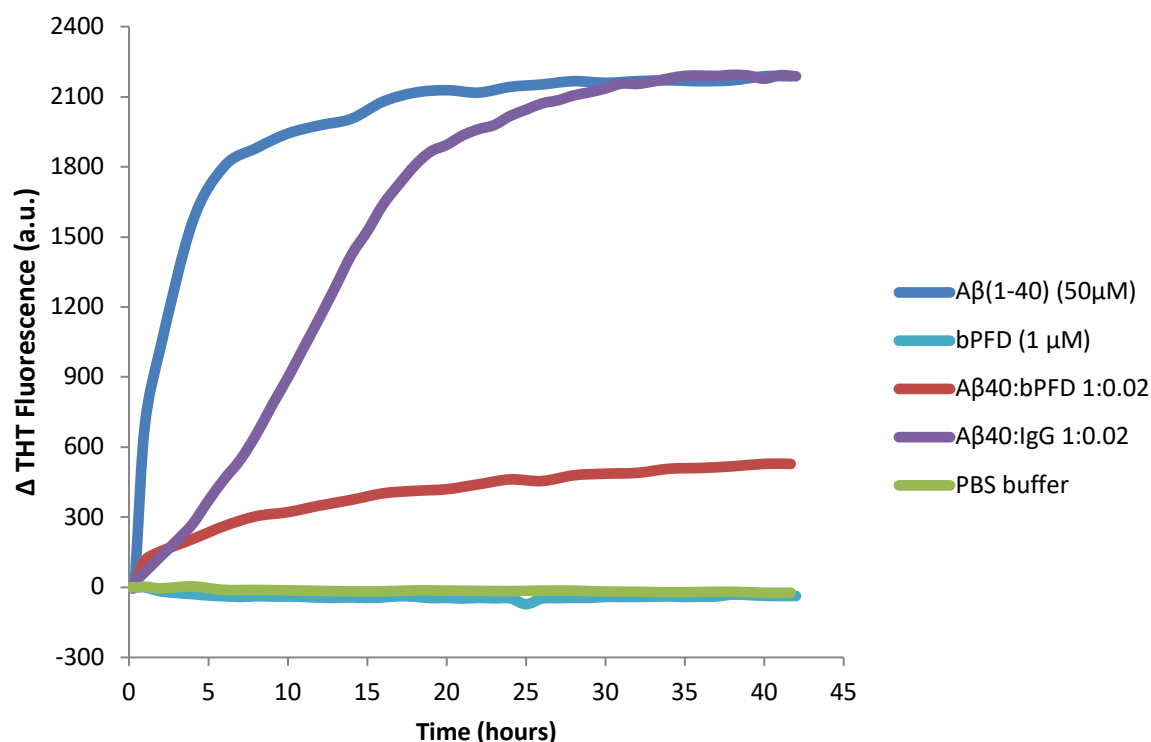


Figure 3.12: *b*PFD inhibits A $\beta$ (1-40) fibril formation. A $\beta$ (1-40) (50  $\mu$ M) was incubated at 37 °C with ThT in the absence & presence of *b*PFD and IgG at a concentration of 1  $\mu$ M. *b*PFD (1  $\mu$ M) was also incubated independently. Representative data are shown from three independent experiments performed in triplicate, which were seen to exhibit similar trends. The lag phase for all incubations was approximately 30 minutes.

Similar to the co-incubation of A $\beta$ (1-40) and *b*PFD, the addition of *h*PFD subunit 5 to A $\beta$ (1-40) appears to have a profound effect on the fibril formation of the prefibrillar peptide. The addition of substoichiometric amounts of *h*PFD subunit 5 (1:0.02 ratio *m*A $\beta$ (1-40):subunit 5) observably alters the elongation phase and the plateau phase of fibril formation, as seen in Figure 3.13. Although the lag phase of the fibril formation does not change, the rate of fibril elongation has drastically decreased with the addition of *h*PFD subunit 5. For example, at 10 hours the change in fluorescence differs significantly between the incubation of A $\beta$ (1-40) without *h*PFD subunit 5 and with the addition of the protein at the ratio of 1:0.02, with  $\Delta$ ThT at 1500 a.u. and 1100 a.u. respectively. A $\beta$ (1-40) incubated alone plateaus at approximately 25 hours, indicating that structural interconversions have reached equilibrium. In comparison, upon addition of *h*PFD subunit 5 at a ratio of 1:0.02, the plateau is observed at approximately 35 hours with a lower overall  $\Delta$ ThT fluorescence. As the assay utilises A $\beta$ (1-40) from the same stock, the fibril formation can be quantified relative to the batch it is run in, suggesting *h*PFD subunit 5 is acting as a molecular chaperone towards A $\beta$ (1-40).

Increasing the chaperone concentration of *h*PFD subunit 5 to a ratio of 1:0.1 is seen to further affect the lag time, the elongation rate and the plateau of fibrillation. The nucleation of A $\beta$ (1-40) was delayed for approximately one hour following the addition of *h*PFD subunit 5 at the ratio of 1:0.1. The rate of fibrillogenesis is seen to also significantly decrease and at plateau,  $\Delta$ ThT is less than a third in comparison to A $\beta$ (1-40) incubated independently. A $\beta$ (1-40) incubated with the chaperone *h*PFD subunit 5 at higher ratios of 1:0.2 and 1:0.5 reveal no increase in ThT fluorescence throughout the incubation period (85 hr) suggesting the abolishment of amyloid fibril formation by A $\beta$ (1-40).

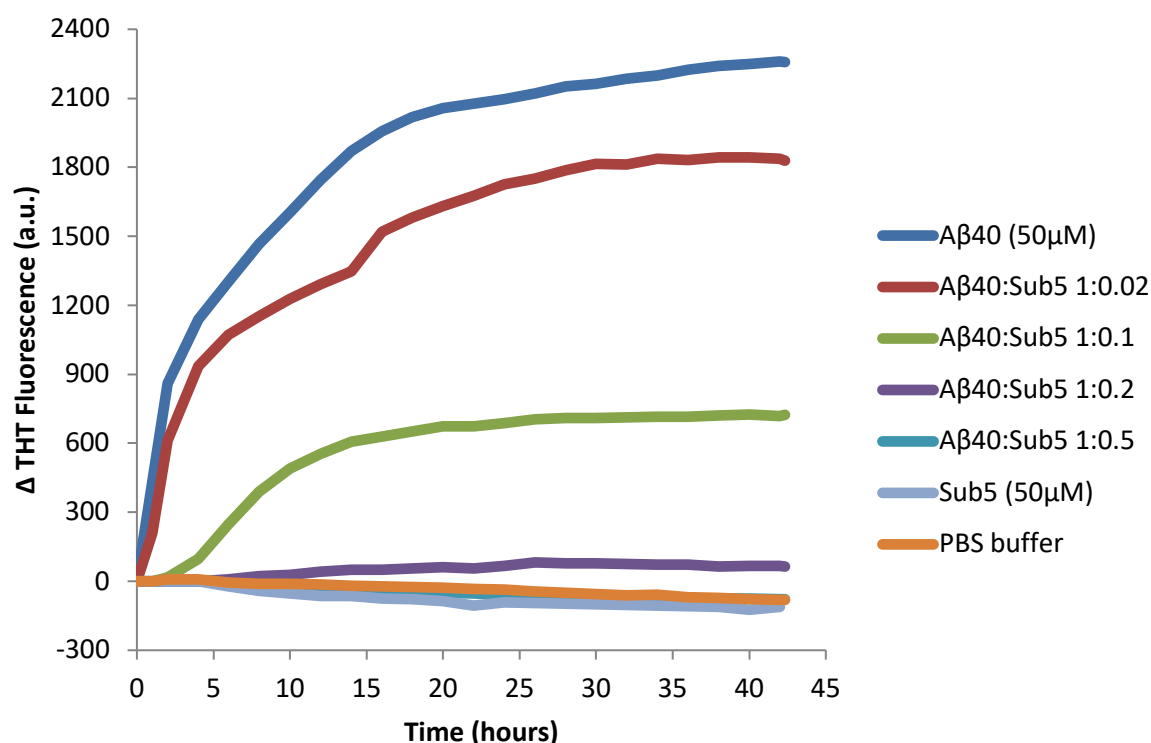


Figure 3.13: *hPFD* subunit 5 inhibits Aβ(1-40) fibril formation. Aβ(1-40) (50 μM) was incubated at 37 °C with ThT in the absence & presence of varying concentrations of *hPFD* subunit 5. Representative data are shown from three independent experiments performed in triplicate, which were seen to exhibit similar trends. The lag phase for all incubations was approximately 30 minutes.

The addition of *hPFD* subunit 6 to monomeric Aβ(1-40) in aggregating conditions at low concentrations was found to have an even greater effect on the formation of β-sheet rich fibrils (Figure 3.14) than *hPFD* subunit 5. The addition of *hPFD* subunit 6 to Aβ(1-40) at a ratio of 1:0.02 (Aβ(1-40):subunit 6) was found to significantly reduce the rate of fibrillogenesis, and at plateau, ΔThT is seen to be reduced to less than a third than that of the incubation of Aβ(1-40) alone. Increasing the concentration of *hPFD* subunit 6 to the ratio of 1:0.1 and higher was found to completely inhibit the misfolding pathway of the disordered amyloid peptide. This indicates that *hPFD* subunit 6 has a high chaperone propensity for the misfolding peptide and was able to directly interact with the prefibrillar monomer and inhibit the formation of β-sheets.



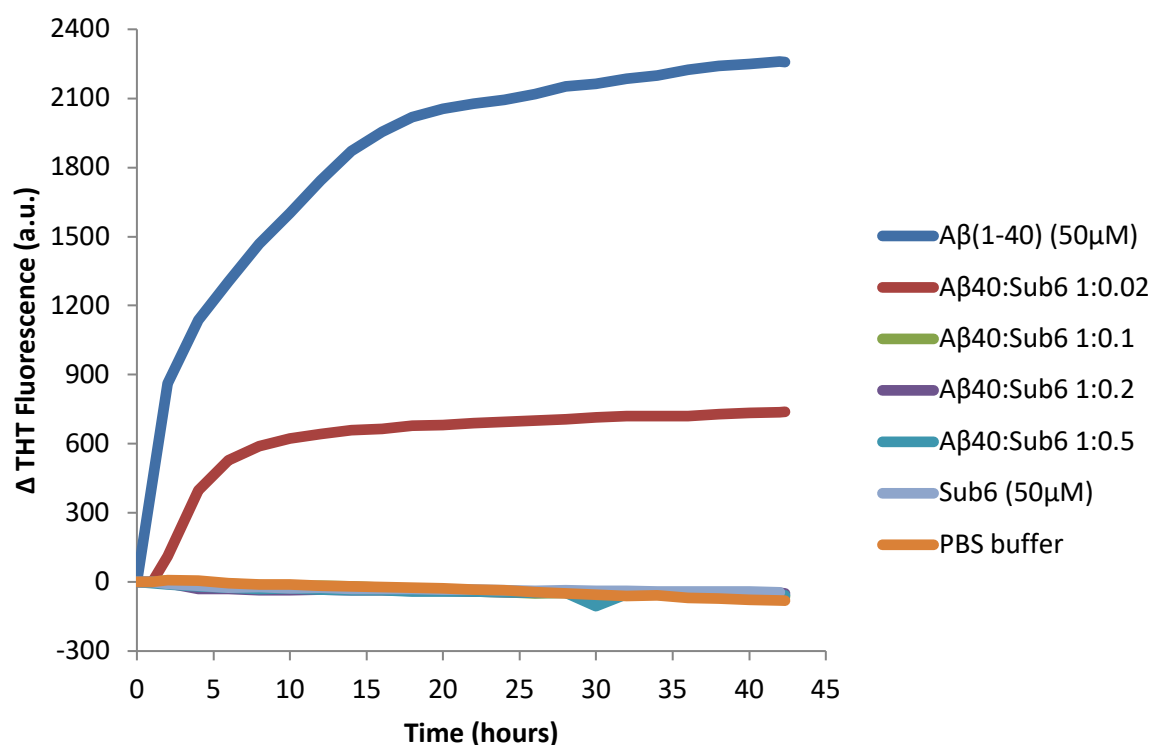


Figure 3.14: *h*PFD subunit 6 inhibits Aβ(1-40) from forming fibrils. Aβ(1-40) (50 μM) was incubated at 37 °C with ThT in the absence & presence of varying concentrations of *h*PFD subunit 6. Representative data are shown from three independent experiments performed in triplicate, which were seen to exhibit similar trends. The lag phase for all incubations was approximately 30 minutes.

Although sub-stoichiometric concentrations of *b*PFD were used in this assay to conserve the protein for further assays, there is clear indication of fibril inhibition. To compare the effectiveness of the intact complex and the individual subunits 5 and 6 against the prefibrillar Aβ(1-40), the plateau phase of each co-incubation with Aβ(1-40) at the ratio of 1:0.02 is plotted in Figure 3.15. ΔThT fluorescence at plateau was normalised against the fluorescence observed for unaccompanied Aβ(1-40) at structural equilibrium and it can be seen from Figure 3.15 that *b*PFD was able to inhibit Aβ(1-40) from forming fibrils to a greater extent than the individual subunits, with a 76 % reduction in fibril formation compared to 67 % and 18 % for *h*PFD subunit 5 and 6 respectively.

In an effort to determine the free Aβ(1-40) and the possible complex formed following *b*PFD and *h*PFD subunit chaperoning, samples from the ThT assays were briefly centrifuged (1000 *g*, 5 minutes) and the supernatant analysed by ESI-MS. Unfortunately, the spectra obtained showed only salts within the buffer and attempts at desalting resulted in spectra without protein, suggesting protein precipitation following buffer exchange.

ThT assays were also conducted on preformed fibrils to determine the disruptive behaviour of *b*PFD and *h*PFD subunits. It was found that the chaperone was unable to disengage the structured fibrils (data not shown). However, the addition of each protein at a ratio of 1:0.5 ( $A\beta(1-40)$ :protein) during the elongation phase of  $A\beta(1-40)$  was found to inhibit further fibril formation (data not shown). However, as the preformed fibrils remained present in the assays, disruption of fibrils by *b*PFD and *h*PFD subunits was not thoroughly explored.

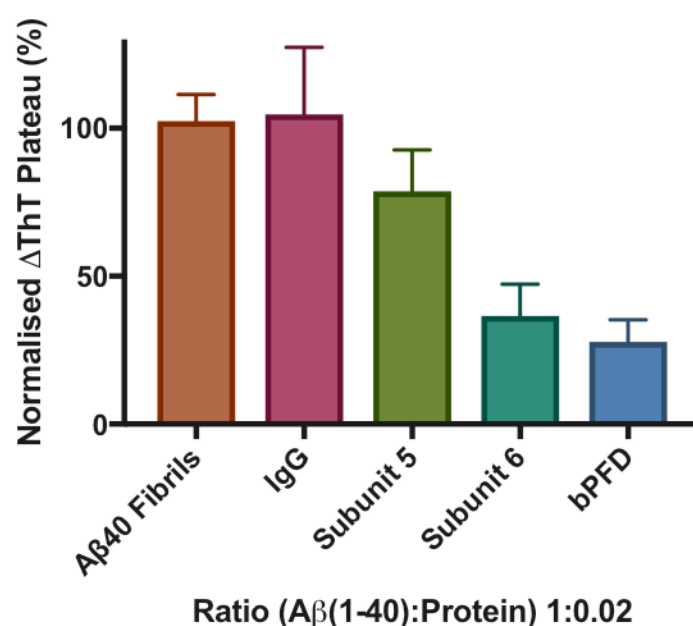


Figure 3.15: Comparison of the  $\Delta$ ThT plateaus of the ratio ( $A\beta(1-40)$ :Protein) 1:0.02 normalised against the plateau of  $A\beta(1-40)$  fibrils formed. The data presented here are the calculated normalised  $\Delta$ ThT mean  $\pm$  SD from 3 independent repeats of the experiment. Post-hoc test following ANOVA 1 way analysis revealed all differences were statistically significant with the exception of column 1 & 2 ( $A\beta$  Fibrils and IgG, data not shown).

To demonstrate the concentration dependent inhibitory effect of *h*PFD subunits 5 and 6 on the fibril formation of  $A\beta(1-40)$ , surrogate rates of fibrillogenesis were formulated and graphed as seen in Figure 3.16. The fibrillogenesis rate is calculated from the linear slope during the elongation phase, and averaged over three independent experiments. This surrogate rate of fibrillogenesis is therefore calculated as the initial rate of fibril formation via the  $\Delta$ ThT (a.u.) over time. Figure 3.16 exposes the difference in chaperone ability the *h*PFD subunits have for the amyloidogenic peptide and the concentration dependence of the rate of fibril formation. At the molar ratio of 1:0.02, the addition of *h*PFD subunit 5 decreases the rate of conversion of the random coil  $A\beta(1-40)$  to structured  $\beta$ -sheets from approximately 110 to approximately 90  $\Delta$ ThT(a.u.)/hr. In comparison, the addition of *h*PFD subunit 6 at the same concentration decreases to approximately 50. When the ratios are increased

to 1:0.1, fibril formation is observed for *h*PF $\Delta$  subunit 5 at an initial rate of approximately 50  $\Delta$ ThT(a.u.)/hr, however the co-incubation of *h*PF $\Delta$  subunit 6 and A $\beta$ (1-40) at this ratio shows no fibril formation. At higher ratios of co-incubation of 1:0.2 and 1:0.5 for both *h*PF $\Delta$  subunits, fibril formation is not observed.

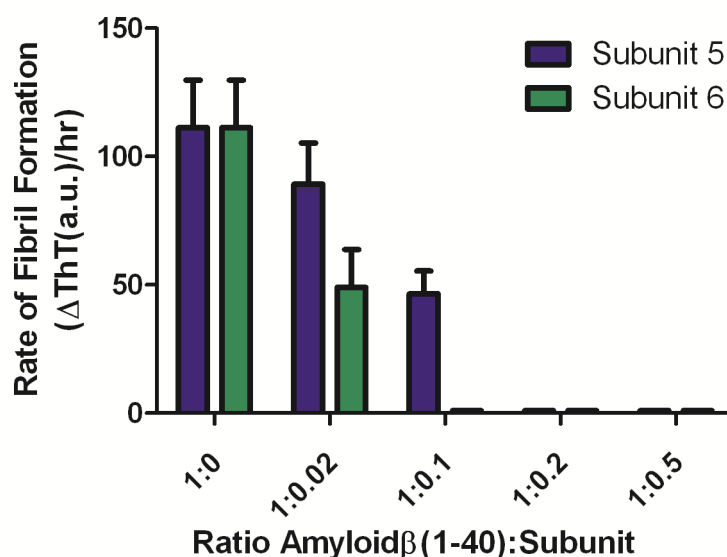


Figure 3.16: Fibril formation rate as measured by  $\Delta$ ThT over time for A $\beta$ (1-40) & varying concentrations of *h*PF $\Delta$  subunits 5 & 6. Rates were extracted from the initial linear elongation phase in the ThT assays. Fibril formation was not observed for co-incubation of A $\beta$ (1-40) & *h*PF $\Delta$  subunit 5 at ratios of 1:0.2, 1:0.5 & A $\beta$ (1-40) & *h*PF $\Delta$  subunit 6 1:0.1, 1:0.2, 1:0.5. The data presented here are the calculated mean from the rate of fibril formation,  $\pm$  SD from 3 independent repeats of the experiment.

The ThT assays reveal that the formation of fibrils by A $\beta$ (1-40) was significantly reduced and in some cases completely abolished with the co-incubation of A $\beta$ (1-40) and *b*PF $\Delta$  or one of the *h*PF $\Delta$  subunits. It is hypothesised that the inhibitors interfere with the fibril formation by interacting with the prefibrillar species, by binding to and redirecting the folding of the protein to afford off-pathway monomers and oligomers. Although the mechanism by which this is achieved is not well understood, it is known that many different types of intermolecular contacts may mediate chaperone-protein interactions, including interactions with exposed aromatic residues and amino groups [326], [327]. It is therefore postulated that the inhibition of A $\beta$ (1-40) fibril formation by *b*PF $\Delta$  and its subunits are mediated through non-specific interactions that take place as the misfolding of the amyloidogenic peptide exposes regions of hydrophobicity that interact with PFD and initiates the remodelling processes. Previous studies of PFD have determined the amphipathic  $\alpha$ -helical tentacles of the jelly fish like structure to be the fundamental component for its chaperone activity [91], [93], [328]. It is

suggested that these  $\alpha$ -helical extensions have regions of hydrophobicity that are normally buried, but are exposed in the vicinity of the misfolding peptide. The hexameric complex encapsulates the misfolding peptide and their hydrophobic regions bind and interact. This ability to recognise segments of exposed hydrophobic amino acid residues, which are normally buried in the interior of the natively folded protein is similar to the chaperoning function of HSP70 [71].

The *in vivo* biological activity of eukaryotic PFD has been found to include chaperoning the misfolded actin to CCT for refolding [95], [329], and it is likely that the chaperone acts in a local manner rather than inducing large conformational changes, as also observed for HSP70 [71], [329]–[331]. The transient release of the bound misfolded peptide is necessary for refolding, and the interactions of the chaperone to the hydrophobic regions of the substrate prevents aggregation and reduces the concentration of free folding intermediates [68]. Since ATP was not included in the ThT assays conducted in this Chapter, it is revealed that the chaperoning mechanism of PFD is independent of energy input from ATP hydrolysis. Chaperones that protect the substrates from aggregation by holding them in their cavities, do not directly facilitate refolding, and are independent of ATP are referred to as ‘holdases’ and hence PFD is identified as a holdase chaperone [332].

As the proposed mechanism of identifying the misfolded substrate is through exposed hydrophobic residues, the ability to identify these normally buried residues for binding is likely to be a contributing factor in the efficiency of chaperoning. Therefore, flexibility in the complex PFD and its substrates may be a crucial factor to the efficiency of chaperoning.

The data in this Chapter shows that the intact complex *b*PFD was found to be the most efficient at inhibiting A $\beta$ (1-40) from fibrillation, followed by *h*PFD subunit 6 and then *h*PFD subunit 5. Recognition of the misfolded protein is a prerequisite for efficient chaperone activity and although flexibility in the regions of hydrophobicity is likely to be a protein characteristic required to efficiently recognise misfolded substrates; it is likely that there are other contributing factors. It can be proposed that intact *b*PFD has a higher chaperoning efficiency than *h*PFD subunit 5 and 6 due to the higher number of flexible regions dedicated to the identification of the misfolded polypeptides. As all 6 subunits construct the intact complex, and it is presumed that each PFD subunit has its own region of misfolding recognition; *b*PFD therefore has 5 more ‘zones’ than *h*PFD subunits 5 and 6 available for identifying and binding to exposed hydrophobic residues of misfolded polypeptides.

The data here shows that *h*PFD subunit 6 is more efficient at inhibiting fibril formation than *h*PFD subunit 5, and this suggests that the  $\beta$  type subunits are better chaperones than  $\alpha$  type structures. As the modelled hexameric PFD structure consists of 4 outer  $\beta$  type subunits and 2 inner  $\alpha$  type

subunits, it is likely that the outer  $\beta$  type structures contribute more to the overall binding and chaperone efficiency.

It is also suggested that cellular temperature can affect the efficiency of biological chaperones such as PFD. A published study of *phPFD* revealed the archaeal chaperone assisted in the correct refolding of denatured lysozyme at temperatures of 20 °C and suppressed refolding of lysozyme at higher temperatures of 40 °C [333]. Although these findings seem to establish chaperone efficiency at lower temperatures, the paper reported that the interactions of the chaperone with the denatured lysozyme at temperatures of 40 °C was significantly stronger than at 20 °C. In this case, refolding of lysozyme to its native conformation was prevented at 40 °C due to the strong interactions between chaperone and substrate resulting in high concentrations of intermediate states. Although not studied in this work, this strongly suggests that temperature could be a key influence on chaperone activity of eukaryotic PFD and it is proposed that the binding interactions of *bPFD*, *hPFD* subunits 5 and 6 may have had an increased chaperoning effect on misfolding A $\beta$ (1-40) at higher temperatures.

Although upregulation of the intact hexameric chaperone has not been observed in studies of thermal aggregation [334], thermal stress studies of *pyrococcus furiosus* (*pf*) showed that the gene encoding the *pfPFD*  $\beta$  subunit was down regulated upon heat shock [335] and studies of *mjPFD* was identified to have 2 putative types of  $\alpha$  subunits with an upregulation of 20-fold for one of the  $\alpha$  subunits following heat shock [86], [336]. In this particular study, it was unclear whether the subunit could function as a homo-oligomer or needed to interact with the  $\beta$  subunit to assist in protein refolding. However, the opposing regulation of  $\alpha$  and  $\beta$  archaeal PFD subunits in response to heat shock suggests the subunits have different roles in response to elevated cell temperatures and this may be reflected in the results obtained here. As *hPFD*  $\beta$  subunit 6 was found to be a more efficient chaperone than *hPFD*  $\alpha$  subunit 5 at temperatures of physiological conditions, in conditions of thermal stress the upregulation of  $\alpha$  type subunits may occur so that the overall number and function of  $\alpha$  type subunits match the efficiency observed by  $\beta$  type subunits.

Although there are no eukaryotic studies that detail upregulation of one type of PFD subunit occurring *in vivo*, knockdown studies of PFD subunits 2 and 5 in mice revealed an overall reduction in expression of the other remaining subunits with differences in quantities observed between the subunits [102], [316]. This supports the notion of the individual *hPFD* subunits having separate functions *in vivo* and being regulated differently in response to heat shock. A study that has observed upregulation of the intact hexameric complex found that the protein concentration increased when levels of A $\beta$  in mice were increased [82]. Our studies undoubtedly prove the chaperoning efficiency of *bPFD* and the fibril inhibition effect it has on A $\beta$ . Therefore, the

combination of these results proves the importance of PFD in the accumulation of misfolded A $\beta$  and the possible potential for therapeutics in this signalling pathway.

#### 3.2.4.3: TEM analysis of the incubated A $\beta$ (1-40), *b*PFD and *h*PFD Subunits 5 and 6

To further verify the presence of amyloid fibrils, TEM images were collected at the conclusion of the ThT assays (Figure 3.17). Samples were also further incubated for a week to determine any long term effects of the fibril forming conditions and TEM images were again collected, however no difference was observed. In samples containing aggregated A $\beta$ (1-40), the fibrils observed were in high abundance, with long stretches reaching to approximately 5  $\mu$ m in length and many were highly branched. In comparison, the incubated samples of A $\beta$ (1-40): *b*PFD at a molar ratio of 1:0.02, reveal a significant reduction in the number of fibrils formed by A $\beta$ (1-40) (Figures 3.17B&C). Morphologically, these fibrils resembled the rod-like structures formed by A $\beta$ (1-40) alone (Figure 3.17) although they were shorter, reaching approximately 1  $\mu$ m in length and were observed in the presence of small rounded aggregates (Figure 3.17B&C). Interestingly, the fibrils seen in the incubated samples of A $\beta$ (1-40):subunit 5 at a molar ratio of 1:0.02 (Figure 3.17D), were found to exhibit similar morphology to the fibrils of the co-incubation of A $\beta$ (1-40): *b*PFD. Although higher abundance of fibrils was seen in the TEM samples of A $\beta$ (1-40) and *h*PFD subunit 5, these fibrils were also found to reach approximately 1  $\mu$ m in length and observed in the presence of aggregates (Figure 3.17D). Similar images were also recorded for incubation of A $\beta$ (1-40) and *h*PFD subunit 6 (Figure 3.17E) at the same concentration. Although the appearance of the small aggregates and fibrils were similar to the co-incubated samples of A $\beta$ (1-40) with either *b*PFD or *h*PFD subunit 5, fibrils were found to reach lengths of approximately 3  $\mu$ m. The abundance of fibrils was observed to be less than those seen for A $\beta$ (1-40):subunit 5, although higher than the amount of fibrils upon co-incubation of A $\beta$ (1-40): *b*PFD. No fibrils were observed following incubation of A $\beta$ (1-40) and *h*PFD subunits at high concentrations (Figure 3.17F), consistent with the results obtained from the ThT assays, further validating the chaperone activity of the *b*PFD and *h*PFD subunits.

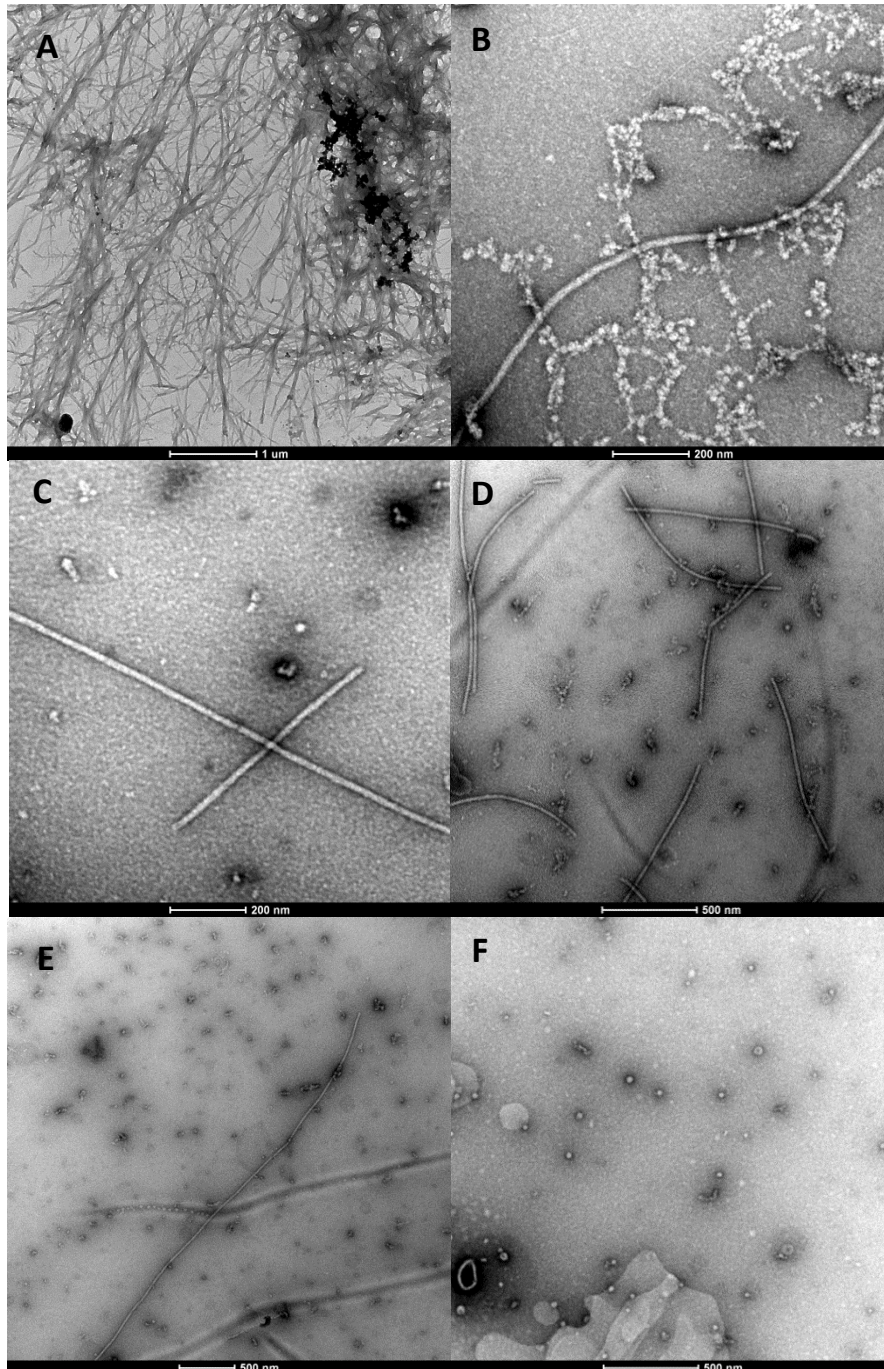


Figure 3.17: TEM images of Aβ(1-40) peptide following ThT assays. (A) Fibrils formed from peptide Aβ(1-40) (50 μM). (B&C) Aβ(1-40) (50 μM) incubated in the presence of bPFD (1 μM). (D) Fibrils formed from peptide Aβ(1-40) (50 μM) in the presence of hPFD subunit 5 (1 μM). (E) Fibrils formed from peptide Aβ(1-40) (50 μM) in the presence of hPFD subunit 6 (1 μM). (F) Aβ(1-40) (50 μM) incubated in the presence of hPFD subunit 6 (25 μM). Scale is as indicated in figures.

The morphology of the fibrils following the co-incubation of the amyloidogenic peptide and bPFD, hPFD subunit 5 or hPFD subunit 6 is revealed in the TEM images of Figure 3.17B-F. Although it has been previously postulated that the chaperone interacts with the prefibrillar peptide to prevent fibril formation, it is curious to see by TEM that the fibrils formed in the co-incubations are short in length



and do not reach spans seen in the incubation of A $\beta$ (1-40) alone. This raises the possibility of chaperone interaction at the regions of hydrophobicity present at the ends of the fibrils. It is possible that the amphipathic coils of the chaperone are attracted to the exposed hydrophobic residues of the fibril end whilst simultaneously attracted to the exposed hydrophobic residues of the misfolded prefibrillar species. The concept that these chaperones may bind to fibrils at the ends or along the surface to prevent further elongation aligns with the recent work seen for sHsps [337]–[339].

It was previously proposed that the biological mechanism of eukaryotic PFD was through a bind and release cycle to guide misfolded amyloidogenic peptides to the correct form. It is postulated here that the chaperone also has the ability to bind to the end of fibrils through the same local bind and release mechanism. However, as it has been found that PFD is unable to disrupt preformed fibrils, it is likely that the chaperone competes with free form misfolded peptides to bind to the fibril, therefore limiting further extension of the fibril.

#### 3.2.4.4: Cytotoxicity assays of A $\beta$ (1-40) and *h*PFD Subunits 5 and 6

Many previous studies have demonstrated that the fibrillar form of proteins display enhanced cytotoxicity relative to the un-aggregated form [49], [50]. However, more recent studies have emerged to reveal that the soluble oligomeric species preceding fibril formation are the likely cytotoxic species [47], [51], [57], [340].

To compare the cytotoxicity of the soluble A $\beta$ (1-40) monomer and the structured  $\beta$ -sheet fibrils, both were added at a final concentration of 1  $\mu$ M to cultured pheochromocytoma-12 (PC12) cells, a neuronal cell line that is widely used to study amyloid toxicity. To determine if *h*PFD subunits also extended their chaperone activity to prevent the toxicity of A $\beta$ (1-40) to cells, the pre-incubated samples from ThT assays were diluted with PBS and added to the PC-12 cells at a final peptide concentration of 1  $\mu$ M. As ThT was present in these samples, cells were also treated with a ThT-containing control and independent *h*PFD subunit 5 or subunit 6 to see if these components had any inherent cytotoxicity, however no toxicity was observed in these experiments. Treated cells were incubated for 48 hours at 37 °C and cell survival was assessed by an MTT assay. Healthy cells produce succinate dehydrogenase, a mitochondrial enzyme which has ability to reduce the MTT to formazan, a purple substance. Cell viability is therefore quantified by measuring the mitochondrial activity through the absorbance of formazan at 570 nm [341]. The cell viability assays were normalised against a control of buffer treated PC12 cells. Values 3 standard deviations away from the median were treated as outliers and excluded from the overall calculation due to possible external contamination. Cytotoxicity assays with the intact *b*PFD complex were not included in this



study to preserve the material for structural characterisation. It should also be noted that previous studies have demonstrated that the fibrillar species are maintained in cell culture media and fibril formation proceeds at a reduced rate [341], however the *h*PFD subunits are still able to chaperone the misfolding substrates.

Figure 3.18 demonstrates the relative cellular toxicity of the different forms of A $\beta$ (1-40) and the influence of *h*PFD subunit 5. Cells treated with monomeric A $\beta$ (1-40) exhibited the highest cytotoxicity with a decrease in viability of approximately 50%. Pre-incubated, fibrillar A $\beta$ (1-40) was found to have the lowest cytotoxicity with approximately 80% cell viability. Cells treated with the pre-incubated mixture of A $\beta$ (1-40) and varying concentrations of *h*PFD subunit 5 exhibited cell viability values between those of the fibrillar A $\beta$ (1-40) and monomeric A $\beta$ (1-40).

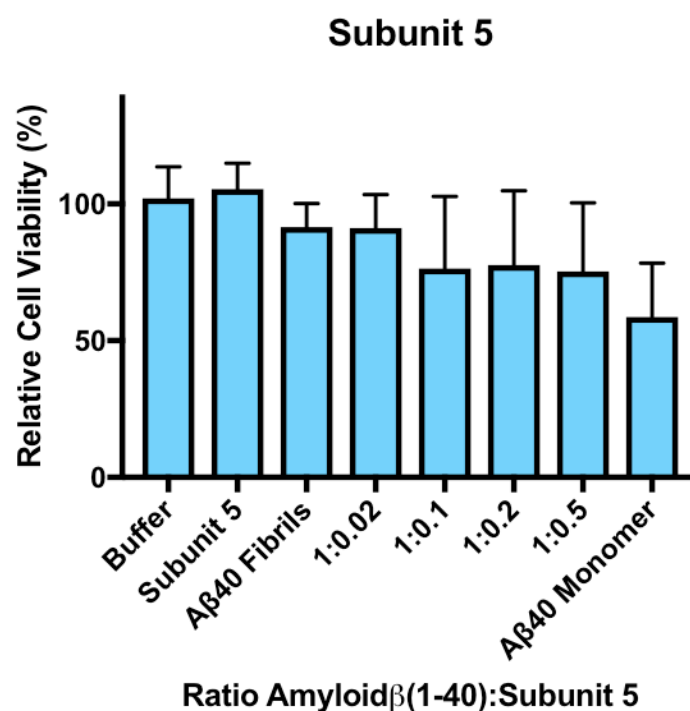


Figure 3.18: Cytotoxicity studies of A $\beta$ (1-40) in the absence & presence of varying concentrations of *h*PFD subunit 5. Mitochondrial activity was measured through the absorbance of formazan at 570 nm and were then normalised against the buffer control. Samples containing healthy cells (PBS buffer) and samples of A $\beta$ (1-40) & *h*PFD subunit 5 from the previous ThT assays were diluted with PBS. The diluted samples of Incubated A $\beta$ (1-40) (1  $\mu$ M final concentration) in the absence or presence of *h*PFD subunit 5 (0-0.5  $\mu$ M final concentration), & monomeric A $\beta$ (1-40) were added to cell culture media of PC12 cells, the cells incubated at 37 °C for 48 hr & cell viability was determined by an MTT assay. The data presented here are the calculated mean  $\pm$  SD from 3 independent repeats of the experiment normalised against the buffer control. Statistical significance between conditions is summarised in Appendix F.

To ascertain if the difference observed between groups of cell viability values were significant, each set of data within the cytotoxicity assay was subjected to a post-hoc test following a one-way Analysis of variance (ANOVA). Post-hoc following an ANOVA test compares pairwise differences among independent variable groups, based on the assumption that the populations are Gaussian [342], [343]. A small P value suggests that the variances are different and data sets are determined to be significantly different when  $P < 0.05$ . Post-hoc tests after a one-way ANOVA analysis of the cytotoxicity studies of A $\beta$ (1-40) co-incubated with different ratios of hPFD subunit 5 are summarised in Appendices F and G. MTT assays for the individual hPFD subunits independently reveal no statistical difference in cell viability relative to cells incubated with PBS buffer controls (Appendix F and G) consequently revealing no inherent cytotoxicity exhibited by the chaperone subunits.

The cell viability assays for A $\beta$ (1-40) fibrils co-incubation with A $\beta$ (1-40) and hPFD subunit 5 at ratios of 1:0.1, 1:0.2, and 1:0.5 show no statistical difference, but they are different to monomeric A $\beta$ (1-40). This suggests that hPFD subunit 5 is able to chaperone A $\beta$ (1-40) into a product that is less toxic than monomeric A $\beta$ (1-40), however, only to the same extent of cytotoxicity displayed by fibrillar A $\beta$ (1-40).

Addition of pre-incubated hPFD subunit 6 and A $\beta$ (1-40) to PC12 cells reveal a dose-dependent decrease in cell viability as observed in Figure 3.19. The cell viability assays of A $\beta$ (1-40):hPFD subunit 6 again exhibited cytotoxicity to a greater extent than fibrillar A $\beta$ (1-40) and a lesser extent than monomeric A $\beta$ (1-40).

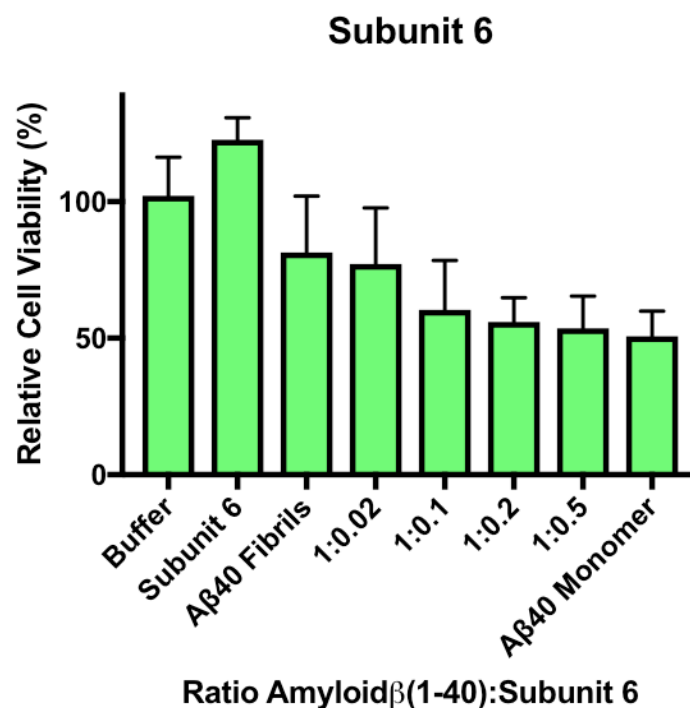


Figure 3.19: Cytotoxicity studies of A $\beta$ (1-40) in the absence & presence of varying concentrations of hPFD subunit 6. Mitochondrial activity was measured through the absorbance of formazan at 570 nm and were then normalised against the buffer control. Samples containing healthy cells (PBS buffer) and samples of A $\beta$ (1-40) & hPFD subunit 6 from the previous ThT assays were diluted with PBS. The diluted samples of incubated A $\beta$ (1-40) (1  $\mu$ M final concentration) in the absence or presence of hPFD subunit 6 (0-0.5  $\mu$ M final concentration), & monomeric A $\beta$ (1-40) were added to cell culture media of PC12 cells, the cells incubated at 37 °C for 48 hr & cell viability was determined by an MTT assay. The data presented here are the calculated mean  $\pm$  SD from 3 independent repeats of the experiment normalised against the buffer control. Statistical significance between conditions is summarised in Appendix G.

Similar to the cytotoxicity assays for A $\beta$ (1-40) and hPFD subunit 5, the cellular viability data sets of A $\beta$ (1-40) and hPFD subunit 6 were subjected to one-way ANOVA analysis to determine statistical significances between data sets. The results are summarised in Appendix G.

Figure 3.19 and Appendix G reveals that there is no statistical difference between the cytotoxicity of A $\beta$ (1-40) fibrils and the co-incubation of A $\beta$ (1-40) and hPFD subunit 6 at a ratio of 1:0.02. It is also observed that the cytotoxicity of cells incubated with A $\beta$ (1-40) is not statistically different between the co-incubation of the amyloid peptide and hPFD subunit 6 at ratios of 1:0.1, 1:0.2, 1:0.5, however these do show a significant difference to the cells incubated with A $\beta$ (1-40) fibrils.

It is seen in Figure 3.18 and 3.19 that the PC12 cells incubated with fibrillar A $\beta$  exhibit fluorescence at a slightly smaller degree than those incubated with buffer or independent protein. It was previously stated in Chapter 2 that MTT assays quantify the metabolic activity of cells and are not direct representations of cell death. Due to the plethora of studies that suggest the decreased toxicity and possible null effect of fibrils [41], [47], [340], it is proposed that the decrease in

fluorescence observed by fibrils and the products of A $\beta$ (1-40) and PFD subunits at low concentrations is mediated through a reduction in cellular proliferation. It is possible that the fibrils present in these samples affect the rate of cell mitosis rather than inducing cytotoxicity.

From the MTT assays, it is also apparent that the co-incubation of the cells with A $\beta$ (1-40) with either *h*PFD subunit 5 and 6 at higher concentrations, induces cytotoxicity between the ranges observed by fibrillar and monomeric A $\beta$ (1-40). In Section 3.2.4.1, it was suggested that PFD is a holdase chaperone; inhibiting aggregation of A $\beta$ (1-40) through the encapsulation of the misfolding peptide and refolding of the substrate is initiated through cycles of chaperone interaction and release. This indirect chaperone participation is suggested to be a possible mechanism to explain the data seen here, specifically the results observed for the co-incubation of *h*PFD subunit 6 and A $\beta$ (1-40). It is possible that the *h*PFD subunits are able to chaperone and 'hold' the substrate at a particular intermediate product that is cytotoxic to cultured PC12 cells. It is suggested that this intermediate product has a high attraction for the cellular membrane and as the *h*PFD subunits release the intermediate from its cavity to initial refolding, the transient release allows membrane intercalation resulting in cellular death.

To investigate the mechanism of cytotoxicity, microscopic digital photos were taken before and after the incubation of PC-12 cells with the products of A $\beta$ (1-40). Figure 3.20A reveals a well filled with healthy cells taken before the addition of monomeric A $\beta$ (1-40). Colourless, large cells with high fluid content are indicative of healthy cells with the ability to divide and release NAD(P)H-dependent cellular oxidoreductase enzymes. In comparison, Figure 3.20B was taken following the incubation of monomeric A $\beta$ (1-40). Small rounded cells and black inclusion bodies within the cytosol are indicative of unhealthy dying cells. Microscopic views of the cells incubated with A $\beta$ (1-40) and *h*PFD subunit 5 or 6 revealed the same small rounded cells and black inclusion bodies within the cytosol (data not shown).

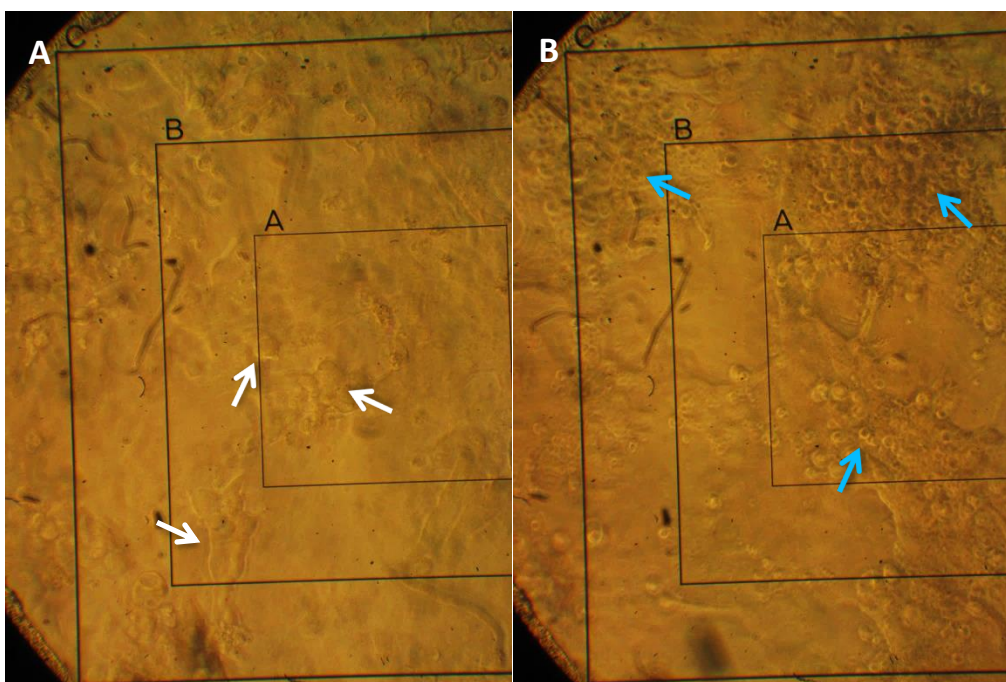
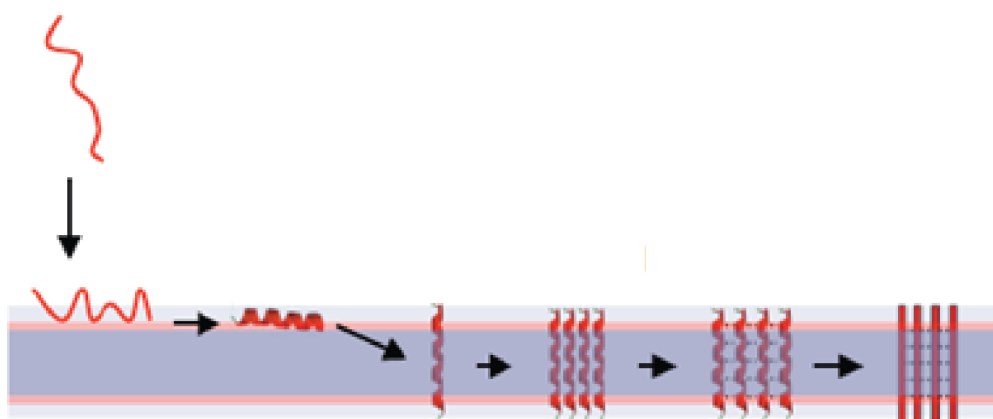


Figure 3.20: Digital photographs of (A) healthy cells before incubation with monomeric A $\beta$ (1-40) & (B) unhealthy cells following incubation of taken through a microscope lens viewing the cells in a well from a 96 well plate. Fat, large, transparent cells are indicative of healthy cells as indicated by the white arrows & small, nodular cells with black inclusion bodies are indicative of necrosis & unhealthy cells as indicated by the blue arrows.

As the cells incubated with A $\beta$ (1-40) and *h*PF $\beta$  subunit 5 or 6 at high ratios revealed the same morphology as the cells incubated with monomeric A $\beta$ (1-40), this suggests that the toxic mechanisms exhibited by the chaperoned peptide and monomeric A $\beta$  are similar. Unfortunately the cytotoxicity mechanism exhibited by the products of the co-incubation of the chaperone to A $\beta$  has yet to be determined however, several plausible suggestions are presented here. A mechanism thought to occur for monomeric A $\beta$  that is a possible scenario for the products of the co-incubation of the *h*PF $\beta$  subunit and A $\beta$ (1-40) is the intercalation of the substrates into the lipid membrane. As mentioned earlier, PF $\beta$  is likely to chaperone A $\beta$  through bind and release cycles. Unfortunately the transient release of the misfolded A $\beta$ (1-40) to allow for correct refolding also allows the substrate to be electrostatically attracted to the negatively charged membrane of the cellular surface as seen in Figure 3.21. It is possible that the refolding peptide embeds itself into the membrane to initiate the folding of amphipathic  $\alpha$ -helices. This is followed by the insertion of the peptide into the membrane at an axis perpendicular to the bilayer and further accumulation of substrates into the lipid bilayer allows dimerisation to complete traversal of the lipid bilayer membrane. It is possible that the anionic phospholipids of the membrane are then reorientated. The embedded  $\alpha$ -helical A $\beta$  substrates may congregate and align in the lipid bilayer as seen in Figure 3.21, where it is possible that they interact to undergo a structural shift to form  $\beta$ -sheet conformations and  $\beta$ -sheet rich

oligomers introducing leakages into the cell membrane [344]–[346]. It is also possible that the oligomerisation of the A $\beta$  substrates within the membrane can result in a linear transmembrane structure that occurs when the hydrophobic residues of the amphipathic A $\beta$  peptides interact with the lipid bilayer to force the hydrophilic face of the membrane to curve [340], [347]. This possible intercalation mechanism of the products of the co-incubation of A $\beta$ (1-40) and the *h*PFD subunits results in an increase in permeability of the cell membrane resulting in cytotoxicity through cellular chemical imbalances [348], [349].



**Figure 3.21: Schematic diagram showing peptide binding to a lipid bilayer, followed by insertion & oligomerisation.**  
Adapted from [344].

The co-incubation of PFD or its subunits with monomeric A $\beta$ (1-40) is proposed to result in intermediate structures of monomeric A $\beta$ (1-40), however it is also likely that amyloid oligomers are produced from the chaperoning interactions of PFD to the amyloid peptide as misfolded A $\beta$  is released from PFD to permit refolding. Cellular death induced by A $\beta$  oligomers has been thoroughly explored and is proposed to be mediated through their ability to bind to multiple receptors eventually resulting in the disruption of various signaling pathways. This lack of receptor specificity is reportedly due to the oligomers 'sticky' characteristic that allows it to latch onto receptors. Figure 3.22 illustrates the receptors thought to be susceptible to amyloid oligomers. This includes nerve growth factor (NGF) receptors, prion protein (PrPC) receptors, N-methyl-D-aspartate receptors (NMDAR), insulin receptors, and Frizzled (Fz) receptors [51], [340], [350].

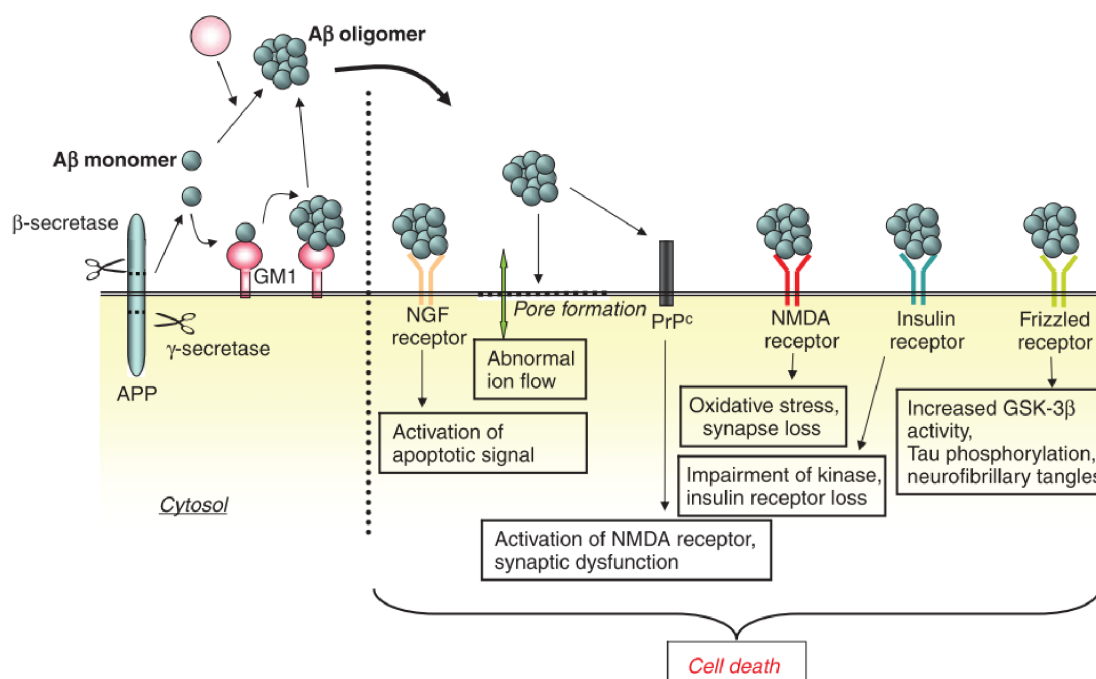


Figure 3.22: Schematic diagram showing the mechanism of Aβ oligomers on several different receptors. The 'sticky' property of the molecule results in extracellular receptor binding. Reproduced with permission from [51].

Although the mechanism of PFD and the effects of its products are yet to be fully elucidated, work described in this Chapter has been able to give an insight into the inhibitory effects of *b*PFD and its subunits towards fibril formation by Aβ(1-40). It was found that intact *b*PFD has a high efficiency for inhibiting the formation of fibrils at substoichiometric concentrations. Higher concentrations of *h*PFD subunits 5 and 6 were needed to effectively inhibit the formation of fibrils, with *h*PFD subunit 6 exhibiting better chaperoning abilities than *h*PFD subunit 5 at the same concentrations. This was reflected in the cell viability assays, where the cytotoxicity of the products from the co-incubation of *h*PFD subunit 5 and Aβ were closer to the viability observed by fibrillar Aβ and the co-incubation products of *h*PFD subunit 6 and Aβ were closer to the viability expressed of the monomeric Aβ.

### 3.3: Experimental

All materials were purchased from Sigma Aldrich (Missouri, United States) unless otherwise stated. FPLC columns were purchased from GE Healthcare (Little Chalfont, United Kingdom). Luria broth-ampicillin agar plates (containing 100  $\mu$ L of ampicillin stock (100 mg/mL) per 100 mL agar), ampicillin stock (100 mg/mL) and luria broth (LB: 10 g tryptone, 5 g yeast extract, 5 g NaCl, 1 L H<sub>2</sub>O) were prepared in house using items purchased from Merck Millipore (Billerica, Massachusetts, United States).

#### 3.3.1: BLAST Needleman-Wunsch Global Alignment

BLAST Needleman-Wunsch global alignments were computed through the National Centre for Biotechnology Information in the U.S. National Library of Medicine in the National Institute of Health Parameters (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) for the algorithm Existence 11, Extension: 1. An open source available for the generation of random sequences (<http://random-ize.com/word-scrambler/smb.php>)

#### 3.3.2: Extraction of bPFD

Bovine testes (500-600 g) were obtained from Austral Meats Butcher (South Australia). The skin of the testes was stripped and the remaining flesh cut into small pieces and then blended with a handheld blender until the meat had a smooth consistency. Cold homogenization buffer (20 mM Tris-HCl, pH 7.2, 10 mM KCl, 5 mM MgCl<sub>2</sub>, 1 mM EGTA, 1 mM PMSF, cocktail inhibitor tablet (1 mg/mL, Roche, Switzerland), 1 mM DTT) was added to the blended tissue at a 1 mL/g ratio and this was further blended at full speed for 5 minutes. The solution was centrifuged (4000 *g*) for 30 minutes and the pellet discarded. The supernatant was then further ultracentrifuged (80,000 *g*) for an hour. The supernatant was decanted and filtered under vacuum with a double layer of Whatman No. 1 filter paper. The solution was then cooled to 0 °C on ice and (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> was slowly added over 30 minutes until 20% saturation (117 g/L) was reached. This was stirred on ice for a further 30 minutes. The precipitate was collected by centrifugation (4000 *g*) for 10 minutes. Precipitation of the remaining supernatant and collection of the precipitate was repeated at saturations of 40%, 60% and 80% (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>. Samples of the (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> pellets were then subjected to shotgun proteomics analysis as described in Section 3.3.4.

The 40-60 % (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> pellet was dissolved by soft swirling into a NaH<sub>2</sub>PO<sub>4</sub> buffer (10 mM NaH<sub>2</sub>PO<sub>4</sub>, 1 mM EGTA, 1 mM DTT, 200 mL, pH 7.2) and this solution was then dialysed with this NaH<sub>2</sub>PO<sub>4</sub>



buffer using a snakeskin membrane (10 kDa MW cut off, ThermoFisher, Massachusetts) overnight with 3 buffer changes. The dialysed extract was centrifuged (7,000 *g*) to remove precipitate and the supernatant was syringe filtered (0.2  $\mu$ m, GE Healthcare, Little Chalfont, United Kingdom). Using an automated FPLC system (GE Healthcare, Little Chalfont, United Kingdom), this was then loaded onto a pre-equilibrated Q-Sepharose HP 5/25 column (10 mM NaH<sub>2</sub>PO<sub>4</sub>, 1 mM EGTA, 1 mM DTT, pH 7.2). The column was washed with this buffer at a rate of 10 mL/min until the UV absorbance plateaued. The protein was then eluted with a step gradient using NaH<sub>2</sub>PO<sub>4</sub> buffers (100, 200, 300, 400 mM NaH<sub>2</sub>PO<sub>4</sub>, 1 mM EGTA, 1 mM DTT, pH 7.2) at a rate of 10 mL/min. Fractions (100 mL) were collected and samples of each fraction were analysed by shotgun proteomics. Fractions containing *b*PFD subunits were combined, concentrated and buffer exchanged (20 mM Tris-HCl, 1 M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 1 mM MgCl<sub>2</sub>, 1 mM EGTA, 1 mM DTT, pH 7.2) using an Amicon ultrafiltration cell fitted with a membrane (10 kDa, Merck, Massachusetts) to approximately 5 mL.

The concentrated solution containing the *b*PFD extracts was then loaded onto a pre-equilibrated phenyl Superose HP G20 column (20 mM Tris-HCl, 1 M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 1 mM MgCl<sub>2</sub>, 1 mM EGTA, 1 mM DTT, pH 7.2). The column was washed with this buffer at a flow rate of 2.5 mL/min until the UV absorbance plateaued and protein was then eluted with a linear gradient (100 mL) (20 mM Tris-HCl, 1 mM MgCl<sub>2</sub>, 1 mM EGTA, 1 mM DTT, pH 7.2) at a flow rate of 2.5 mL/min. Fractions (5 mL) were collected and analysed by shotgun proteomics. Fractions containing *b*PFD subunits were combined, concentrated and buffer exchanged (20 mM MES, 20 mM NaCl, 2 mM MgCl<sub>2</sub>, 1 mM EGTA, 1 mM DTT, pH 6.0) using an Amicon ultrafiltration cell fitted with a membrane (10 kDa, Merck, Massachusetts) to approximately 2 mL.

The concentrated solution containing the *b*PFD extracts was loaded onto a pre-equilibrated mono S column (20 mM MES, 20 mM NaCl, 2 mM MgCl<sub>2</sub>, 1 mM EGTA, 1 mM DTT, pH 6.0). The column was washed with this buffer at a flow rate of 1 mL/min until the UV absorbance plateaued and protein was then eluted with a linear gradient (25 mL) at a flow rate of 0.5 mL/min with a high salt buffer (20 mM MES, 250 mM NaCl, 2 mM MgCl<sub>2</sub>, 1 mM EGTA, 1 mM DTT, pH 6.0). Fractions (1 mL) were collected and analysed by Western blotting. The washings were found to contain *b*PFD subunits and were therefore combined and concentrated using an Amicon ultrafiltration cell fitted with a membrane (10 kDa, Merck, Massachusetts) to approximately 1 mL.

The concentrated solution containing the *b*PFD extracts was loaded onto a pre-equilibrated ceramic hydroxylapatite type II column (20 mM NaH<sub>2</sub>PO<sub>4</sub>, 20 mM KCl, 1 mM MgCl<sub>2</sub>, 0.1 mM CaCl<sub>2</sub>, 1 mM DTT, pH 6.8). The protein was eluted into a buffer of sodium phosphate (250 mM NaH<sub>2</sub>PO<sub>4</sub>, 20 mM KCl, 1 mM MgCl<sub>2</sub>, 0.1 mM CaCl<sub>2</sub>, 1 mM DTT, pH 6.8) at a flow rate of 1 mL/min. Fractions (1 mL) were

collected and analysed by electrophoresis PAGE and Western blotting. Fractions containing PFD were combined and concentrated using an Amicon ultrafiltration cell fitted with a membrane (10 kDa, Merck, Massachusetts) to approximately 500  $\mu$ L.

The concentrated solution containing the prefoldin extracts was loaded onto a pre-equilibrated Superdex 200 column (100 mM AA, pH 7.4). The protein was eluted into a buffer of ammonium acetate (100 mM AA, pH 7.4) at a flow rate of 0.4 mL/min. Fractions (1 mL) were collected and analysed by electrophoresis PAGE, Western blotting and proteomics. Fractions containing prefoldin were combined and concentrated with an Amicon ultrafiltration cell fitted with a membrane (10 kDa, Merck, Massachusetts) to approximately 300  $\mu$ L.

### 3.3.3: *h*PFD Subunit Expression

*h*PFD subunits were expressed as a fusion protein with a His-tag using the pPROX HTb bacterial expression vector (Aldevron). A 6 x His tag and a TEV protease site was included between the His-tag and prefoldin subunit sequences. The oligonucleotides encoding the PFD subunit sequences and the translated amino acid sequences are shown in Appendix D. The cDNA sequence corresponding to *h*PFD subunits 1-6 were cloned into the restriction sites BamHI (345bp) and NotI (388bp). Upon successful sequencing and cloning of the inserted vector, the plasmid was transformed into *Escherichia coli* BL21(DE3).

Transformation of the DNA encoded plasmids was performed using an aliquot of *E. coli* cells (BL21(DE3)). The competent cells (50  $\mu$ L) and the plasmid DNA (pPROX HTb) (1  $\mu$ L) were mixed together and incubated on ice for 20 minutes. The temperature of the transformation mixture was then rapidly raised to 42 °C for 30 seconds in a water bath and then cooled on ice for 2 minutes. The transformation mixture was added to LB broth and these cells were grown with shaking (225 rpm) for an hour at 37 °C. The cells were then pelleted (1000 g) for a minute and resuspended in LB broth (200  $\mu$ L). This was then spread over a LB-amp agar plate and cells were left to grow overnight at 37 °C. The next day a single colony was chosen from the plate and was inoculated into LB broth (2 mL) with ampicillin (2  $\mu$ L) and cells were grown with shaking (250 rpm) for 3 hours at 37 °C. When the OD<sub>600</sub> reading reached 0.8-1.0, the inoculum (2 mL) was added to LB broth (10 mL) and ampicillin (10  $\mu$ L) with shaking (250 rpm) for 1-2 hours at 37 °C until OD<sub>600</sub> reached 1. 80% glycerol (25 mL) was added to prepared culture and was slowly mixed. This solution was then aliquoted, flash frozen and stored at -80 °C until required for overexpression.

The overexpression of cells was performed using a thawed aliquot of transformed DNA (1  $\mu$ L) and this was streaked onto an LB-amp agar plate using a sterile metal hoop. The cells were left to grow

overnight at 37 °C. The next day a single colony was chosen from the plate and was inoculated into LB broth (50 mL) containing ampicillin (50 µL), and cells were grown with shaking (200 rpm) overnight at 37 °C. The next morning an aliquot of the overnight culture (10 mL) was inoculated into LB broth (500 mL) containing ampicillin (500 µL) and grown with shaking (200 rpm) at 37 °C in bunged conical flasks for approximately 3 hours until OD<sub>600</sub> reached 0.6-0.8. IPTG (250 µL, 1M) was added to the medium and the culture continued to grow with shaking for another 5 hours. Cells were then harvested by centrifugation (5000 g) for 10 minutes. The supernatant was discarded and the pellets were suspended in lysis buffer (50 mM Tris, 250 mM NaCl, 1 mM EDTA, 1 mM PMSF, 0.1 % triton X 100 and protease inhibitor cocktail tablet (1 mg/mL, Roche, Switzerland)). Lysozyme (0.1 mg/mL) was added to the solution and this was then agitated on ice for an hour. The solution was then frozen in a cold mixture of acetone and dry ice for 15 minutes. The frozen mixture was then thawed in lukewarm water until completely melted (approximately 20 minutes). This freeze thaw procedure was repeated for a total of 4 cycles. Deoxyribonuclease (0.1 µg/mL) was added to the viscous solution and the mixture was agitated on ice for 20 minutes. The solution was then centrifuged (7000 g) and the supernatant was dialysed into an imidazole washing solution (20 mM NaH<sub>2</sub>PO<sub>4</sub>, 250 mM NaCl, 30 mM imidazole) overnight with 3 changes of buffer using a snakeskin membrane (10 kDa MW cut off, ThermoFisher, Massachusetts). The supernatant was then syringe filtered (0.2 µm, GE Healthcare, Little Chalfont, United Kingdom) and loaded onto a pre-equilibrated nickel column at a flow rate of 4 mL/min. The column was washed (20 mM NaH<sub>2</sub>PO<sub>4</sub>, 250 mM NaCl, 30 mM imidazole, approximately with 200 mL) at a flow rate of 4 mL/min until the UV reading plateaued. The bound proteins were then eluted from the column with a buffer high in imidazole content (20 mM NaH<sub>2</sub>PO<sub>4</sub>, 250 mM NaCl, 1 M imidazole) and fractions (5 mL) were collected. Fractions containing protein visualised by UV were combined and concentrated with an Amicon ultrafiltration cell fitted with a membrane (10 kDa MW, Merck, Massachusetts) until the volume was approximately 0.5 mL.

The concentrated solution containing the His-tagged subunits was loaded onto a pre-equilibrated Superdex 200 column (100 mM AA, pH 7.4 or PBS, pH 7.4). The protein was eluted into a buffer of PBS (pH 7.4) at a flow rate of 0.4 mL/min. Fractions (1 mL) were collected and fractions containing the hPFD subunits of interest were combined and concentrated with an Amicon ultrafiltration cell (10 kDa MW, Merck, Massachusetts).

### 3.3.4: Shotgun Proteomics Analysis of Protein Fractions

Approximately 10 µg of protein was removed from the sample (as measured through a Bradford coomassie assay (ThermoFisher Scientific, Massachusetts)). The sample was prepared in a solution of

urea (100  $\mu$ L, 7 M) and  $\text{NH}_4\text{HCO}_3$  (100 mM). DTT (10  $\mu$ L, 1 M) was added, briefly vortexed, briefly centrifuged and incubated at room temperature for an hour. IAA (10  $\mu$ L, 1 M) was added, briefly vortexed, briefly centrifuged and incubated at room temperature in darkness for 15 minutes. The sample was diluted with  $\text{NH}_4\text{HCO}_3$  (600  $\mu$ L, 100 mM) until the urea concentration was approximately 1 M. Trypsin (1  $\mu$ L, 200 ng/ $\mu$ L) was added at a 1:50 ratio. The sample was briefly vortexed, briefly centrifuged and then incubated at 37 °C overnight. Following incubation, the sample was cleaned using a C18 reverse phase column (ThermoFisher Scientific, Massachusetts) and solvents removed by SpeedVac (Savant, Farmingdale, NY). The sample was reconstituted into water and 0.1 % FA and run through a LC-MS/MS system, a nanoLC 3000 RSLC coupled to LTQ Orbitrap XL ETD hybrid mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA).

### 3.3.5: Electrophoresis and Staining

15 % SDS lower resolving gels were cast by adding acrylamide (5 mL, 30 %),  $\text{H}_2\text{O}$  (2.5 mL), resolving gel stock (2.5 mL, Stock: 1 M Tris base, 0.4 % SDS), APS (50  $\mu$ L, 10 %) and TEMED (10  $\mu$ L). This was mixed thoroughly and poured into clean casting plates. The gel was levelled with excess  $\text{H}_2\text{O}$  (approximately 1 mL). This was left at room temperature for 30 minutes to set. The water was removed and the stacking gel was made by the addition of acrylamide (560  $\mu$ L, 30 %),  $\text{H}_2\text{O}$  (3.2 mL), stacking gel stock (1.25 mL, Stock: 0.5 M Tris base, 0.4 % SDS), APS (20  $\mu$ L, 10 %) and TEMED (10  $\mu$ L). This was poured on top of the lower resolving gel and a gel comb was inserted into the stacking gel. This was left at room temperature for 30 minutes to set.

SDS (1 g), glycerol (1 mL), Tris HCl (6.25 mL, 0.5 M, pH 6.8),  $\beta$ -mercapto-ethanol (100  $\mu$ L), bromophenol blue (0.05 %) and MQ water was added to make 10 mL of 5x stock loading buffer. Tris base (15.1 g), glycerine (72 g), SDS (5 g) and MQ water were mixed to make 1 L of running buffer.

Electrophoresis apparatus was set up with 15 % SDS PAGE gels. Proteins subjected to analysis (20  $\mu$ L) were prepared with the addition of 5x stock loading buffer (5  $\mu$ L). These solutions were vortexed and briefly centrifuged before having DTT (1  $\mu$ L, 500 mM) added. The solutions were vortexed, centrifuged and warmed at 60 °C for 10 minutes. IAA (1  $\mu$ L, 500 mM) was added and the solutions were vortexed, centrifuged and incubated at room temperature in darkness for 10 minutes. The comb was removed from the pre-cast gel and the protein sample solutions were then pipetted into the wells. The running buffer was poured into the electrophoresis apparatus and the voltage was set to 130 V. This was run for approximately 85 minutes.

Gels were either stained by coomassie, silver or transferred to a membrane for Western blotting. Coomassie staining of the SDS electrophoresis gel was carried out using a coomassie blue R-250

buffer (0.25% coomassie blue R-250, 50 % MeOH, 10 % HoAC, 40 % H<sub>2</sub>O) with agitation at room temperature overnight. Following staining, the gel was removed and placed into a destaining solution (30 % MeOH, 5 % acetic acid, 65 % H<sub>2</sub>O) until the bands containing protein were visualised.

For silver staining, the silver staining kit SilverQuest, manufactured by Invitrogen (California, United States) was used. Solutions were prepared prior to staining; fixing solution (40 % EtOH, 10 % AcOH, 50 % H<sub>2</sub>O), washing solution (30 % EtOH, 70 % H<sub>2</sub>O), sensitising solution (30 % EtOH, 10 % sensitiser, 60 % H<sub>2</sub>O), staining solution (1 % stainer, 99 % H<sub>2</sub>O), developing solution (10 % developer, 0.1 % developer enhancer, 89.9 % H<sub>2</sub>O). Following electrophoresis, the gel was agitated in fixing solution (100 mL) for 20 minutes at room temperature. The fixative solution was decanted and the gel was washed for 10 minutes in the washing solution. The washing solution was decanted and sensitising solution (100 mL) was added to the gel. This was agitated for 10 minutes at room temperature. The solution was then discarded and the gel was washed in the washing solution (100 mL) for 10 minutes. This was followed by a wash with H<sub>2</sub>O (100 mL) for 10 minutes. The water was discarded and the gel was placed into the staining solution (100 mL) and agitated for 15 minutes. Following staining, the solution was decanted and the gel was washed with H<sub>2</sub>O (100 mL) for approximately 40 seconds. The gel was then incubated in developing solution (100 mL) until the bands of protein were visualised. Stopping solution (10 mL) was then directly added to the gel in the developing solution.

For Western Blotting, following electrophoresis, the gel was incubated at 4 °C in blotting transfer buffer (100 mL, 25 mM Tris, 190 mM glycine, 20 % methanol) for 20 minutes. A polyvinylidene difluoride (PVDF) membrane cut to the size of the gel was activated with 100 % methanol and incubated at 4 °C in blotting transfer buffer (100 mL) for 20 minutes. The gel and membrane were assembly together with filter papers and loaded into the transfer apparatus. Voltage was set at 100 V and the transfer was run for approximately 70 minutes. Following transfer, the membrane was blocked with 5 % skim milk in tris-buffered saline (TBS, 60 mL) at room temperature for an hour. The membrane was then incubated at room temperature with the primary antibody (PFD subunit 3, PFD subunit 5, PFD subunit 6 purchased from Santa Cruz, 5 µL in 10 mL TBS) for 3 hours. The membrane was then washed with TBS (100 mL) with 0.1 % of Tween (TBST) for 10 minutes. This washing was repeated 3 more times with a new solution of TBST. The membrane was then incubated with the secondary antibody (anti-goat, 2 µL in 10 mL TBS). The membrane was then washed with TBST (100 mL) for 10 minutes at room temperature and this was repeated 3 more times with a new solution of TBST. The membrane was then cut into strips, incubated with GLO (3 mL) and then imaged with a ChemiDoc MP (Bio-rad, California).

### 3.3.6: Protein Concentration Determination

Protein concentrations were determined by UV-vis absorption measured at 280 nm using a Nano 2000 (Varian Ltd, Australia), using extinction coefficients calculated with online open source software (<http://web.expasy.org/protparam/>; calculated extinction coefficients summarised in Appendix E, or through the ThermoFisher Scientific (Massachusetts, United States) Bradford assays. Bradford assays were conducted using bovine serum albumin (BSA) as a protein standard to construct the calibration curve and the stock solution (2000 µg/mL) was diluted to varying concentrations between 2000-0 µg/mL with the appropriate solvent (milliQ water, 100mM AA, PBS). Each solution of different BSA concentrations (10 µL) and solutions of the unknown proteins (10 µL) were added to Coomassie Plus Reagent (300 µL) at room temperature and briefly vortexed. Protein concentrations were determined by the Nano 2000 and averaged over at least 3 reads. A calibration curve was constructed using at least 6 different concentrations of BSA. Concentrations of the unknown proteins were averaged over 3 readings.

### 3.3.7: Aβ(1-40) Aggregation and Inhibition ThT Assays

A stock solution of Aβ(1-40) was prepared by dissolving the peptide in 0.5 % NH<sub>4</sub>OH at a concentration of 0.5 µg/µL. The solution was filtered using a spin filter (10,000 MW, Merck, Massachusetts, United States) and the solution with the peptide was then diluted to the appropriate concentration with PBS.

Samples for the in situ ThT assay were prepared by diluting the appropriate concentrated protein and peptide stock solution into PBS solution at the appropriate concentrations. IgG was extracted from crude ThT dye was added to the proteins (40 µM) of which 30 µL solutions were loaded into a black, clear bottom 384-well plate (Greiner Bio-One, Stonehouse, UK) that were sealed with clear film (Sigma Aldrich, Castle Hill, NSW) to prevent evaporation. Proteins were incubated at 37 °C, without shaking for 83 hours. The ThT fluorescence intensities of the samples were recorded using a Fluostar Optima plate reader (BMG Labtechnologies, Offenburg, Germany) fitted with 440/490 nm excitation/emission filters. Data were normalized by plotting the change in ThT fluorescence (F, arbitrary units [a.u.]) from the initial fluorescence reading. Experiments were averages of 3 independent experiments with at least 4 replicates per experiment.

### 3.3.8: Cell Cytotoxicity Assay

An MTT assay was used to determine the cytotoxicity of the peptides and proteins. Rat phenochromocytoma cells (PC-12 cells) were seeded the day prior to treatment with a density of

approximately 100 cells per well in 2 96-well plates. Non-fibrillar peptides and proteins were dissolved and diluted in PBS and used immediately. Incubated A $\beta$ (1-40) and *h*PFD subunits were taken from ThT assays and were diluted to the desired concentration in PBS, snap frozen on dry ice and stored at -20 °C until use. After treatment of the cells with the appropriate peptide or protein, the cells were incubated for 72 hours. The tissue culture medium was then aspirated from the wells and a serum free media containing MTT (0.25 mg/mL) was added. The plate was then incubated for 3 hours at 37 °C under a 5 % CO<sub>2</sub> atmosphere. The medium was removed and DMSO (100  $\mu$ L) was added. The plate was gently agitated for 10 seconds and the resultant absorbance of formazan was measured using a Fluostar Optima platereader (BMG Laboratories, UK) at 570 nm. Cell viability values are an average of 3 independent experiments with at least 8 replicates per experiment. Cell viability was assessed as percentage absorbance relative to the vehicle control (PBS only). These calculated relative values were analysed for statistical significance using a standard post-hoc test following ANOVA one way analysis using Graph pad prism (San Diego, California).

# Chapter 4:

## Far-UV Circular Dichroism

### Spectroscopy Analysis of the

### Human Prefoldin Subunits 5 and 6

---

#### 4.1: Introduction

Circular Dichroism (CD) spectroscopy has the capacity to determine the predominant secondary structure of peptides and proteins through the characteristic spectra observed as mentioned earlier in Chapter 2 [153]. In previous studies utilising this technique, far-UV CD spectroscopy has successfully monitored the secondary structure changes of proteins with small changes in sample conditions [150], [157], [351]–[353]. The pH and temperature of a solution can affect the protein conformation, and CD assays as a function of pH or temperature, at the characteristic  $\lambda$ , allows the protein's change of state and therefore unfolding and unfolding intermediates to be examined [153]. The heat capacity change can also be estimated from the acquired CD data if the unfolding transition is also perturbed as a function of protein concentration or salt [157]. The thermodynamics of protein unfolding can be investigated through the acquisition of spectra as a function of temperature by monitoring the ellipticity at a single wavelength. However, CD spectra acquired over the  $\lambda$  range of 190–260 nm allows more information to be conveyed and the residual structure achieved following equilibrium unfolding may be better characterized [112], [158]. It should be noted however, that care must be taken in the analysis of proteins with multiple unfolding transitions due to the collective CD spectra obtained and the difficulty in separating the overlapping transitions [157], [354].

Previous studies of PFD suggest the composition of the chaperone complex is primarily  $\alpha$ -helices, with each individual subunit containing a primary helical domain [84], [93]. These constructed models of PFD have been found to be consistent across all origins of the hexameric complex [84], [91], [95], [98]. However, these studies also suggest the presence of  $\beta$ -sheets and turns located at the top of the jelly-fish like structure of PFD. The presence of different secondary structures will ultimately result in a convoluted CD spectrum. The dedicated software DichroWeb can deconvolute



CD spectroscopy data of proteins to reveal the different proportions of secondary structure. It has been pre-determined that the CDSSTR algorithm available in DichroWeb produces the most accurate deconvolution of CD spectra and can produce results where other methods fail to analyse the data [160]. This method provides a superior fit of the conformation of globular proteins compared to other algorithms, particularly for estimates of  $\alpha$ -helical content. The output of the analysis provides a summary of the best calculated deconvolution, the details in which the deconvolution occurred, an overlaid plot of the experimental and reconstructed CD data and a standard goodness-of-fit parameter, the normalized root mean square deviation (NRMSD). The NRMSD provides an indication of how closely related the constructed basis curve produced from the predicted secondary structures is to the experimental spectrum. Conformation estimates are generally rejected if the NRMSD is higher than 0.05 [160]–[162], [164].

#### 4.1.1: Aim

This chapter aims to confirm the predominant secondary structure of *h*PFD subunits 5 and 6 at physiological conditions, to gain further insight into the PFD constituent subunits.

Secondly, due to the solutions of varying pH required for NMR and cross-linking analysis, the effect of pH on the *h*PFD subunits is investigated by far-UV CD. This work aims to determine if minor changes in pH affect the secondary structure of *h*PFD subunit 5 and 6 as they are analysed.

In addition, measuring CD spectra as a function of temperature can determine the stability of the protein. Thermostability assays of the *h*PFD subunits aim to investigate the thermal denaturation of the proteins.

Finally, structural investigation of the products following the incubation of A $\beta$ (1-40) and the respective subunit is undertaken to gain further insight into the chaperoning activity of the proteins.

## 4.2: Results and Discussion

### 4.2.1: Far-UV CD Spectroscopy of *h*PFD Subunits 5 and 6

The far-UV CD spectra of *h*PFD subunits 5 and 6 were recorded. Figure 4.1 reveals a similar overall secondary structure between the 2 subunits, with minima observed at 222 nm and 208 nm and a maximum observed at 193 nm. As previously indicated in Section 2.2.3, this is indicative of  $\alpha$ -helices and therefore it can be concluded from this analysis that *h*PFD subunits 5 and 6 are comprised predominantly of  $\alpha$ -helices.

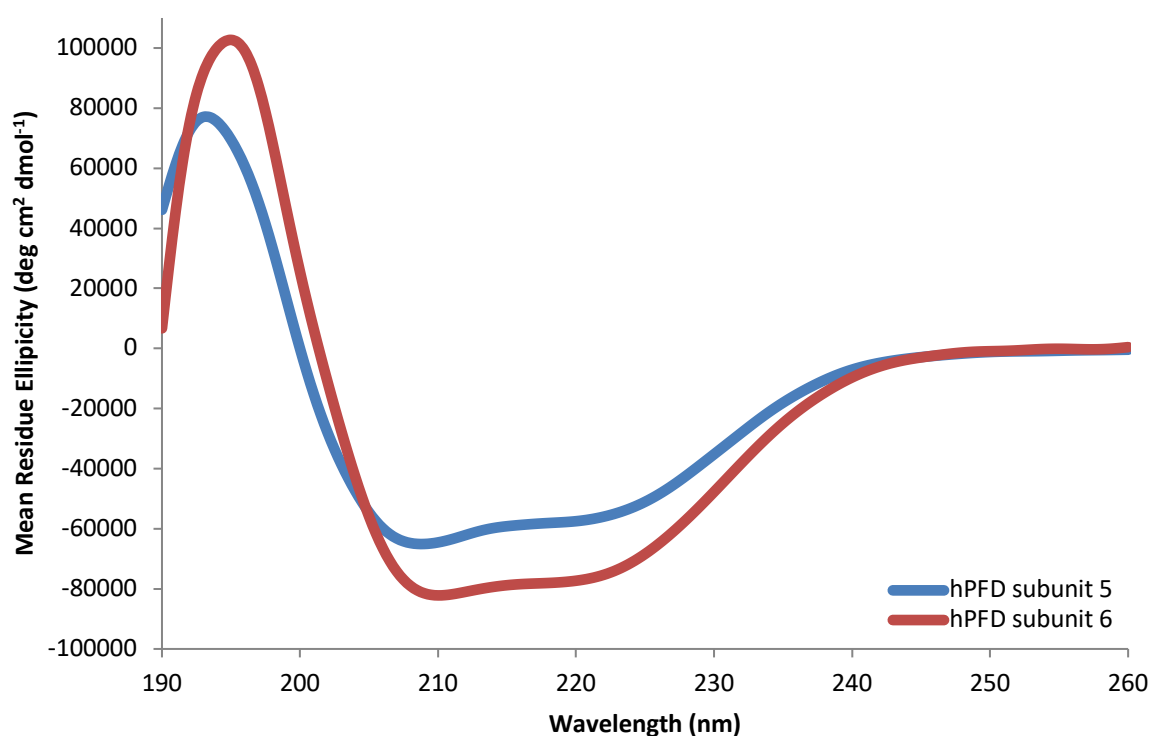


Figure 4.1: Far-UV CD spectra of *h*PFD subunit 5 (blue) & *h*PFD subunit 6 (red) (10  $\mu$ M) in a solution of 10 mM NaPO<sub>4</sub>H<sub>2</sub> at pH 7. Representative data are shown from at least two independent experiments performed.

Although Figure 4.1 reveals *h*PFD subunit 5 and 6 are similar in secondary structure, the magnitude of MRE differs between the constituents, with a higher absorbance observed for *h*PFD subunit 6 than *h*PFD subunit 5. As the raw data has been adjusted to MRE, this suggests that *h*PFD subunit 6 has a higher degree of absorbing chromophores. Although some amino acid residue side chains absorb light strongly in the far-UV region (Tyr, Trp, Phe and Met), the most important contributor is the peptide bond. The contribution of any individual chromophore to the total absorbance of the

protein will depend, to some extent at least, upon its environment [355]. It is subsequently concluded from these results that there is a higher content of structured  $\alpha$ -helices in *h*PDF subunit 6 than in *h*PDF subunit 5 at the same concentration.

Some publications also detail that the 222:208 ratio can indicate a coiled-coil structure [356], [357]. A quick calculation of the ratios from Figure 4.1 reveals that *h*PDF subunit 5 and 6 have ratios of 0.95, respectively. Depending on the literature, single helices have a ratio of 0.9 while coiled coils have a ratio of greater than 1.10. Unfortunately, variations in the specific numbers were seen across different publications [358]. Consequently, this ratio was not insightful in this instance.

#### 4.2.1.1: DichroWeb Deconvolution of CD Spectroscopy of *h*PDF Subunit 5 and 6

To help determine the relative proportion of the secondary structures of *h*PDF subunit 5 and 6, the open software DichroWeb was employed to deconvolute the CD spectra. As previously mentioned, the algorithm implemented is CDSSTR. To maximise the accuracy of the output, the reference set SMP180 was chosen [359]. Of the options available, this reference set has the highest number of CD data of globular proteins available. Figure 4.1 revealed that *h*PDF subunit 5 and 6 have CD curves with characteristic  $\alpha$ -helical content consistent with previously modelled structures of the subunits [89], [91], [328]. For comparison, the CD spectrum of standard bovine serum albumin (BSA) was also recorded and is shown here. Figure 4.2 reveals the reconstructed basis curves using CDSSTR and the reference set SMP 180 of *h*PDF subunit 5, *h*PDF subunit 6 and standard BSA.

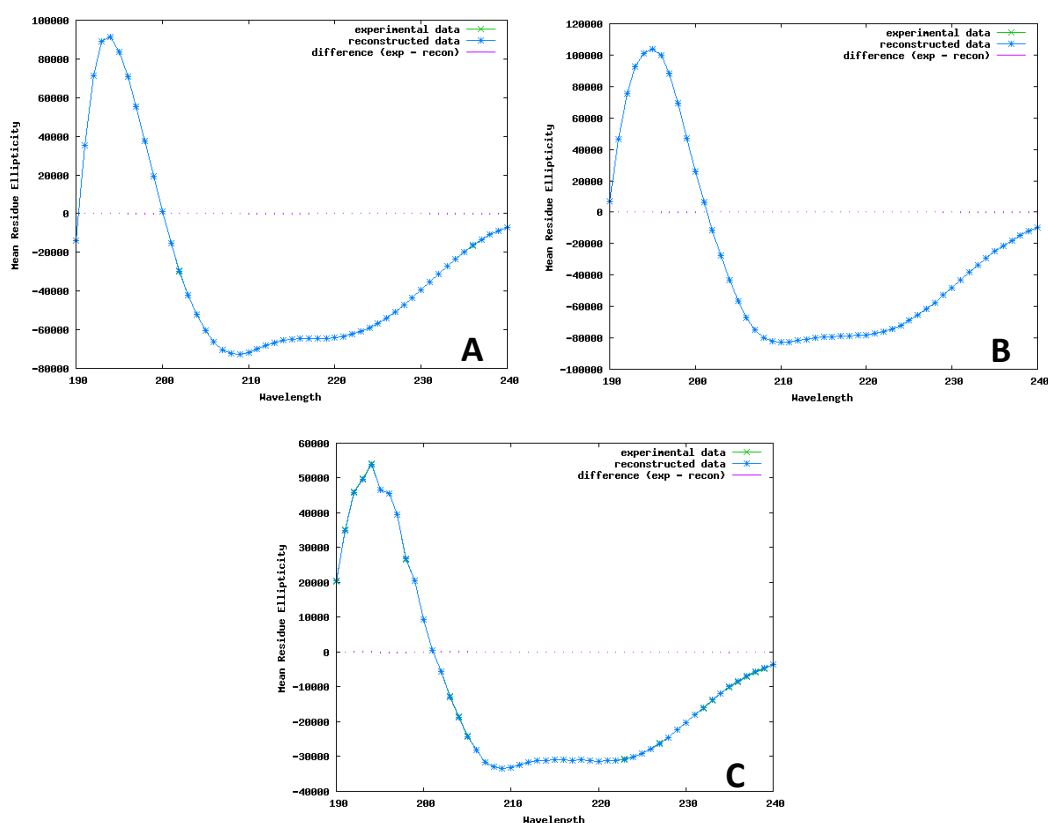


Figure 4.2: CDDSTR reconstruction of the experimental CD spectroscopy data of (A) *hPFD* subunit 5, (B) *hPFD* subunit 6, & (C) standard BSA utilising DichroWeb from CD data obtained in solutions of 10 mM  $\text{NaPO}_4\text{H}_2$  at pH 7.

As seen in Figure 4.2, there are small to almost no differences observed between the experimental data to the reconstructed data and this is reflected in the small NRMSD value summarised in Table 4.1. The minimal amount of difference reflected in the spectra is an indication of a good reconstructed basis curve using the algorithm CDDSTR and the reference set SMP180. Table 4.1 summarises the CDSSTR deconvolution of *hPFD* subunit 5, *hPFD* subunit 6, and BSA.

Table 4.1: Calculated secondary structure fractions & spectral fitting of *hPFD* subunits 5, 6 and BSA utilising the software available at DichroWeb. The range is calculated from two replicate experiments.

	Alpha helix	Beta sheet	Turns	Unordered	NRMSD
<i>hPFD</i> subunit 5	0.65-0.58	0.16-0.14	0.09*	0.19-0.10	0.003-0.001
<i>hPFD</i> subunit 6	0.76-0.61	0.11-0.08	0.13-0.08	0.13-0.08	0.003-0.001
Standard BSA	0.75-0.73	0.12-0.09	0.07-0.06	0.10*	0.003*

\*The same value was calculated for both deconvolutions.

It can be seen in Table 4.1 that the CD spectra deconvolution of *hPFD* subunit 5 and 6 reveal similar content of secondary structures. CDSSTR deduced that both subunits have approximately 60 %  $\alpha$ -

helical structure, with *h*PF<sub>2</sub>D subunit 6 having a slightly higher  $\alpha$ -helical content. *h*PF<sub>2</sub>D subunit 5 was also calculated to have a slightly higher fraction of  $\beta$ -sheet, approximately 5 % higher than *h*PF<sub>2</sub>D subunit 6. A higher percentage range was also observed for unordered structures for *h*PF<sub>2</sub>D subunit 6, while *h*PF<sub>2</sub>D subunit 5 was calculated to have higher percentages of turns. This aligns with the previously mentioned proposition that *h*PF<sub>2</sub>D subunit 6 must have a higher content of structure.

It should be noted that in Section 3.2.3, details of the synthesis of *h*PF<sub>2</sub>D subunits 5 and 6 required additional amino acids for purification and translation. It is not known what structure these additional amino acids fold into and it is possible that the contribution of these tags to the spectra may affect secondary structure estimations and this must be noted. However, as both synthesised subunits have identical extensions, it is likely the extensions will behave similarly therefore allowing direct comparisons between the subunits.

Estimates of CD deconvolution errors due to inaccuracies in the CD experiment have been cited to be approximately  $\pm 5$  % [360]. A simple test utilising a protein standard was therefore conducted to verify the uncertainty established. The CD deconvolution results in Table 4.1 show that pure BSA is composed of 75-73 %  $\alpha$ -helices, 12-9 %  $\beta$ -sheets, 7-6 % turns and 10 % random structures. In previous studies of BSA using far-UV CD spectrometry, the domains were calculated to be predominantly  $\alpha$ -helical between 66-70 % [361]–[363]. These published results would also have error margins of  $\pm 5$  %, and consequently overlap with the results seen here. This therefore suggests that our CD deconvolution results are relatively consistent. All deconvoluted data were therefore treated with a  $\pm 5$  % uncertainty and differences higher than  $\pm 5$  % were deemed as significantly different.

It can also be seen that the NRMSD for *h*PF<sub>2</sub>D subunits 5 and 6 had low scores ranges of 0.003-0.001 and BSA was determined to have a NRMSD of 0.003. Although high scores of NRMSD ( $>0.05$ ) can contribute to errors associated with the calculated deconvolution, a low NRMSD alone does not necessarily mean that an analysis is accurate. However, DichroWeb calculates this parameter methodically for each analysis and therefore it provides a comparison of results between the different data sets [160]–[162]. It should also be noted that accuracy of concentration, precise measurement of cell pathlength, minimum acquisitions of 190 nm and proper CD data reconfiguration are stated to be essential for valid deconvolution and minor inconsistencies in these can result in large differences in calculations [160]. Fortunately, the majority of the deconvolutions seen here appear relatively consistent to previous CD analyses and structure determination techniques and subsequently DichroWeb deconvolutions were conducted for all the CD data acquired in this Chapter.

#### 4.2.2: Far-UV CD Spectroscopy of the *h*PFD Subunits 5 and 6 at Varying pH

As previously explained, far-UV CD spectroscopy is able to determine the predominant secondary structure of proteins in differing solutions of low salt concentration. Therefore, this technique can also be utilised to study the effects of extrinsic conditions on a protein's secondary structure. Far-UV CD spectroscopy was used to examine the effect of changing pH in solution on the conformation of *h*PFD subunits 5 and 6. This was performed to determine if any secondary conformational changes result due to the slight change in conditions necessary for other structural investigations.

*h*PFD subunit 5 was diluted to 10  $\mu$ M in solutions of 10 mM NaPO<sub>4</sub>H<sub>2</sub> at pH 6, 7 and 8. Changes in the CD spectra were observed from the marginally acidic to the alkaline solution (Figure 4.3). The CD spectra measured at pH 6 reveals the MRE is lower at 222 nm than for *h*PFD subunit 5 at physiological pH. This suggests a higher content of structured  $\alpha$ -helices in the slightly more acidic solution. It is also noted that a shift of the maxima is observed between the solutions. At pH 6, *h*PFD subunit 5 is seen to exhibit a maximum at 198 nm, rather than at the typical 193 nm for  $\alpha$ -helices, and this suggests chromophores of structured  $\beta$ -sheets. In comparison, a maximum is exhibited at 193 nm for *h*PFD subunit 5 at pH 7 and 8 (Figure 4.3). At pH 8, a weaker absorption of the chromophores is observed with a smaller change in MRE compared to the curve at pH 6 and 7.

A change in pH is thought to disrupt weak ionic bonds, thus affecting salt bridges and hydrogen bonds of the protein and resulting in the unfolding of the structure [364]. However, structural changes based on small pH differences must also take into consideration the pI of the protein [365]. The computationally calculated pI of *h*PFD subunit 5 is 5.98 (refer to Appendix E). Solutions with pH values at the protein pI results in an overall primary net charge equalling 0 and therefore weak non-covalent bonds are not compromised. Typically, proteins have a tendency to precipitate out of solution when the pH equals the pI [365], [366]. This was fortunately not directly observed. As the pH moves away from the pI of 5.98, the primary net charge of the protein decreases with an increase in aqueous HO<sup>-</sup> in solution. However, although the non-covalent interactions are marginally disrupted with conformation reconfiguration observed with the change in pH,  $\alpha$ -helical structure is seen to be relatively preserved as seen from the minima of 208 and 222 nm in Figure 4.3.

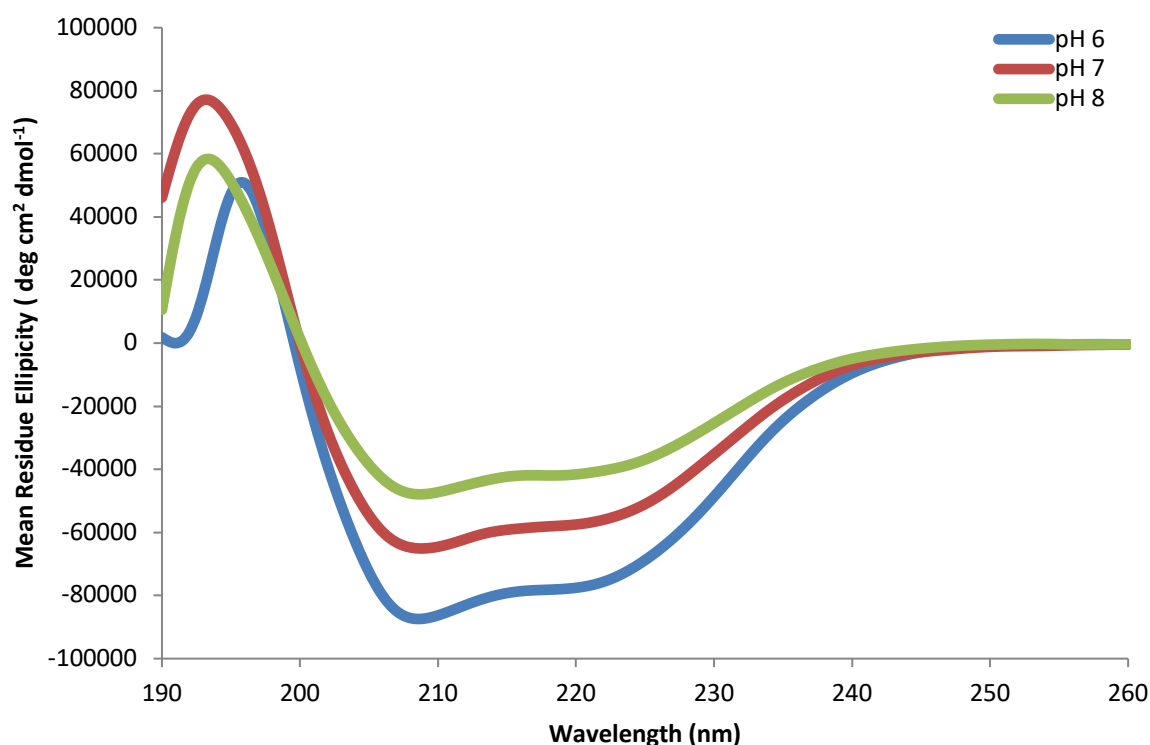


Figure 4.3: Far UV CD spectra of subunit 5 (10  $\mu$ M) in 10 mM NaPO<sub>4</sub>H<sub>2</sub> at pH 6 (blue), pH 7 (red) & at pH 8 (green). Representative data are shown from at least two independent experiments performed.

Deconvolution of the CD curves by CDSSTR are summarised in Table 4.2 and reveal a decrease in  $\alpha$ -helical content and an increase in unordered structure with the increase of pH.

Table 4.2: Calculated secondary structure fractions & spectral fitting of *h*PF<sub>D</sub> subunit 5 in varying solutions of pH utilising the software available at DichroWeb. The range is calculated from two replicate experiments.

pH	Alpha helix	Beta sheet	Turns	Unordered	NRMSD
6	0.76-0.71	0.08-0.06	0.10-0.08	0.13-0.10	0.006-0.001
7	0.65-0.58	0.16-0.14	0.09*	0.19-0.10	0.003-0.001
8	0.61-0.48	0.12-0.09	0.15-0.12	0.27-0.16	0.028-0.004

\*The same value was calculated for both deconvolutions.

Figure 4.3 reveals that the CD spectrum obtained for *h*PF<sub>D</sub> subunit 5 in a solution of pH 6 has the highest signal at  $\lambda$  222 nm implying the highest amount of  $\alpha$ -helical content and this is reinforced in the CDSSTR deconvolution. At pH 6, the calculated secondary structure content was 76-71 %  $\alpha$ -helical content and 8-6 %  $\beta$ -sheets. It is seen in Table 4.2 that the increase in pH decreases the calculated  $\alpha$ -helical percentage range and increases the  $\beta$ -sheet and unordered fractions, while the number of turns remained relatively consistent at different pH. These results align with Figure 4.3,

however it must be noted that the calculated deconvolutions represent the fractions of secondary structure present in the sample at the time of CD data acquisition. DichroWeb does not take into account the magnitude of the spectra and therefore it subsequently does not quantify the protein in the different solutions. This ultimately means that although a direct comparison of the overall secondary structure content can be compared through the deconvolution, overall protein concentrations cannot be compared. Precipitation was not directly observed in the solutions following CD data acquisition, but this cannot be completely ruled out as attempts to examine concentrations following the experimental were not conducted. Some biophysical analyses require different pH buffer conditions and the CD results obtained here are able to show the effects of this change on the secondary structure. It can be summarised from Table 4.2 that an increase in pH adversely affects the conformation of *h*PFD subunit 5 as the  $\alpha$ -helical content decreases by approximately 20 % when pH is increased from 6 to 8.

Analogous to the assays conducted for *h*PFD subunit 5, CD spectra of subunit 6 in the 3 solutions at differing pH are seen to have a predominant  $\alpha$ -helical secondary structure with minimum at 208 and 222 nm. Similar to the CD curves seen in Figure 4.3, a decrease in pH gives rise to increasing MRE signal intensity at  $\lambda$  208 and 222 nm, consistent with an increase in  $\alpha$ -helical content. The CD spectrum of subunit 6 at pH 8 reveals maxima at 190 and 200 nm (Figure 4.4). These maxima are not observed for *h*PFD subunit 6 in solutions at pH 7 or 8. Instead, the maxima are seen at 199 and 190 nm for pH 7 and 8 respectively. It is likely that with the increase in pH, a shift to more  $\beta$ -sheet rich structure occurs as the  $\alpha$ -helices are disrupted.



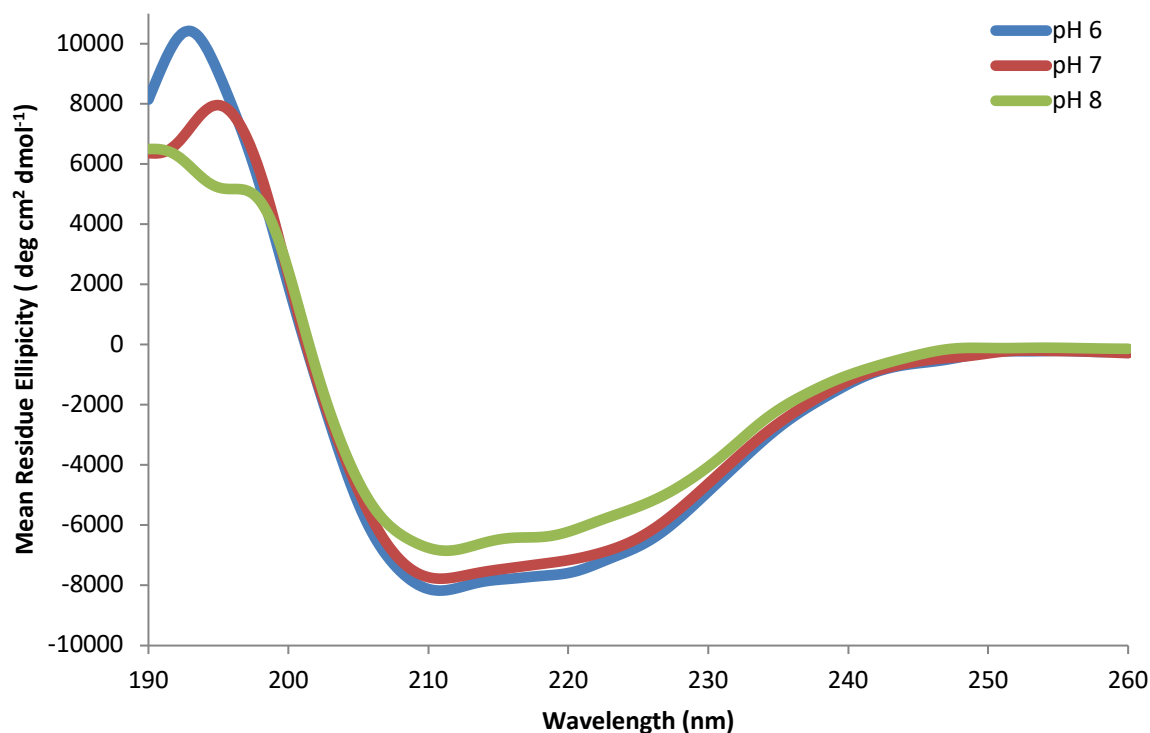


Figure 4.4: Far-UV CD spectra of subunit 6 (10  $\mu$ M) in 10 mM  $\text{NaPO}_4\text{H}_2$  at pH 6 (blue), pH 7 (red) & at pH 8 (green). Representative data are shown from at least two independent experiments performed.

The pI for subunit 6 is computationally calculated to be 6.60 (Refer to Section 5.3.3). As the CD spectra seen for *h*PF<sub>2</sub>D subunit 6 are relatively similar to the spectra of *h*PF<sub>2</sub>D subunit 5 (Figure 4.3), the same mechanisms previously proposed are also suggested here and can explain the difference in CD signal magnitude exhibited for *h*PF<sub>2</sub>D subunit 6 between the different pH solutions. Figure 4.4 reveals that the CD data acquired for *h*PF<sub>2</sub>D subunit 6 in solutions with pH 7 and 6 have marginal differences in signal intensities. As the pI of *h*PF<sub>2</sub>D subunit 6 is calculated to be in between these values, it is therefore unsurprising that there is little difference between the curves. The further away the pH of the solution is to the pI value of 6.60, the higher the chances of non-covalent disruption occurring to *h*PF<sub>2</sub>D subunit 6. However, the small changes observed in the CD spectra suggest subtle adjustments in the environment of numerous residues. Deconvolution of the spectra by CDSSTR reinforces the changes observed in the spectra and the calculated secondary structures are summarised in Table 4.3.

**Table 4.3: Calculated secondary structure fractions & spectral fitting of *h*PF<sub>D</sub> subunit 6 in varying solutions of pH utilising the software available at DichroWeb. The range is calculated from two replicate experiments.**

pH	Alpha helix	Beta sheet	Turns	Unordered	NRMSD
6	0.76-0.71	0.10-0.09	0.08-0.07	0.12-0.08	0.003*
7	0.76-0.68	0.11-0.08	0.13-0.08	0.13-0.08	0.003-0.001
8	0.65-0.61	0.18*	0.08-0.07	0.15-0.09	0.003-0.002

\*The same value was calculated for both deconvolutions.

The calculated secondary structure fractions of *h*PF<sub>D</sub> subunit 6 in a solution of pH 6 is seen in Table 4.3 to show  $\alpha$ -helical content of 76-71 %,  $\beta$ -sheet content of 10-9 % and turns and unordered content at 8-7 and 12-8 %, respectively. The calculated secondary structure percentages with an increase in pH to 7, is seen to retain similar content, with an approximate 3-5 % difference observed in  $\beta$ -sheet and turns content. The deconvoluted structures of *h*PF<sub>D</sub> subunit 6 at pH 7 and 8 therefore have marginal differences, indicating that there are no significant conformational changes in structure at these pH levels. Further pH increases to pH 8 is seen to decrease the  $\alpha$ -helical content by 15-11 %, increase the  $\beta$ -sheet content by approximately 10 %, while turns and disordered structure remain relatively the same. Table 4.3 ultimately reveals that the increase in pH to 8 adversely affects the structure of the *h*PF<sub>D</sub> subunit 6, as seen by the decrease in calculated  $\alpha$ -helices and increase in  $\beta$ -sheets. Therefore it can be concluded that in solutions of pH 6, both *h*PF<sub>D</sub> subunits are calculated to exhibit chromophores with high percentages of  $\alpha$ -helices. The increase in pH solution to 7, is seen to affect *h*PF<sub>D</sub> subunit 5 while the calculated percentages for *h*PF<sub>D</sub> subunit 6 remain the same. CD analysis of the subunits in a solution of pH 6 reveals the acidity of the solution is seen to have an adverse effect on the  $\alpha$ -helical structure of both subunits.

To determine the magnitude of  $\alpha$ -helical protein structure change with the change of pH, the differences observed at 222 nm for *h*PF<sub>D</sub> subunits 5 and 6 were normalised against the maximum MRE value obtained (shifts seen in solutions of pH 6) and plotted onto a graph (Figure 4.5). Figure 4.5 reveals that increasing the pH decreases the overall amount of chromophores exhibiting  $\alpha$ -helical signals for both *h*PF<sub>D</sub> subunits, although it can be seen that *h*PF<sub>D</sub> subunit 5 has a bigger decrease in  $\alpha$ -helical content than *h*PF<sub>D</sub> subunit 6. This ultimately means that *h*PF<sub>D</sub> subunit 5 is more susceptible to the pH change. As the calculated pI of the *h*PF<sub>D</sub> subunit 5 is 5.98, it is likely that this subunit has the least compromised structure in the data acquired with the solution of pH 6. In comparison, the calculated pI of *h*PF<sub>D</sub> subunit 6 is 6.60. This protein is found to exert little change in conformation in solutions of pH 6 and 7, suggesting a higher resilience against external pH conditions than *h*PF<sub>D</sub> subunit 5. Nevertheless,  $\alpha$ -helical secondary structures are consistently seen to be the predominant conformation adopted by the subunits in these solutions of different pH. It is

concluded that the variations in pH necessary for several protein structure determination techniques may induce some conformational differences and will be considered with caution.

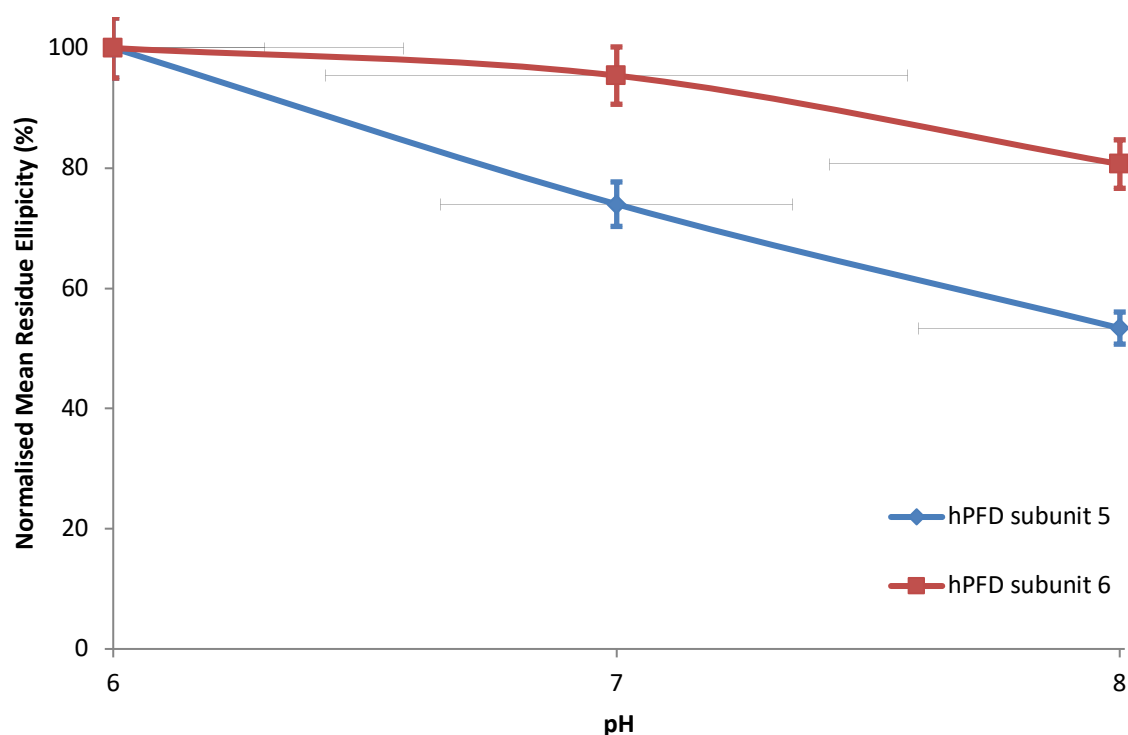


Figure 4.5: pH profiles of *hPFD* subunit 5 (blue) & *hPFD* subunit 6 (red) at varying pH conditions of aqueous buffers 10 mM NaPO<sub>4</sub>H<sub>2</sub>. The data presented here are changes in ellipticity at 222 nm as a function of pH and the values were normalised against the MRE observed at pH 6, with +/- SD from 2 independent repeats of the experiment.

### 4.2.3: Thermostability of *hPFD* Subunits 5 and 6 as monitored by Far-UV CD Spectroscopy

As a protein is thermally unfolded and denatured, it loses its secondary structure. The temperature range over which this occurs depends on the inherent thermostability of the protein. Therefore, by measuring the changes in a protein's far UV spectrum as a function of temperature, its thermostability can be assessed. In this way the thermostability of *hPFD* subunit 5 and *hPFD* subunit 6 was investigated. Due to the highly  $\alpha$ -helical nature of the *hPFD* subunits, the far UV signal between 190-260 nm was measured and the change at 222 nm was plotted as a function of temperature, over a temperature range of 20-90 °C.

*h*PFD subunit 5 was diluted to a concentration of 10  $\mu$ M into a solution of 10 mM NaPO<sub>4</sub>H<sub>2</sub> at pH 7 and subjected to thermal stress prior to data acquisition. The CD data acquired is plotted in Figure 4.6.

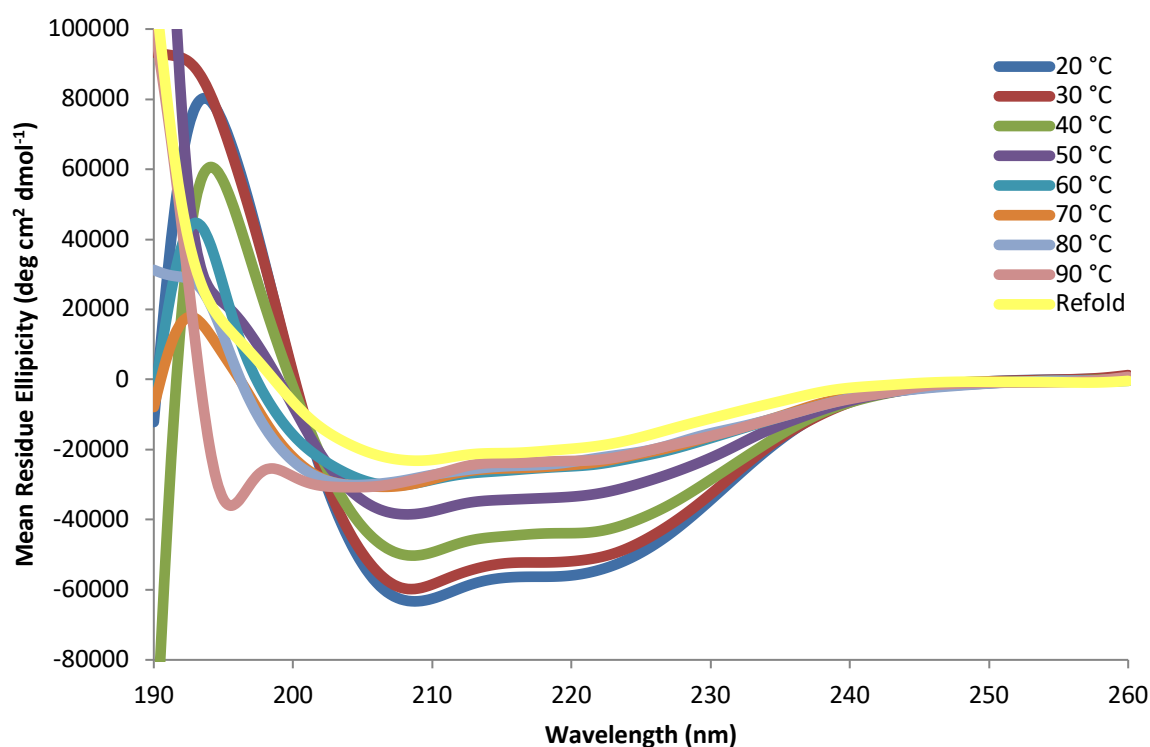


Figure 4.6: Far-UV CD spectra of *h*PF subunit 5 (10  $\mu$ M) in 10 mM NaPO<sub>4</sub>H<sub>2</sub> at pH 7, between the temperature range of 20-90 °C. Following thermal denaturation, the sample was incubated on ice for 24 hr & a far-UV CD spectrum was measured (refold). Representative data are shown from at least two independent experiments performed.

Figure 4.6 shows that over the temperature range of 20-80 °C, minima are observed at  $\lambda$  208 and 222 nm, which decrease in magnitude with the increase in temperature. As these signals are indicative of  $\alpha$ -helices, the decreasing magnitude is therefore an indicator of decreasing  $\alpha$ -helix structure, and possibly signifies thermal denaturation. As an overall loss of signal is observed, it is likely that some protein precipitation has occurred with the increase in temperature. A maximum observed for the CD curves at temperatures of 20-50 °C is seen at 198 nm, which suggests the presence of  $\beta$ -sheet. At the temperature range of 60-80 °C the maximum appears to have shifted to 193 nm, the characteristic absorption of  $\alpha$ -helices. Protein structure typically does not convert from  $\beta$ -sheets to  $\alpha$ -helices with thermal aggregation, therefore a deconvolution of the data is required to help explain the absorption profiles seen here. However, it is proposed that the increase in solution induces the conversion of the structured conformation to disordered aggregates as an isodichroic

point is seen at 203 nm. The presence of this point suggests the presence of a local 2 state population transition [367]–[369]. In this instance the states are of  $\alpha$ -helical and unordered structures due to the position of the isodichronic point established between the characteristic shifts of 222 and 195 nm. Little change is observed in the CD spectra of *h*PFD subunit 5 between the temperature range of 60–90 °C, suggesting that the protein has fully denatured at temperatures higher than 60 °C. Although the CD data obtained at 90 °C suggests some residual  $\alpha$ -helical structure with minima at 208 and 222 nm observed, the largest feature for this curve appears at 195 nm. The minimum observed at this  $\lambda$  suggests the structure of *h*PFD subunit 5 at 90 °C is primarily disordered.

Following thermal aggregation, the solution was cooled on ice for 24 hours to allow refolding. Precipitate was removed from the solution and a CD spectrum was recorded. The characteristic absorbances for disordered structure were not observed and instead only characteristic minima for  $\alpha$ -helix were noted. This does not necessarily suggest that refolding of *h*PFD subunit 5 has occurred, as the MRE measured at 222 nm following cooling is of significantly decreased magnitude. It is proposed that partial retention of the secondary structure has occurred throughout the thermal stress and subsequent cooling. Precipitate formed following thermal stress observed in the solution is likely to have formed from the disordered aggregates.

The CD curves were deconvoluted using the algorithm CDSSTR with the reference set SMP180 and results are summarised in Table 4.4. It should be noted that the CD deconvoluted result for 50 and 80 °C reveals NRMSDs over a range higher than 0.05. It was stipulated in Section 4.2.1.1 that high scores of NRMSD have errors associated with the calculation, consequently the deconvoluted fractions for *h*PFD subunit 5 at these temperatures are rejected from the analysis. The deconvoluted results for  $\alpha$ -helix and unordered structure are seen to follow a particular trend while turns seem to remain relatively constant throughout the assay, ranging between 9–14 %. It is reminded that the calculated deconvoluted secondary structures are a reflection of only the soluble proteins present in the sample.

**Table 4.4: Calculated secondary structure fractions & fitting parameters of *h*PFD subunit 5 incubated at different temperatures, utilising the software available at DichroWeb. The range is calculated from two replicate experiments.**

Temperature	Alpha helix	Beta sheet	Turns	Unordered	NRMSD
20 °C	0.65-0.58	0.16-0.14	0.09*	0.19-0.10	0.003-0.001
30 °C	0.77-0.65	0.11-0.06	0.11*	0.14-0.06	0.005-0.001
40 °C	0.58-0.54	0.15-0.13	0.11-0.09	0.21-0.17	0.004-0.002
50 °C	0.51-0.42	0.30-0.21	0.09-0.07	0.22-0.20	0.081-0.003 <sup>i</sup>
60 °C	0.54-0.47	0.19-0.18	0.11-0.10	0.23-0.18	0.004-0.002
70 °C	0.44*	0.17-0.11	0.14-0.13	0.31-0.26	0.014-0.007
80 °C	0.34-0.16	0.34-0.15	0.14-0.13	0.37*	0.058-0.031 <sup>i</sup>
90 °C	n/r	n/r	n/r	n/r	n/r
Refold	0-0.03	0.58-0.43	0.08*	0.49-0.31	0.026-0.024

n/r = no results obtained from the deconvolution. \*The same value was calculated for both deconvolutions. <sup>i</sup>NRMSD >0.05 suggest deconvolution errors and results are rejected.

It has been previously established that increasing the temperature of the experiment is seen to decrease the chromophores present, ultimately resulting in a decrease in overall protein concentration. This was reinforced through the precipitate found in the solution following the temperature aggregation assays. The deconvoluted summary of the spectra as seen in Table 4.4 reveals there is a clear trend in the calculated  $\alpha$ -helix secondary structure. At 20 °C, *h*PFD subunit 6 has 65-58 % of  $\alpha$ -helical content and with the increase in temperature to 30 °C has a slightly higher calculated percentage range. Further temperature increases is seen to slowly decrease the helical content of the subunit and at 70 °C is seen to reach 44 %. In comparison, an increase in unordered content is observed following thermal aggregation. *h*PFD subunit 5 is calculated to have 19-10 % of unordered structure at 20 °C and this calculated secondary structure is seen to either increase or remain constant with the increase of CD analysis temperature. At 70 °C, the disordered structure is calculated at 31-26 %. Unfortunately the deconvolution of the spectrum at 90 °C could not be obtained due to the low MRE recorded. Following thermal aggregation, the sample was incubated on ice for 24 hours and a far UV CD analysis was conducted. The deconvoluted calculated  $\alpha$ -helical content of the cooled sample was 0-3 %, unordered structure was calculated to be 49-31 % and  $\beta$ -sheet content was calculated to be 58-43 %. It is interesting to see that *h*PFD subunit 5 has adopted a structure with high  $\beta$ -sheet content, especially as the calculated  $\beta$ -sheet content following thermal aggregation is seen to fluctuate between 6-19 % with no clear trend. It is interesting to see that the percentage of turns remains constant throughout the assay, retaining between 9-14 %. It is concluded that the deconvoluted results of *h*PFD subunit 5 with increasing temperatures reveal decreasing fractions of  $\alpha$ -helical content which become disordered. This ultimately shows that the conformation of the subunit is sensitive to temperature and heat, in particular the coiled coils of the chaperone.

Similar to the previous thermostability assay, *h*PF<sub>2</sub> subunit 6 was subjected to thermal stress prior to far-UV CD data acquisition. Figure 4.7 reveals that between the temperature ranges of 20-50 °C no change in the CD spectrum is observed, suggesting that the structure of *h*PF<sub>2</sub> subunit 6 is largely unchanged at these temperatures. However, the maximum observed for the curves relating to the solution temperature of 20-30 °C is seen to occur at 198 nm, indicative of some  $\beta$ -sheet structure. The increasing solution temperature to 40-50 °C is seen to shift the maximum to 193 nm. This is similar to the effects seen for *h*PF<sub>2</sub> subunit 5 and it is likely that the structure of *h*PF<sub>2</sub> subunit 6 is unfolding in a similar manner.

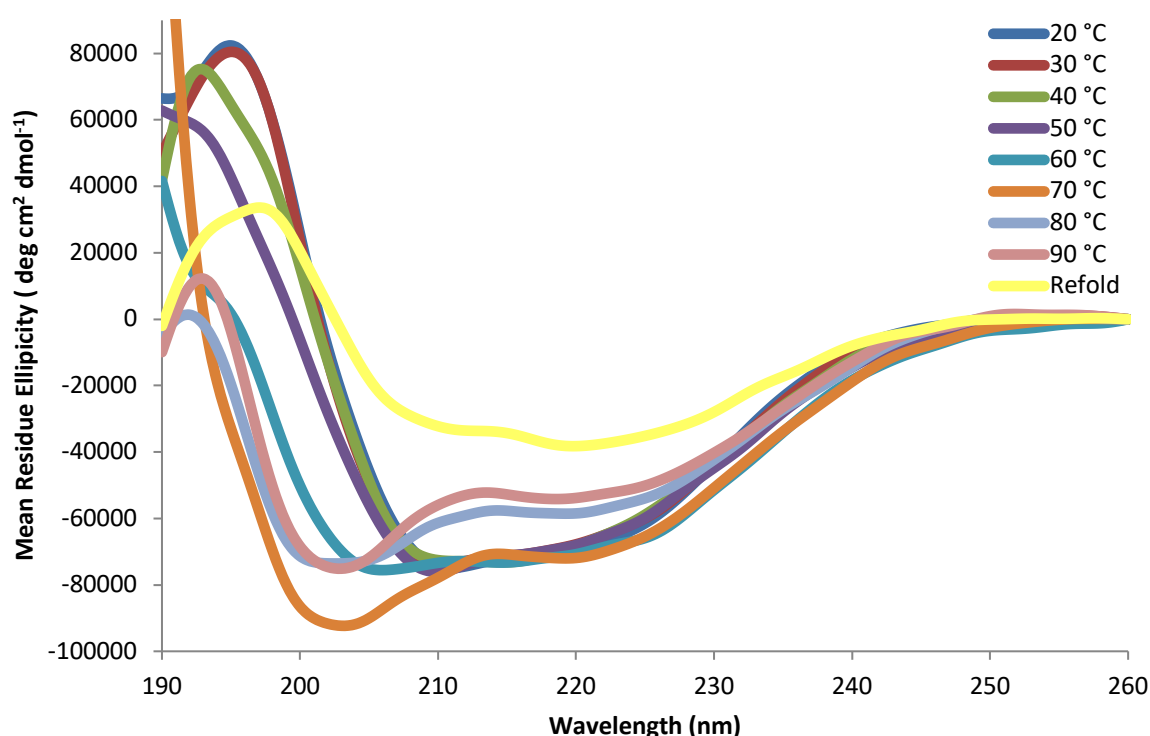


Figure 4.7: Far-UV CD spectra of *h*PF<sub>2</sub> subunit 6 (10  $\mu$ M) in 10 mM NaPO<sub>4</sub>H<sub>2</sub> at pH 7, between the temperature range of 20-90 °C. Following thermal denaturation, the sample was incubated on ice for 24 hr & far-UV CD spectrum was measured (refold). Representative data are shown from at least two independent experiments performed.

Interestingly, the CD curves seen for temperatures 60 °C and beyond reveal a shift of the minimum. A normally observed minimum for  $\alpha$ -helices at 208 nm is now seen reaching  $\lambda$  203 nm and the isodichronic point is seen at approximately 208 nm. Like *h*PF<sub>2</sub> subunit 5, this suggests a local 2 state population, however the disodichronic point is not consistent with reports of  $\alpha$ -helical and unordered structures [367]–[369]. A search of the literature reveals that the minimum exhibited at

203 nm refers to secondary structures that are relatively uncommon such as  $\pi$ -helix,  $\beta$ -bridge, bends, and irregular loops [370].

The magnitude of MRE at  $\lambda$  203 and 222 nm for 80 and 90 °C indicates a decrease in signal intensity, therefore indicating protein precipitation. A maximum at 193 nm is also observed for the the curves relating to 80-90 °C, proposing  $\alpha$ -helical structure as one of the local 2 state populations. The sample was removed following thermal denaturation and cooled on ice for 24 hours. Precipitate was observed in the sample and the pellet removed prior to analysis by CD (Figure 4.7, refold curve). The CD refold curve reveals an overall decreased quantity of chromophores as indicated by the MRE signal intensity. However minima are still observed at 208 and 222 nm and a maximum is seen at 198 nm.

From the data presented here, it is concluded that little structural changes are observed for *h*PF<sub>D</sub> subunit 6 between the temperature range 20-50 °C. At 60 °C the signs of disordered structures appear and at 70-90 °C it is proposed a 2 state population occurs for *h*PF<sub>D</sub> subunit 6. Thermal aggregation of the subunit at 80-90 °C is seen to result in protein precipitation. Similar to *h*PF<sub>D</sub> subunit 5, the refold curve reveals small amounts of structured  $\alpha$ -helix present and chromophores of  $\beta$ -sheets, turns and disordered aggregates are not seen.

To determine the changing conformations of *h*PF<sub>D</sub> subunit 6 through thermal stress, deconvolution of the CD data presented in Figure 4.7 were input into DichroWeb and the CDSSTR analysis revealed calculations of high  $\alpha$ -helical content. The results are summarised in Table 4.5.

**Table 4.5: Calculated secondary structure fractions & fitting parameters of *h*PF<sub>D</sub> subunit 6 incubated at different temperatures, utilising the software available at DichroWeb. The range is calculated from two replicate experiments.**

Temperature	Alpha helix	Beta sheet	Turns	Unordered	NRMSD
20 °C	0.76-0.61	0.11-0.08	0.13-0.08	0.13-0.08	0.003-0.001
30 °C	0.71-0.69	0.13-0.10	0.08*	0.12-0.11	0.003-0.002
40 °C	0.63-0.58	0.16-0.12	0.11-0.07	0.19-0.15	0.003-0.002
50 °C	0.58-0.52	0.25-0.17	0.09-0.07	0.18-0.17	0.004-0.002
60 °C	0.47-0.41	0.27-0.12	0.15-0.07	0.32-0.20	0.018-0.001
70 °C	0.20*	0.34-0.28	0.13-0.09	0.38-0.36	0.046-0.012
80 °C	0.22-0.17	0.27-0.22	0.15-0.13	0.42-0.41	0.050-0.037 <sup>i</sup>
90 °C	0.18-0.17	0.26-0.14	0.15-0.14	0.54-0.42	0.066-0.058 <sup>i</sup>
Refold	0.08-0.07	0.41-0.37	0.14-0.10	0.42*	0.062-0.033 <sup>i</sup>

\*The same value was calculated for both deconvolutions. <sup>i</sup>NRMSD >0.05 suggest deconvolution errors and results are rejected.

The  $\alpha$ -helical content is seen to decrease throughout the thermal aggregation assay, whilst the  $\beta$ -sheet content and unordered structure is observed to increase and the number of turns is seen to remain relatively stable. Unfortunately, the NRMSD range for the data sets obtained at 80 °C, 90 °C



and following the refold are calculated to be equal to and greater than 0.05. This is synonymous with problems in the deconvolution and as a consequence the secondary structure calculations for these are rejected. Table 4.5 reveals that the largest structure change is seen to occur between the temperatures 60-70 °C. This follows the trend seen in Figure 4.7, where the chromophores exhibited for *h*PF<sub>D</sub> subunit 6 at 70 °C is seen to shift the minimum and maximum to the left of the graph. The deconvolution reveals that at this temperature, *h*PF<sub>D</sub> subunit 6 has the lowest fraction of  $\alpha$ -helical chromophores at 20 % and the highest fraction of  $\beta$ -sheet and disordered structures, 34-28 and 38-36 % respectively. Similar to *h*PF<sub>D</sub> subunit 5, the calculated fractions of turns remain fairly constant throughout the thermal aggregation experiment with fluctuations up to 7 %. A previous study has found that the CD spectra of uncommon secondary structures such as  $\pi$ -helix,  $\beta$ -bridge, bends, irregular loops and possible invisible regions of the structure are categorised by CDSSTR into the unordered fraction of the protein, despite having a structured conformation, albeit uncommonly seen [370]. It is therefore possible that the deconvoluted results of 80-90 °C have profiles exhibiting the characteristic  $\lambda$  for these structures and due to the limited reference sets available for CDSSTR, deconvolution of the data at these temperatures are limited.

To follow the unfolding profile of the highly  $\alpha$ -helical structure of the *h*PF<sub>D</sub> subunits, the relative MRE were normalised and plotted as a function of temperature to allow a direct comparison of the thermostability of *h*PF<sub>D</sub> subunit 5 and 6 is seen (Figure 4.8). Protein precipitation is generally characterised by an overall loss in exhibiting chromophores and this was seen to occur for both *h*PF<sub>D</sub> subunits. As the highest  $\alpha$ -helical content calculated for the subunits were in solutions of 20 °C, the succeeding MRE obtained at 222 nm were normalised against this value acquired at room temperature. Figure 4.8 shows that the melting profile of *h*PF<sub>D</sub> subunit 5 is seen to have a steady decrease in normalised MRE signal. This ultimately indicates protein precipitation and the simultaneous decrease of  $\alpha$ -helicity with the increase in temperature. At 60 °C the chromophores of *h*PF<sub>D</sub> subunit 5 reveal the signal has depleted to approximately 40 %. This signal is seen to decrease further at higher temperatures, although to a lesser extent and reaches a constant value at temperatures 80-90 °C. The refold of *h*PF<sub>D</sub> subunit 5 reveals a signal intensity of 35% relative to the *h*PF<sub>D</sub> substrate at 20 °C.

In comparison, *h*PF<sub>D</sub> subunit 6 shows small negligible changes in the relative MRE data between the temperature ranges of 30-50 °C, indicating the retention of  $\alpha$ -helices and no protein precipitation. It is proposed that at these temperatures, the overall stability and structure of *h*PF<sub>D</sub> subunit 6 remains relatively constant. Interestingly, Figure 4.8 shows that the normalised MRE values for *h*PF<sub>D</sub> subunit 6 at 222 nm are seen to increase at temperatures 60-70 °C. This increase in relative MRE is higher

than the exhibited signal at 20 °C, and is anomalous and not observed in the literature. It is suggested that hidden  $\alpha$ -helical chromophores were unveiled following the application of the thermal stimulus to the *h*PFD subunit 6.

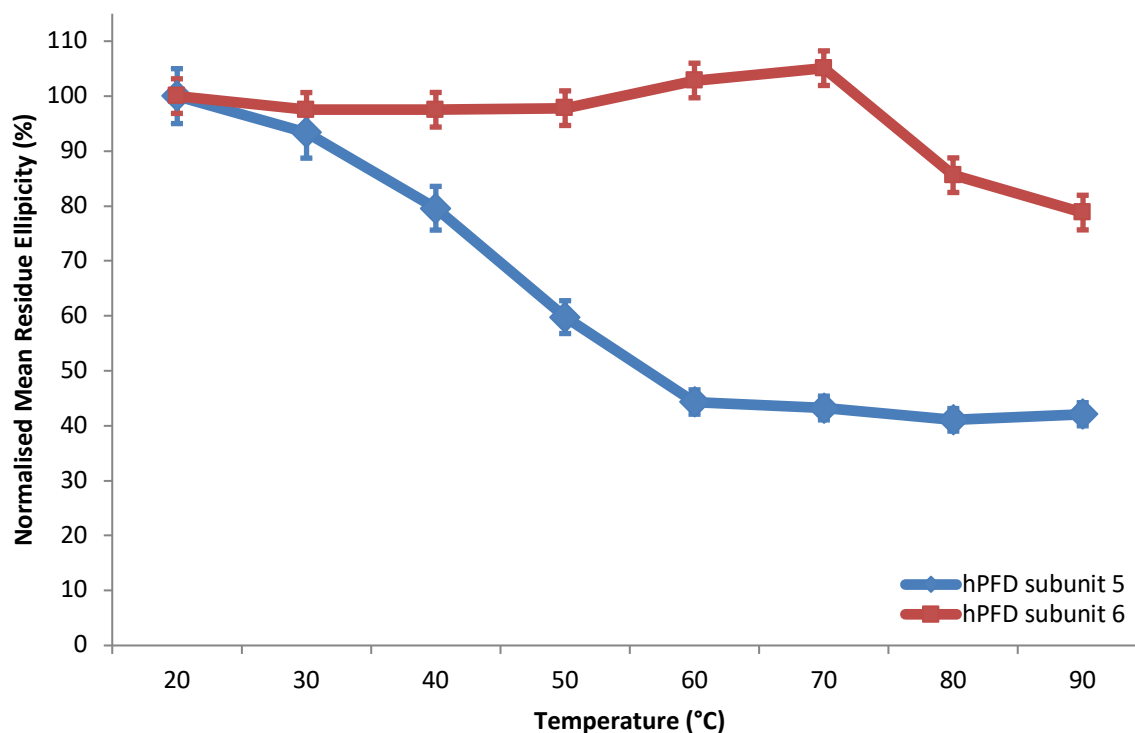


Figure 4.8: Melting profiles of *h*PFD subunit 5 (blue) & *h*PFD subunit 6 (red) observed from the change in ellipticity at 222 nm as a function of temperature. The data presented here are changes in ellipticity at 222 nm as a function of temperature and the values were normalised against the MRE observed at 20 °C, with +/- SD from 2 independent repeats of the experiment.

Figure 4.8 ultimately reveals the resistance of the *h*PFD subunits and in particular subunit 6 to changes in secondary structures with thermal stimulation. The normalised MRE seen at 70 °C for *h*PFD subunit 6 suggests that the protein retains a high content of  $\alpha$ -helices, with little precipitation. The melting profile of *h*PFD subunit 6 suggests that  $\beta$  type subunit sections of the complex are more stable in harsh conditions than  $\alpha$  type subunits. Unfortunately, a HT trace was not conducted to show an increase in scattering with temperature; however it is known that insoluble aggregates were present in the solution due to the precipitation viewed after the experiment. Melting profiles of the different species of prokaryotic PFD seem to differ [85], [86], [94]. Previous temperature dependent studies of *mt*PFD showed *mt*PFD  $\beta$  subunit dissociated before *mt*PFD  $\alpha$  subunit, while *mj*PFD  $\alpha$  subunit was found to unfold before *mj*PFD  $\beta$  subunit, suggesting *h*PFD has a closer

association to *mj*PFD in terms of thermal aggregation. These results propose that the peripheral  $\beta$  type subunits of eukaryotic PFD protect the inner  $\alpha$  type subunits from the effects of thermal stress.

A summary of *h*PFD subunit 6 at temperatures 60-90 °C suggests that at these temperatures the molecule unfolds in a way that maintains some  $\alpha$ -helices of the branching coils but at the same time induces unusual structural chromophores previously unexposed at lower temperatures by CD spectrometry. The isodichronic points at 203 and approximately 208 nm for *h*PFD subunits 5 and 6 reveals a 2 state transition with thermal denaturation and ultimately implies that a 2 state folding transition exists for the substrates. These results align with the transitions seen by MS for thermal denaturation of prokaryotic *mt*PFD subunits [94] and it ultimately gives insight into the folding mechanism of the nascent *h*PFD subunits.

It is curious to see that the deconvoluted results reveal both *h*PFD subunits 5 and 6 have the highest amount of calculated  $\alpha$ -helices and the lowest amount of calculated unordered structure at 30 °C. Ideal physiological temperature for activity is 37 °C and it has previously been proposed that the biological activity of the *h*PFD subunits are likely to be more efficient at slightly higher temperatures (Section 3.2.4.1), however the conformation does not necessarily need to be fully formed to have the most efficient chaperoning abilities due to the suggested chaperoning mechanism. It is likely that at sub-optimal cellular temperatures, the *h*PFD subunit coiled coils adopt their tightly wound  $\alpha$ -helical conformation. Higher temperatures increase the likelihood of protein misfolding and it is proposed that the flexible coiled tentacles lose some  $\alpha$ -helical structure in anticipation of chaperoning. To determine the possible products formed from the chaperoning effect of the *h*PFD subunits, far-UV CD was used following the thermal aggregation of co-incubated A $\beta$ (1-40) and the *h*PFD subunits.

#### **4.2.4: The Products of A $\beta$ (1-40) Following Incubation with *h*PFD Subunits 5 and 6 Monitored by Far-UV Circular Dichroism**

The conformational changes of monomeric A $\beta$ (1-40) following incubation at 37 °C were followed by far-UV CD analysis between 190-260 nm. Solutions of A $\beta$ (1-40) at concentrations of 50  $\mu$ M in PBS were incubated at 37 °C. Before incubation an aliquot was removed and diluted to 10  $\mu$ M with milliQ water. The sample was briefly centrifuged to remove any aggregates and the supernatant was analysed by far-UV CD. This was repeated after 24 hours and 43 hours of incubation following complete fibrillation. Figure 4.9 reveals the different CD spectra observed for the aggregating peptide before, during and after the incubation period.

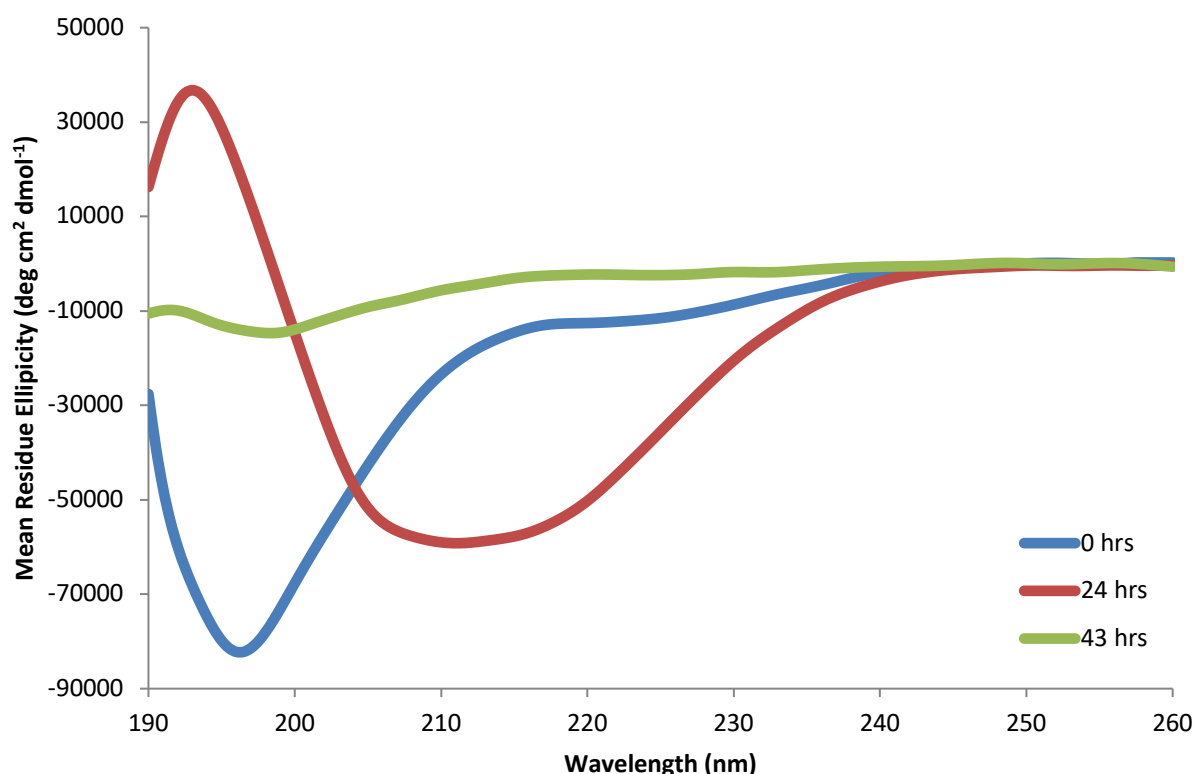


Figure 4.9: Far-UV CD spectra of the soluble products following incubation of A $\beta$ (1-40) (50  $\mu$ M) at 37  $^{\circ}$ C in a buffer of PBS over a 43 hr time period. Representative data are shown from at least two independent experiments performed.

The far-UV CD data of monomeric A $\beta$ (1-40) before incubation (labelled 0 hr) reveals the characteristic features of the intrinsically disordered monomer with a minimum displayed at 195 nm. It should be noted that a small minimum is also observed at 222 nm and this may be due to the A $\beta$ (1-40) adopting a partial  $\alpha$ -helix fold, consistent with previous studies [371].

Following incubation of A $\beta$ (1-40) at 37  $^{\circ}$ C for 24 hours, the peptide is observed to have CD data characteristic of  $\alpha$ -helix and  $\beta$ -sheets with an increase in signal at 222 nm and a maximum and minimum observed at 198 and 215 nm. This is consistent with disordered monomer nucleating to form structured soluble oligomers as detailed in Section 1.2.2. Although it is likely that there are some intrinsically disordered structures still in solution, this is not observed in Figure 4.9, presumably due to the dominance of other secondary structures present in the sample.

Following complete fibrillation of A $\beta$ (1-40) (as followed by ThT), CD measurements were taken of the soluble protein. Figure 4.9 reveals a weak signal following 43 hours of incubation at 37  $^{\circ}$ C. Although the magnitude of MRE is relatively small, a minimum is observed at approximately 195 nm, indicating that a small amount of soluble unordered monomeric A $\beta$ (1-40) remains.

The secondary structures of the products formed by incubation of A $\beta$ (1-40) at 37 °C for 43 hours in the presence of either *h*PFD subunit 5 or 6 at a ratio of 5:1 (A $\beta$ :PFD) were investigated. Precipitation formed was removed by centrifugation and soluble proteins were investigated by far-UV CD, with spectra plotted in Figure 4.10.

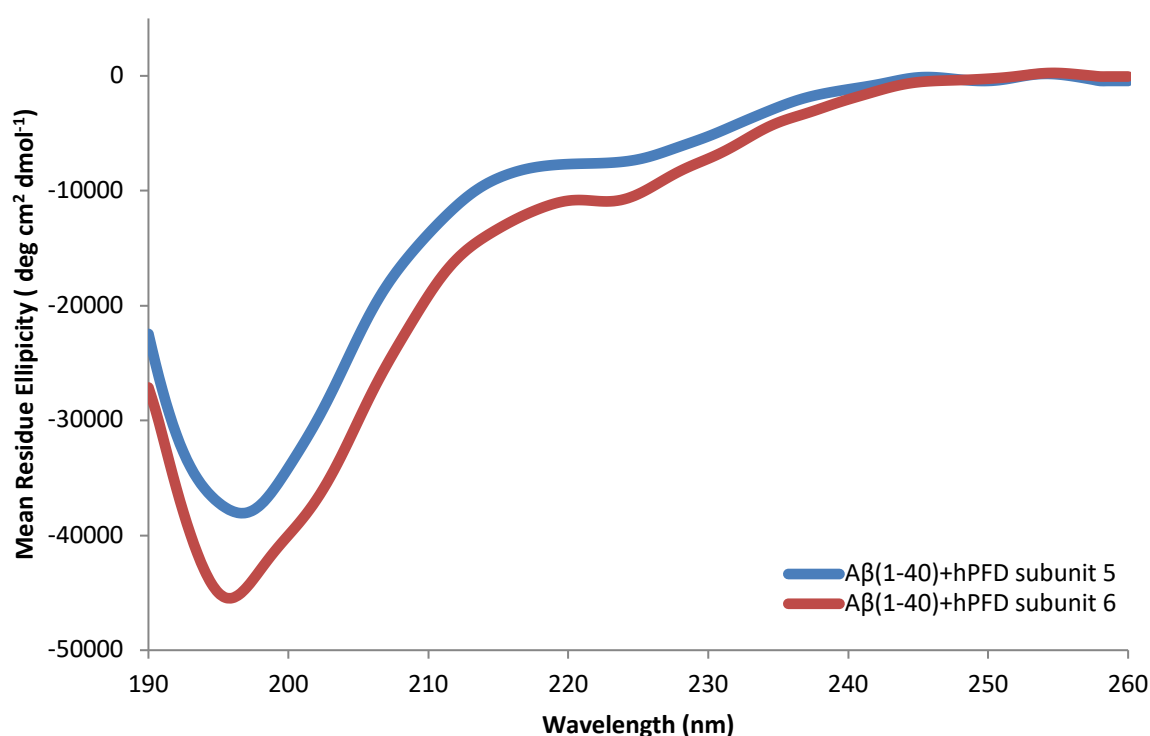


Figure 4.10: Far-UV CD spectra of the soluble products following co-incubation of A $\beta$ (1-40) & *h*PFD subunits 5 & 6 at 37 °C for 43 hrs. Representative data are shown from at least two independent experiments performed.

In the presence of both *h*PFD subunit 5 and 6 minima are observed at 195 and 222 nm, indicative of random coil and  $\alpha$ -helix structures respectively. It is difficult to determine whether the features present in the CD spectra are indicative of A $\beta$ (1-40), the respective *h*PFD subunits or a new product formed by the co-incubation. However, due to the low ratio of the *h*PFD subunits it is assumed that the spectra obtained reflects minimal amounts of the chaperone. It is therefore suggested that the minimum at 222 nm is indicative of the  $\alpha$ -helical structures in chaperoned A $\beta$ (1-40) and that the minimum at 195 nm is indicative of the disordered structure of monomeric A $\beta$ (1-40).

Comparison of the MRE of the products in Figure 4.10 show higher content of absorbing chromophores found in the co-incubation of *h*PFD subunit 6 and A $\beta$ (1-40), possibly indicating a higher concentration of soluble chaperoned substrate. Although, it was established in Section 3.2.4.1

that the chaperones do not allow fibril formation to occur at these ratios, it does not necessarily mean amorphous aggregates and eventual precipitation do not occur. The difference in the MRE magnitude of the chaperoned products (Figure 4.10) suggests that this is a highly likely scenario. It is proposed that while the *h*PFD subunits are able to perform their local chaperoning effects on the misfolding substrate, the suggested bind and release mechanism also does allow the amorphous aggregation to occur.

If the curves in Figure 4.9 are to be compared to the curves of Figure 4.10, then the magnitude of MRE observed in each case can illustrate the relative abundance of disordered monomeric A $\beta$ (1-40) with the co-incubation of a *h*PFD subunit in relation to the fibrillated A $\beta$ (1-40). The magnitude of MRE in the CD curves reveal an overall decrease in the abundance of the intrinsically disordered peptide in comparison to the 0 hour monomeric A $\beta$ (1-40) solution. As mentioned previously, co-incubation of A $\beta$ (1-40) and the respective *h*PFD subunit at the ratios implemented do not result in fibril formation (Section 3.2.4.1). It is therefore likely that small amounts of *h*PFD subunit and some A $\beta$ (1-40) may have aggregated, in a non-fibrillar, amorphous way and precipitated out of solution. It is also apparent that this effect is more prominent for the substrates incubated with *h*PFD subunit 5 than *h*PFD subunit 6.

Deconvolution of the spectra obtained for aggregating A $\beta$ (1-40) and chaperoned A $\beta$ (1-40) by CDSSTR reveals an overall high  $\beta$ -sheet and disordered secondary structure content. Unfortunately, a comparison cannot be made between the different conformations of the amyloid peptide aggregating in the absence and presence of the *h*PFD subunit, as the results summarised in Table 4.6 reveal the NRMSD ranges are over 0.05 and must therefore be rejected.

**Table 4.6: Calculated secondary structure fractions & fitting parameters of *h*PFD subunit 6 from CD spectra obtained at different temperatures. Algorithm CDSSTR & reference set SMP180 are selected for DichroWeb deconvolution.**

	Alpha helix	Beta sheet	Turns	Unordered	NRMSD
A $\beta$ (1-40) monomers	0.12-0.07	0.24-0.17	0.22-0.12	0.55-0.43	0.004-0.003
A $\beta$ (1-40) oligomers	0.62-0.45	0.15-0.12	0.08-0.04	0.34-0.15	0.002-0.001
A $\beta$ (1-40) fibrils	0.08-0.06	0.32-0.28	0.19-0.15	0.50-0.41	0.051-0.027 <sup>i</sup>
A $\beta$ (1-40) + <i>h</i> PFD subunit 5	0.23-0.18	0.33-0.10	0.19-0.11	0.49-0.37	0.140-0.027 <sup>i</sup>
A $\beta$ (1-40) + <i>h</i> PFD subunit 6	0.19-0.18	0.24-0.09	0.17-0.14	0.55-0.43	0.126-0.081 <sup>i</sup>

\*The same value was calculated for both deconvolutions. <sup>i</sup>NRMSD >0.05 suggest deconvolution errors and results are rejected.

The deconvoluted CD results summarised in Table 4.6 for aggregating A $\beta$ (1-40) are consistent with previous structure determinations for the amyloid peptide [58], [151], [152], [155], [156], [371]–

[376]. Although there are papers that report oligomeric A $\beta$ (1-40) has more structured  $\beta$ -sheet content similar to fibrils [57], [377], [378], it is likely that the oligomeric results seen here are the soluble prefibrillar  $\alpha$ -helical structures before conversion to the  $\beta$ -sheet oligomers. It is curious to see that after 43 hours of incubation, the deconvoluted secondary structures of the A $\beta$  fibrils are similar to monomeric A $\beta$ . This is likely due to the remaining soluble monomeric A $\beta$ (1-40) at equilibrium following fibril formation. Unfortunately, fibrils precipitate out of solution and subsequently cannot be observed by CD spectroscopy. This has occurred here, resulting in a rejection of the deconvolution.

The results seen in Table 4.6 for the co-incubation of A $\beta$ (1-40) with either *h*PFD subunit 5 or 6 must also be rejected due to the high NRMSD values. The results here are therefore used only as a guide to demonstrate similar percentages of secondary structure to the monomeric A $\beta$ (1-40). The  $\alpha$ -helical content is calculated to be approximately 20 %, whereas  $\beta$ -sheets were found to be at approximately 30-10 %, turns at 20-10 % and disordered structure at approximately 45 %.

To conclude, the results in seen here validate the previously suggested interaction of eukaryotic PFD intercepting the prefibrillar peptide and obstructing further aggregation. The combination of the data in this Chapter and Chapter 3 suggest that a prefibrillar monomeric intermediate species of A $\beta$ (1-40) chaperoned by *h*PFD is likely to be the primary cause of the cytotoxicity previously observed, rather than the unchaperoned disordered monomers or the pre-fibrillar  $\beta$ -sheet oligomers.

## 4.3: Experimental

### 4.3.1: CD Assays

CD spectra were acquired in the far-UV range using a Jasco-815 spectropolarimeter (Jasco, MD, USA) at room temperature using a 0.1 cm path length quartz cuvette. Samples were diluted to a concentration of 10  $\mu$ M in 10 mM NaPO<sub>4</sub>H<sub>2</sub>. The concentrations of proteins were determined spectrophotometrically by measuring the absorbance at 280 nm and using a calculated extinction coefficient (Refer to Appendix E). Each spectrum was obtained by averaging five scans and subtracting the buffer contribution. For thermal aggregation assays, samples were incubated at the specified temperature for a minimum of 3 minutes for equilibration before acquisition. Assays were repeated a minimum of 2 times to ensure data obtained was consistent.

### 4.3.2: pI Computational Calculation of *h*PFD Subunits

The sequence of *h*PFD subunits 5 and 6 were input into an open server available on uniprot; [http://web.expasy.org/compute\\_pi/](http://web.expasy.org/compute_pi/) to determine the theoretical pI of the protein and is seen in Appendix E.

### 4.3.3: Dichroweb Deconvolution of CD Spectra

Raw data recorded from Jasco-815 spectropolarimeter (Jasco, MD, USA) was input into the open software available at <http://dichroweb.cryst.bbk.ac.uk/html/home.shtml> to deconvolute the spectrum. The analysis programme selected was CDSSTR and the reference set chosen was SMP180. The optional scaling factor was set to 1.0 and the output unit was selected to be of MRE.



# Chapter 5:

## Nuclear Magnetic Resonance

### Analysis of the Human Prefoldin

### Subunits 5 and 6

---

#### 5.1: Introduction

Although structure determination of proteins and nucleic acids through NMR spectroscopic analysis has been insightful into structure-function relationships, this technique is still inherently limited. Many proteins are exempt from being analysed by this technique due to the strict requirements for analysis. Sample preparation for solution NMR requires specific conditions to obtain spectra that are sufficient for analysis and modelling. Typically, the protein of interest is synthesised to high concentrations and buffered into phosphate buffers that allows good NMR spectra, whilst simultaneously stabilising the protein. Depending on the specific information wanted, different NMR experiments are acquired. Following acquisition of good spectra, NMR interpretation can be long and tedious and unfortunately not all outcomes result in clear protein models [369], [379]. Although NMR spectroscopy is a challenging technique for structure determination, it is a desirable method due to the high resolution and information rich data it is able to obtain.

The first step to acquiring good protein NMR spectra is the preparation of the sample. Recently, protein production has been driven by developments in genome-scale proteomics to include vectors, infinity tags and host cell lines. *E. coli* based expression systems are most favoured, due to the rapid cloning and production of proteins at a minimum cost. Although protein production is often a rate-limiting step, these expression systems have made the process of producing and purifying large amounts of protein for structural studies simpler and faster. Depending on the type of NMR experiments conducted, some protein preparations require metabolic enrichment of molecules in the three major isotopes  $^2\text{H}$ ,  $^{15}\text{N}$  and  $^{13}\text{C}$ , achieved by inclusion of isotopically enriched chemicals the protein expression system [380], [381].

Although NMR spectroscopy acquires high resolution data, the technique is intrinsically poor in sensitivity and as a consequence, protein samples need to be at concentrations within the millimolar range. Typically, for a conventional room temperature probe, the sample concentration is ideally around 1 mM, however samples can be as low as 0.5 mM [382]. Fortunately, cryogenically cooled probes are able to acquire data at reduced concentrations of approximately 0.2 mM and depending on the sample can be reduced to as low as 0.05 mM. The concentrated protein sample is generally buffer exchanged into a solution that promotes long-term stability, although there are some buffer salts more convenient for NMR spectroscopy than others [382].

Phosphate buffers are commonly used in sample preparations for NMR spectroscopy as they typically do not interact with the protein of interest and do not give rise to signals in the NMR spectra. Buffers containing covalently bonded protons typically induce sharp and obtrusive signals that compete with the protein signals displayed in the spectra, especially as buffer concentrations are typically 10 fold higher or more than the protein concentrations, usually between 10-50 mM [382]. Stability of the NMR spectrometer is typically maintained through the continuous monitoring of a deuterium solvent peak. Therefore, deuterated water is also added to the buffer at levels of 5-10 % [382]. Although buffer salts with labile protons are able to exchange with the deuterium from  $^2\text{H}_2\text{O}$  added to the solvent, the protons of water are found to have the most intense peak in the spectrum. Also, water alone forms hydrogen bonds with the carbonyl and amide groups of the protein, disrupting the regular intramolecular bonding of secondary structure. To counteract this, salts are added. Although the addition of salts, such as NaCl to the buffer is generally seen to increase protein solubility and decrease aggregation, high salt concentrations above 150 mM have been reported to decrease the signal intensity, particularly for cryogenically cooled probes and at high magnetic fields [383], [384]. This loss in signal coupled to poor sensitivity can be dramatic and as a consequence, buffer salts used for NMR sample preparation must therefore be carefully considered. The pH of the solution is also a factor that requires optimisation. Exchange of the backbone amide proton is base-catalysed and as a consequence, a high pH can result in rapid exchange rates. If the amide proton is swiftly exchanged throughout the NMR pulse sequence, it will not be detected and therefore no signals for the N-H correlations will appear. The exchange rate of backbone amide protons most conducive to observing NMR signals occurs when the pH values are between 3-7 [382], [385]. Unfortunately, protein conformations are affected by pH as observed in Section 4.2.2. As a consequence, the pH of the solution must be carefully considered.

As the solvent for protein samples of NMR are predominantly comprised of water, the proton peak for  $\text{H}_2\text{O}$  is usually the most intense signal in the acquired spectrum. Fortunately, there are several

ways to overcome this. The development of solvent suppression methods have been seen to reduce the intensity of the water signal. WET or WATERGATE sequences are based on selective dephasing of the solvent signals using gradient pulses [385], [386]. Presaturation, on the other hand suppresses water signals during the relaxation delay, through the application of continuous low power radio frequency radiation at the same frequency of the solvent peak. As this is the easiest process available, it is frequently used in protein NMR [143], [386]–[389].

### 5.1.1: Aim

The aim of this chapter was to elucidate the structure of *h*PF<sub>D</sub> subunits 5 and 6 through high resolution 2D and 3D NMR techniques.

## 5.2: Results and Discussion

The development of multidimensional and heteronuclear NMR has allowed structural properties of proteins at molecular weights of 30-40 kDa to be elucidated [108]. As 3D NMR has been proven to be successful at elucidating macromolecules, this technique was applied to *h*PF<sub>2</sub>D subunit 6. However, the 3D NMR studies of the *h*PF<sub>2</sub>D subunit 6 were unsuccessful and 2D NMR advances were then made. Due to the quicker acquisition time of 2D NMR experiments, following the 2D NMR spectroscopy experiments were conducted.

### 5.2.1: 3D NMR Spectroscopy

Due to the smaller size and molecular weight of *h*PF<sub>2</sub>D subunit 6, labelling and subsequent 3D NMR spectroscopy was first conducted on this subunit before further investigations of *h*PF<sub>2</sub>D subunit 5 expression in *E. coli* follows the method detailed in Section 3.3.3 with several key changes required for metabolic enrichment of the isotopes <sup>15</sup>N and <sup>13</sup>C. The large overexpression LB broth media is replaced with a M9 minimal media as detailed in Section 5.3.1. The overnight expressed bacterial cells were re-cultured in the M9 minimal media containing <sup>15</sup>NH<sub>4</sub>Cl and <sup>13</sup>C<sub>6</sub>H<sub>12</sub>O<sub>6</sub> as the sole nitrogen and carbon sources, respectively. As the isotope labelled chemicals required for the media were relatively expensive, expressions of the subunit in minimal media was first acquired with unlabelled chemicals and growing conditions were trialled to determine the maximum amount of possible cell expression. Varying growing conditions such as temperature, shaking, aeration, time of induction, and time of harvest were trialled and eventually the optimal growing conditions were found to be at 30 °C, with shaking at 180 rpm, in unbunged conical flasks, induction occurred at OD<sub>600</sub> readings of approximately 0.4-0.6, and left to grow overnight until the OD<sub>600</sub> reading was approximately 2.0. Following cell harvest, the protein was extracted from the cell and purified as previously detailed in Chapter 3. Details of the vector, DNA transformation and purification of *h*PF<sub>2</sub>D subunit 6 can be seen in Section 3.2.3 and 3.3.3.

To determine the best solvent for solution NMR, several buffers were trialled with unlabelled *h*PF<sub>2</sub>D subunit 6 to determine stability of the protein at room temperature. It was found that in buffers with no salt, aggregation of the *h*PF<sub>2</sub>D subunit occurred and protein precipitation was seen within 1 hour. Concentrations of NaCl were added to the sample at 50 mM, 100 mM and 150 mM and after a 24 hour incubation period, it was found that in the solution containing 50 mM NaCl the protein had precipitated, 100 mM NaCl had minor precipitation and 150 mM NaCl had no precipitation observed. As a consequence it was determined that salt was needed at concentrations of 150 mM or higher to allow protein solubility and stability.

A previous successful study of *ph*PFD probed by  $^{13}\text{C}$  NMR used a solution of 50 mM MES-HCl, 150 mM NaCl, 25 mM  $\text{MgCl}_2$ , 0.02%  $\text{NaN}_3$ , and 10%  $^2\text{H}_2\text{O}$  (v/v) at pH 6 [390] and a variation of this buffer was used. Following labelling of the protein, the sample was placed into a buffer of 20 mM MES, 150 mM NaCl, 25 mM  $\text{MgCl}_2$ , 10%  $^2\text{H}_2\text{O}$  (v/v), pH 6 at a concentration of 1 mM and inserted into a 3 mm NMR tube (Shigemi, Pennsylvania, USA).

The first NMR experiment run for the  $^{15}\text{N}$  and  $^{13}\text{C}$  labelled *h*PFD subunit 6 was the  $^1\text{H}$ - $^{15}\text{N}$  heteronuclear single quantum coherence (HSQC) NMR experiment. This experiment is able to show all H-N correlations and therefore should contain at least one peak for each residue in the protein. Consequently, this spectrum can provide an excellent high-resolution 'fingerprint' of the protein and is predominantly used to assign backbone amide groups [391]. A close up of the H-N region of the spectrum obtained for *h*PFD subunit 6 is shown in Figure 5.1.

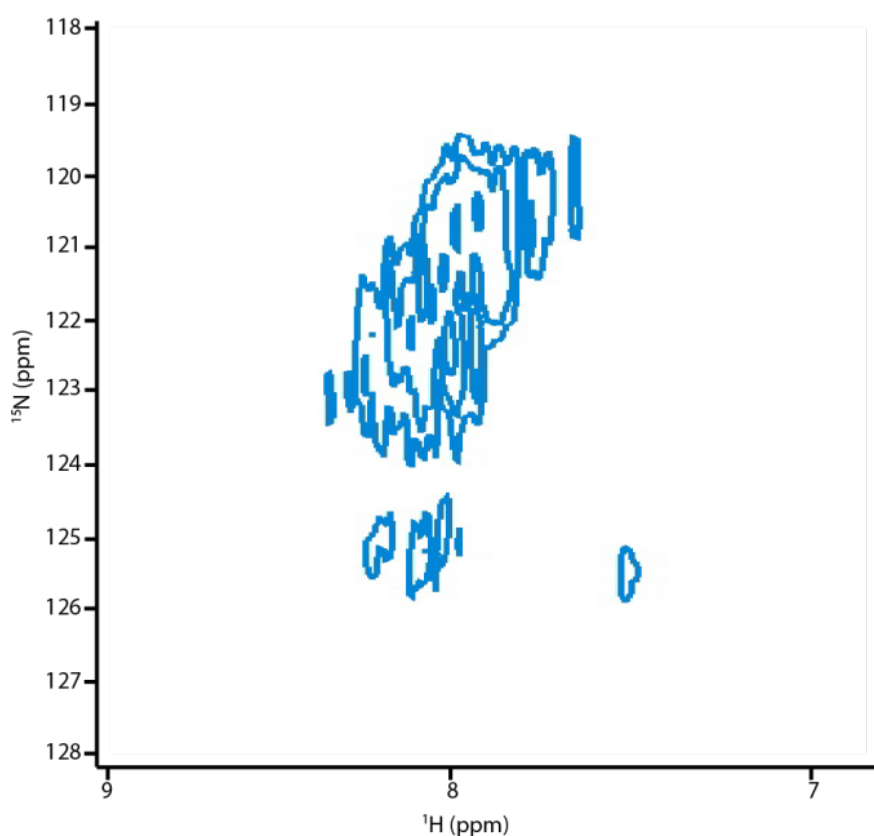


Figure 5.1: 2D  $^1\text{H}$ - $^{15}\text{N}$  HSQC NMR spectrum of  $^{15}\text{N}$  &  $^{13}\text{C}$  labelled *h*PFD subunit 6, close up of the H-N region.

The peaks in the  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrometry were found to be poorly resolved. Figure 5.1 illustrates the section observed with the majority of the peaks, which are seen along the  $^1\text{H}$  axis at

approximately 8 ppm, and along the  $^{15}\text{N}$  axis ranging from approximately 107-130 ppm. Peaks observed at approximately around 110-115 ppm on the  $^{15}\text{N}$  axis typically correspond to Asn and Gln side chain  $\text{NH}_2$  groups and therefore result in 2 peaks along the same  $^{15}\text{N}$  axis due to the 2 attached protons in the amide group, *h*PFD subunit 6 has 9 Arg amino acids and the peaks seen at 111-113 ppm  $^{15}\text{N}$  and 7-8 ppm  $^1\text{H}$  may correspond to these residues. Unfortunately, the clumping of the signals ultimately means that peaks are not well defined and resolved.

A  $^1\text{H}$ - $^{13}\text{C}$  HSQC experiment was also performed following acquisition of  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum. These experiments are similar and magnetisation is transferred from the hydrogen nuclei to the attached  $^{13}\text{C}$  nuclei via J-coupling, instead of on the  $^{15}\text{N}$ . A close up of the observed peaks in the  $^{13}\text{C}$ -HSQC spectrum for *h*PFD subunit 6 is seen in Figure 5.2.

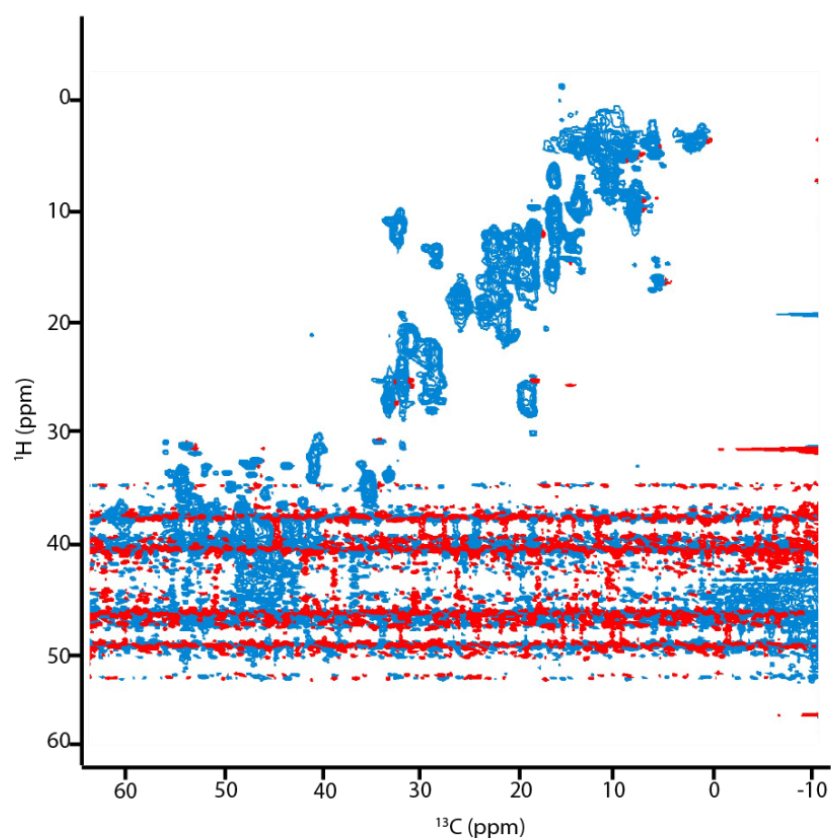
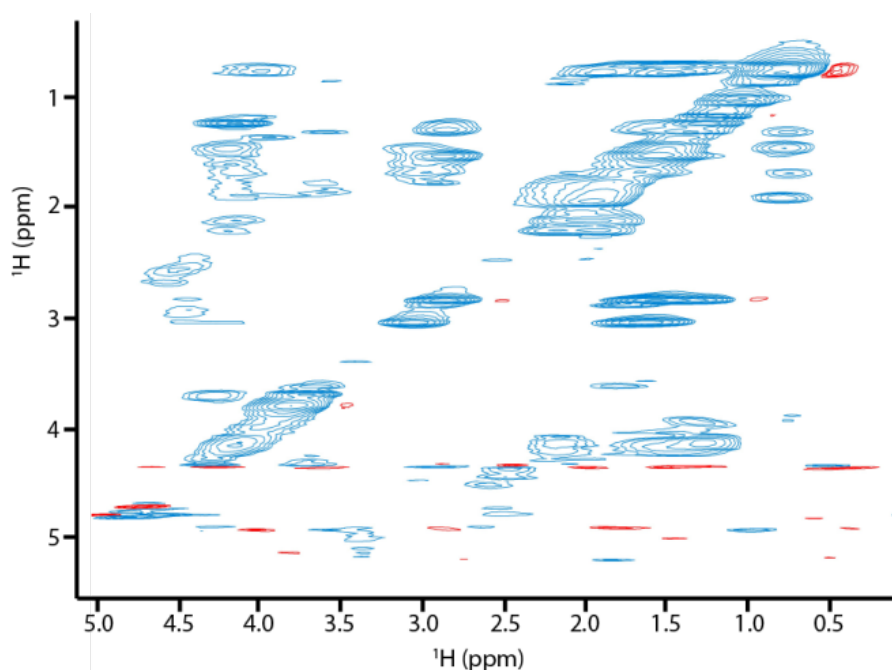


Figure 5.2: 2D  $^1\text{H}$ - $^{13}\text{C}$  HSQC NMR spectrum of  $^{15}\text{N}$  &  $^{13}\text{C}$  labelled *h*PFD subunit 6.

Figure 5.2 reveals that some signals in the  $^1\text{H}$ - $^{13}\text{C}$  HSQC are well defined, while others are overlapped. Unfortunately it can also be observed that over 1/3 of the peaks seen in the spectrum overlap with the water peaks introducing ambiguity to the assignment of the peaks. It must also be

noted that the  $^{13}\text{C}$  HSQC spectrum is theoretically meant to show all H-C correlations in the protein, therefore 180 H-C backbone signals should be seen. Figure 5.2 reveals a low number of signals and clumping of signals similar to Figure 5.1. The low quality spectrum seen here is likely due to a combination of sample and spectrum related problems. A comparison between the  $^{15}\text{N}$  and  $^{13}\text{C}$  HSQC show that the peaks in the  $^{13}\text{C}$  HSQC spectrum are relatively more defined, despite literature stating that the opposite should be true [392].

Following HSQC NMR experiments, a 3D HCCH-TOCSY NMR experiment was performed. This experiment is typically used for the assignment of side-chains as a  $^{13}\text{C}$  mixing scheme correlates all the  $^1\text{H}$  and  $^{13}\text{C}$  nuclei in a spin system. The magnetisation evolves on the side chain  $^1\text{H}$  nuclei to the attached  $^{13}\text{C}$  nuclei which is then mixed and transferred back to the  $^1\text{H}$  for detection. Each carbon frequency strip should therefore display all side chain  $^1\text{H}$  resonances in the spectrum [393].



**Figure 5.3:** 3D HCCH-TOCSY NMR spectrum of  $^{15}\text{N}$  &  $^{13}\text{C}$  labelled *hPFD* subunit 6. The 3<sup>rd</sup> axis  $^{13}\text{C}$  is combined and overlaid to reveal all peaks seen in the 3D spectrum.

Although the peaks in Figure 5.3 are seen to be better resolved than the peaks from the HSQC experiments, the displayed spectrum is overlaid with all peaks displayed along the  $^{13}\text{C}$  axis. This ultimately shows a small number of nuclei decay detected. Shimming of the sample in this experiment was difficult, suggesting sample problems with the experiment.

Problems with shimming are typically associated to sample problems such as protein aggregation. Although shimming in the previous experiment was difficult, a 3D CBCA(CO)NH NMR experiment was run to determine if magnetisation from the backbone could be detected. This experiment involves the transfer of magnetisation between  $^1\text{H}_\alpha$  and  $^1\text{H}_\beta$  to  $^{13}\text{C}_\alpha$  and  $^{13}\text{C}_\beta$  and then to  $^{13}\text{C}=\text{O}$ , followed by transfer to  $^{15}\text{N}-\text{H}$  and then to  $^1\text{H}-\text{N}$  for detection. The chemical shift on  $^{13}\text{C}_\alpha$  and  $^{13}\text{C}_\beta$  appear in one dimension as they are simultaneously evolved, and the chemical shifts of  $^{15}\text{N}-\text{H}$  and  $^1\text{H}-\text{N}$  are evolved in the other 2 dimensions [394].

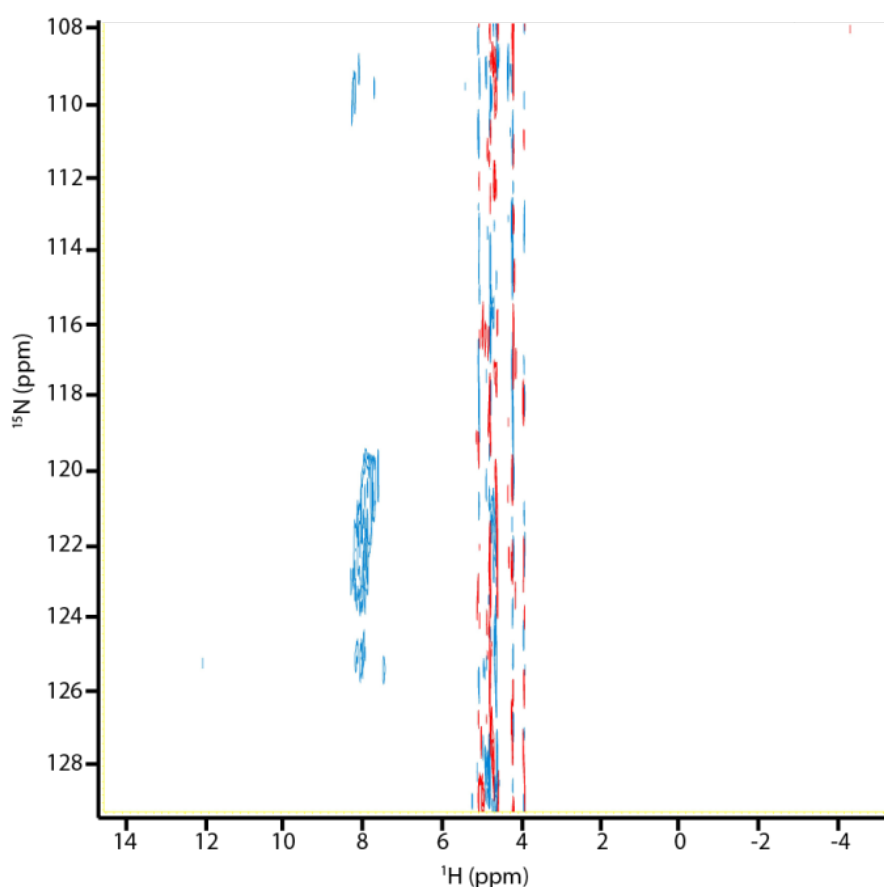
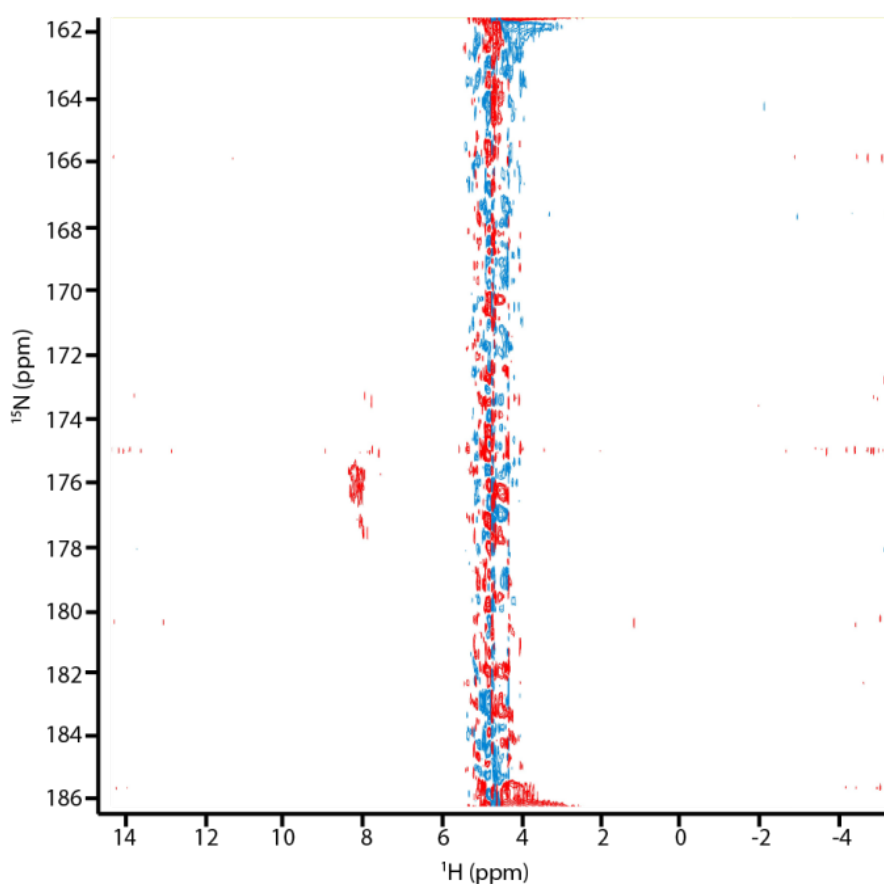


Figure 5.4: 3D CBCA(CO)NH NMR spectrum of  $^{15}\text{N}$  &  $^{13}\text{C}$  labelled *h*PFD subunit 6. The 3<sup>rd</sup> axis  $^{13}\text{C}$  is combined and overlaid to reveal all peaks seen in the 3D spectrum.

Figure 5.4 reveals a clumped signal, this is similar to the peak seen in Figure 5.1, as the shifts between these experiments remain the same. Shimming of this 3D CBCA(CO)NH NMR experiment was found to be more difficult than in the 3D HCCH-TOCSY NMR experiment, and it is apparent from the spectrum that signal intensity is lost. A 3D HNCO NMR experiment following the CBCA(CO)NH NMR experiment was run to confirm possible sample precipitation.



The 3D HNCN NMR experiment was the most sensitive triple-resonance experiment available for 3D NMR spectroscopy and is able to assist in backbone assignment through the magnetisation transfer of backbone  $^1\text{H}$  nuclei to backbone  $^{15}\text{N}$  nuclei and then selectively to the  $^{13}\text{C}=\text{O}$ . Detection occurs when the magnetisation is then passed back through the backbone  $^{15}\text{N}$  nuclei to the  $^1\text{H}$  nuclei, resulting in the chemical shift evolving on all of the 3 nuclei. This NMR experiment is also able to show side chain correlations of Asn and Gln [394].



**Figure 5.5: 3D HNCN NMR spectrum of  $^{15}\text{N}$  &  $^{13}\text{C}$  labelled *hPFD* subunit 6. The 3<sup>rd</sup> axis  $^{13}\text{C}$  is combined and overlaid to reveal all peaks seen in the 3D spectrum.**

Figure 5.5 clearly reveals low protein signal intensities, however high intensities are observed for the water peaks. At this point total acquisition time for the *hPFD* subunit 6 sample was approximately 6 days. When the NMR tube containing the sample was removed from the NMR spectrometer it was clearly seen that precipitation had occurred over the course of the NMR acquisitions. This aligns with the decreasing signal observed in the sequential NMR experiments. This ultimately reveals that the sample conditions were not optimal for the labelled *hPFD* subunit 6 for the duration of the NMR

experiments performed. Although a freshly expressed batch of labelled protein could be used with each 3D NMR experiment conducted, this is a time consuming and expensive endeavour which unfortunately could not be conducted for this thesis.

### 5.2.2: 2D NMR Spectroscopy

Following the attempted 3D NMR spectroscopy of *h*PFD subunit 6, it was decided that a series of 2D  $^1\text{H}$ - $^1\text{H}$  NMR experiments for both *h*PFD subunits would be attempted. 2D  $^1\text{H}$ - $^1\text{H}$  NMR experiments have only 2 frequency dimensions with less evolution and mixing periods than 3D NMR experiments and as a consequence it is faster to obtain spectra. This ultimately means that the overall acquisition time for the 2D NMR experiments of the *h*PFD subunits should be relatively shorter than the acquired 3D NMR experiments and therefore can be acquired before protein precipitation is seen to affect the data acquired.

#### 5.2.2.1: 2D NMR Spectroscopy of *h*PFD Subunit 5

*h*PFD subunit 5 was obtained through recombinant protein expression, purified through an affinity column and buffer exchanged into 20 mM MES, 200 mM NaCl, 20  $\text{MgCl}_2$  and 10 %  $\text{D}_2\text{O}$  (v/v) at pH 6 at a concentration of 1 mM and inserted into a 3 mm Shigemi tube. A series of 2D  $^1\text{H}$ - $^1\text{H}$  experiments were run and Figure 5.6 illustrates the N-H to  $\text{H}_\alpha$ ,  $\text{H}_\beta$  and  $\text{H}_\gamma$  region of the 2D  $^1\text{H}$ - $^1\text{H}$  TOCSY NMR spectrum obtained.

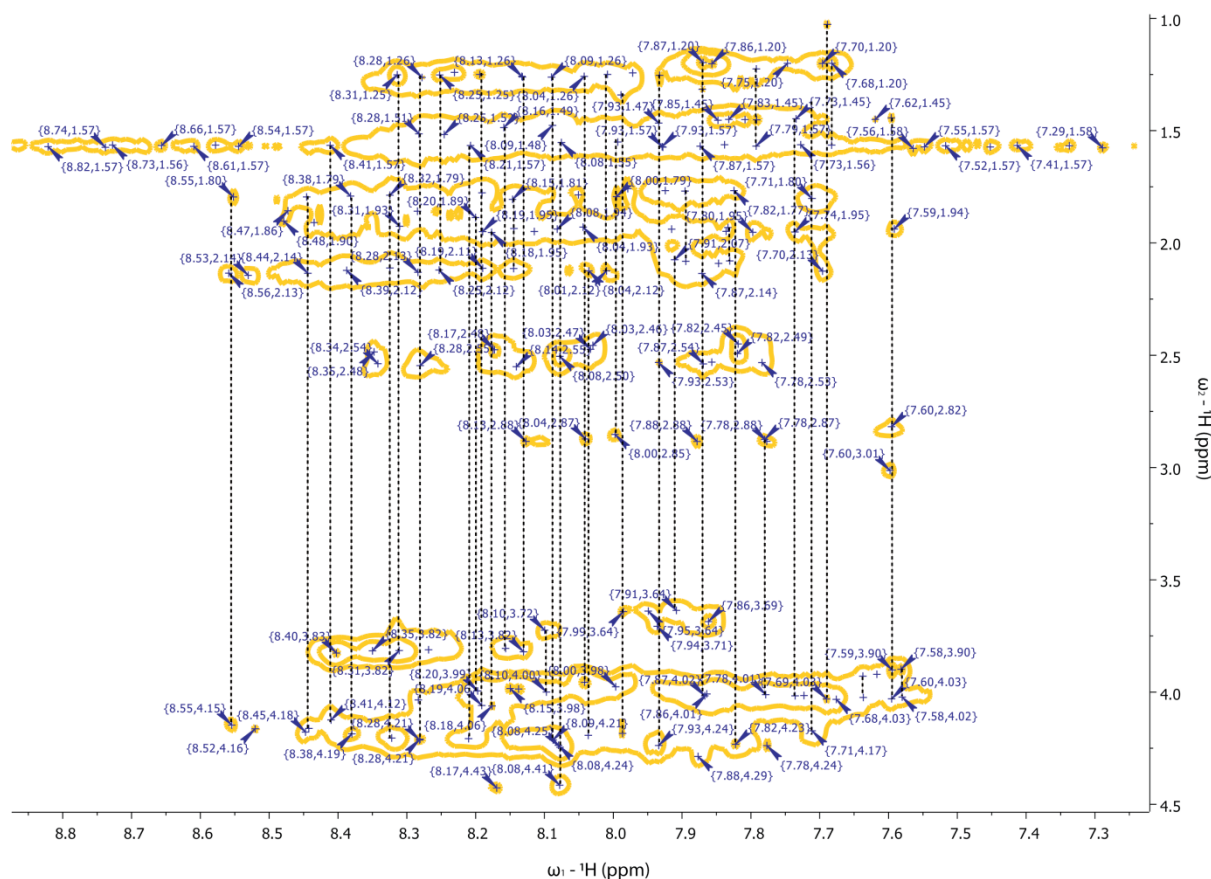


Figure 5.6: 2D  $^1\text{H}$ - $^1\text{H}$  TOCSY NMR spectrum of *hPFD* subunit 5, the N-H to  $\text{H}_\alpha$ ,  $\text{H}_\beta$  &  $\text{H}_\gamma$  region is shown with correlating signals of a spin system coordinated as indicated by the dashed line.

TOCSY NMR spectra are theoretically meant to show correlations of all protons in the amino acid residue, however experimentally, not all signals are observed in the spectra. It can be seen in Figure 5.6 that the detected peaks for the TOCSY experiment are well defined and extraction of the chemical shifts for each individual signal is relatively easy. However, it is also seen that there are 50 possible correlating spin signals. There are a total of 205 amino acids for *hPFD* subunit 5, and with the exception of Pro residues which are not displayed in this region, the spectra ultimately reveals peaks that are lost in background noise or are missing. There are also peaks that were not found to correlate to any other peaks in the spin system, although their presence generally signifies partial spin system detection. Unfortunately, the lack of correlated signals makes interpretation of the spectrum difficult. The absence of NMR signals can be due to a number of things. Although size was considered, the *hPFD* subunits are smaller than some previously NMR analysed proteins [108]. It is possible that signals of some residues could be overlapping, resulting in a peak for two possible amino acids. The likelihood of protein precipitation during acquisition is also conceivable, as previously seen in Section 5.2.1 despite the short acquisition time.

Despite less than a quarter of *h*PFD subunit 5 spin correlations shown and ambiguity present in the interpretation of the spectrum, assignment of resonances may give an insight into the structure amino acids. As previously mentioned in Section 2.2.2, sequential assignment of 2D NMR spectra is initiated through the determination of each spin system through comparison to published chemical shifts of random coil structure. To assist in residue determination of *h*PFD subunit 5, the signals detected in the 2D TOCSY NMR experiment were extracted and assigned to potential  $^1\text{H}$  nuclei, as summarised in Table 5.1. Unfortunately, the number of signals could mean that there are many Ala residues, but more likely, there are a significant number of signals that are missing.

Table 5.1 reveals all the correlating signals assigned to a possible corresponding proton signal in the spin system. As there are many missing signals, the assignment of residue to the spin system resonance is difficult, especially when not all resonances are observed in the TOCSY spectrum. 2D  $^1\text{H}$ - $^1\text{H}$  COSY NMR experiment was then performed; however the spectrum observed revealed peaks already established in the TOCSY and therefore did not assist in the assignment of the peaks (data not shown).

It is detailed in Section 2.2.2 that secondary shifts ( $\Delta\delta$ ) assist in the determination of secondary structures. The difference between the observed chemical shift and published random coil shifts can help determine the predominant structure for *h*PFD subunit 5.  $\alpha$ -helical structures have reported negative  $\Delta\delta$  values of  $^1\text{H}_\alpha$  and N-H, while  $\beta$ -sheet structures display positive  $^1\text{H}_\alpha$  and N-H  $\Delta\delta$  values [142]. The N-H chemical shifts published for residues in an unstructured random coil range between 8.0-8.42 ppm [142]. Despite the inability to assign the spin systems of *h*PFD subunit 5, secondary structure can be determined for resonances with N-H shifts outside the 8.0-8.42 ppm range. The first 3 spin systems in Table 5.1 have N-H shifts higher than 8.42 and are therefore likely to be of  $\beta$ -sheet structure. The last 24 residues in Table 5.1 have observed chemical shifts for N-H below 8.0 ppm resulting in a negative  $\Delta\delta$  and are therefore of  $\alpha$ -helical structure.

In the same manner, the  $\Delta\delta$  for  $\text{H}_\alpha$  can also be analysed. In an unstructured random coil, the chemical shift of  $\text{H}_\alpha$  was found to be between 3.96-4.74 ppm [142] and although there are no shifts in Table 5.1 above 4.74 ppm, there are 7 different spin systems with shifts below 3.96, resulting in at least 7 negative  $\Delta\delta$  values of  $\text{H}_\alpha$ . This ultimately means that there are 7 spin systems with  $\Delta\delta$  correlating to  $\alpha$ -helical secondary structure. If the spin systems already deduced to be of  $\alpha$ -helical structure from the  $\Delta\delta$  of N-H chemical shift are negated from the analysis, this leaves 4 spin systems with  $\alpha$ -helical characteristics. These 4 spin systems are therefore added to the 24 residues already determined to have  $\alpha$ -helical structure and in total, 28 spin systems correspond to residues that are in  $\alpha$ -helical formation. As there are a total of 50 separate spin systems identified in Figure 5.6, the

majority of the residues detected by  $^1\text{H}$ - $^1\text{H}$  2D TOSCY NMR spectroscopy of *h*PF<sub>D</sub> subunit 5 are of  $\alpha$ -helical structure. Unfortunately, the same method cannot be applied to the shifts seen for  $\text{H}_\beta$ , as the published chemical shifts range from 1.39-4.24 ppm and this range is too large to analyse the data.

**Table 5.1: Correlated  $^1\text{H}$  chemical shifts for *h*PF<sub>D</sub> subunit 5 extracted from 2D NMR TOSCY experiment.**

Chemical Shift (ppm)			
N-H	$\text{H}_\alpha$	$\text{H}_\beta$	Other H
8.55	4.15	1.8	
8.45	4.18	1.8	
8.44	4.16	2.14	
8.41	4.12	1.57	
8.38	4.19	1.79	
8.35	3.82	2.48	
8.32	4.2	2.11	1.79
8.31	3.82	1.93	1.25
8.28	4.21	4.04	2.55, 2.13, 1.86, 1.51, 1.26
8.25		2.12	1.52, 1.25
8.21	4.21	1.57	
8.2	3.99	1.89	
8.19	4.06	2.11	1.95, 1.25, 1.78
8.18	4.06	1.95	
8.17	4.43	2.48	
8.16	3.81	1.49	
8.15	3.98	1.81	
8.14	3.99	2.55	2.11, 1.94
8.13	3.82	2.88	1.26
8.1	4.00	3.72	
8.09	4.21	1.48	1.26
8.08	4.41	4.25, 4.24	2.50, 1.55, 1.94
8.04	4.19	3.96	2.87, 2.12, 1.26, 1.93
8.03		2.46, 2.47	
8.01		2.12	1.25
8	3.98	2.85	1.79
7.99	4.18	3.64	1.55, 1.34
7.93	4.24	2.53	1.47, 1.25, 1.57
7.92		1.94	1.77
7.91	3.64	2.07	
7.88	4.29	2.88	
7.87	4.02	2.54	2.14, 1.57, 1.32, 1.2
7.86	4.01	3.69	2.53, 1.2
7.85		2.09	1.45
7.84		1.95	1.56
7.83		2.08	1.45, 1.93
7.82	4.23	2.49	2.45, 1.77
7.79		1.57	1.45, 1.23
7.78	4.24	4.01	2.87, 2.88, 2.53
7.74	4.02	1.95	
7.73		1.56	1.45
7.71	4.17	1.8	
7.7		2.13	1.2
7.69	4.03	1.03	
7.68	4.03	1.56	1.2
7.64	4.02	3.93	
7.62	3.92	1.45	
7.6	4.03	3.01	2.82, 1.44
7.59	3.9	1.94	
7.58	4.02, 3.9		



### 5.2.2.1: 2D NMR Spectroscopy of *h*PFD Subunit 6

As there were many ambiguous peaks observed in the 2D NMR spectra for *h*PFD subunit 5, it was thought that the smaller *h*PFD subunit 6 would give better resolution in the 2D NMR spectra. In an identical manner as *h*PFD subunit 5, purified *h*PFD subunit 6 was subjected to 2D NMR TOSCY as seen in Figure 5.7.

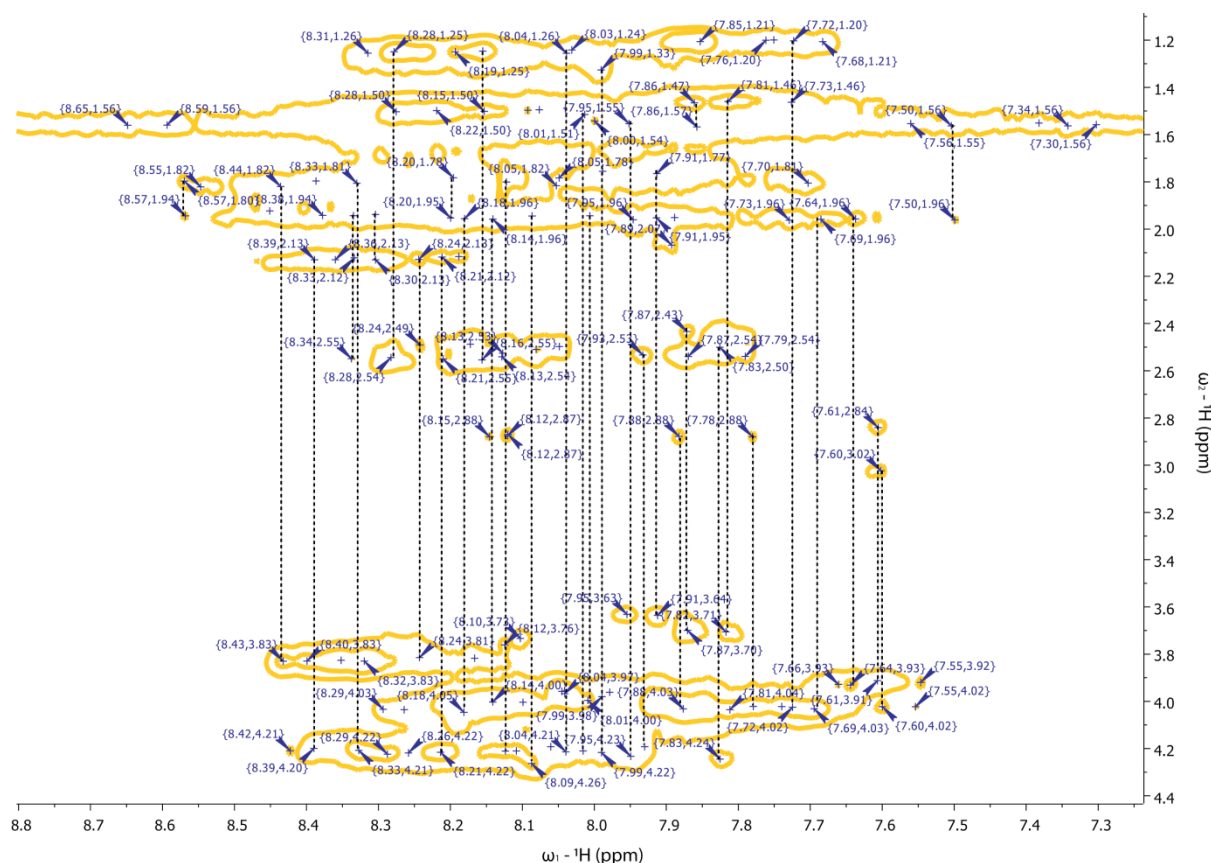


Figure 5.8: 2D  $^1\text{H}$ - $^1\text{H}$  TOCSY NMR spectra of *h*PFD subunit 6, the N-H to  $\text{H}_\alpha$ ,  $\text{H}_\beta$  &  $\text{H}_\gamma$  region is shown with correlating signals of a spin system coordinated as indicated by the dashed line.

Figure 5.8 reveals 34 correlating spin systems and these could be partial or full spin systems relating to the amino acids of *h*PFD subunit 6. The full sequence of the chaperone consists of 180 amino acids and therefore only 1/5 of correlating resonances are shown. This can be compared to the number of signals seen for *h*PFD subunit 5, which had approximately 1/4 of its residues displaying NMR resonances. Although the NMR analysis of both subunits reveals low numbers of detected NMR signals, the samples and NMR experimental parameters were prepared in the exact same conditions for and as a consequence it can be deduced that *h*PFD subunit 5 is a more soluble and

compatible substrate for analysis through NMR spectroscopy. The chemical shifts seen in Figure 5.8 were extracted and sorted into the possible correlating amino acid  $^1\text{H}$  nuclei signal. These are summarised in Table 5.2.

**Table 5.2: Correlated  $^1\text{H}$  chemical shifts for *h*PFD subunit 6 extracted from the 2D NMR TOSCY experiment.**

Chemical Shift (ppm)			
NH	H $_{\alpha}$	H $_{\beta}$	Other H
8.57		1.94	1.8
8.43	3.83	1.82	
8.39	4.2	1.8	
8.34		2.55	1.94
8.33	4.21	1.81	
8.3		2.13	
8.28		1.5	1.25
8.24	3.81	2.13	
8.21	4.22	2.26	2.12
8.2		1.95	1.78
8.18	4.05	1.96	
8.16		2.25	1.25
8.14	4	1.96	
8.12	4.21	3.76	2.87, 1.81
8.09	4.26	1.5	
8.08		2.51	1.49
8.04	4.21	3.97, 3.96	
8.01	4	1.51	
7.99	4.22	3.98, 3.96	1.77, 1.33
7.95	4.23	3.63	1.96, 1.55
7.92	4.18	2.53	
7.91	3.64	2.53	
7.88	4.03	2.88	
7.87	3.7	2.54	
7.86		1.57, 1.47	
7.83	4.24	2.5	
7.81	4.04	1.46	
7.78	4.01	2.88	
7.72	4.02	1.2	
7.69	4.03	1.96	
7.64	3.93	1.96	
7.61	3.91	2.84	
7.6	4.02	3.02	
7.5		1.96	1.56

In a similar manner to the analysis of *h*PFD subunit 5 by 2D TOSCY NMR spectroscopy, the assignment of residues cannot be completed due missing signals. However, from the N-H secondary shifts observed secondary structure of the partial spin systems can be determined. Table 5.2 reveals only one value above the highest chemical shift seen for an unstructured random coil amino acid. On the other hand, there are 16 N-H chemical shifts below the 8.0 ppm for random coil Ile [142]. Regardless of the residues identity, for the signals outside the range of 8.0-8.42 ppm, the  $\Delta\delta$  can be



calculated and therefore secondary structure can be assigned to these residues. The 16 N-H chemical shifts seen for TOSCY analysed *h*PF<sub>D</sub> subunit 6 below 8.0 ppm are calculated to have negative  $\Delta\delta$  and therefore are in residues with  $\alpha$ -helical structure.

Lastly, the  $H_\alpha$  chemical shift were analysed and 6  $H_\alpha$  shifts in Table 5.2 are identified to be below 3.96, therefore resulting in negative  $\Delta\delta$  for  $H_\alpha$ . 4 of these spin systems were already identified to be in residues of  $\alpha$ -helical content. Therefore, 2 more spin systems with  $\alpha$ -helical secondary shifts are added to the 16 spin systems, resulting in 18 total systems belonging to residues in  $\alpha$ -helical secondary structures. Like *h*PF<sub>D</sub> subunit 5, from these results it can be concluded that the majority of the residues detected by 2D NMR spectroscopy are within  $\alpha$ -helical structures and by extension, it implies and reiterates that *h*PF<sub>D</sub> subunit 6 is predominantly conformed to an  $\alpha$ -helix.

To see if further information could be extracted a 2D  $^1H$ - $^1H$  NOESY was obtained of *h*PF<sub>D</sub> subunit 6 following the TOSCY experiment and this is displayed in Figure 5.9. It is unsurprising to see here that Figure 5.9 reveals a spectrum with poor signal to noise, as the previous TOSCY spectrum was indicative of experimental problems for subsequent 2D NMR experiments.

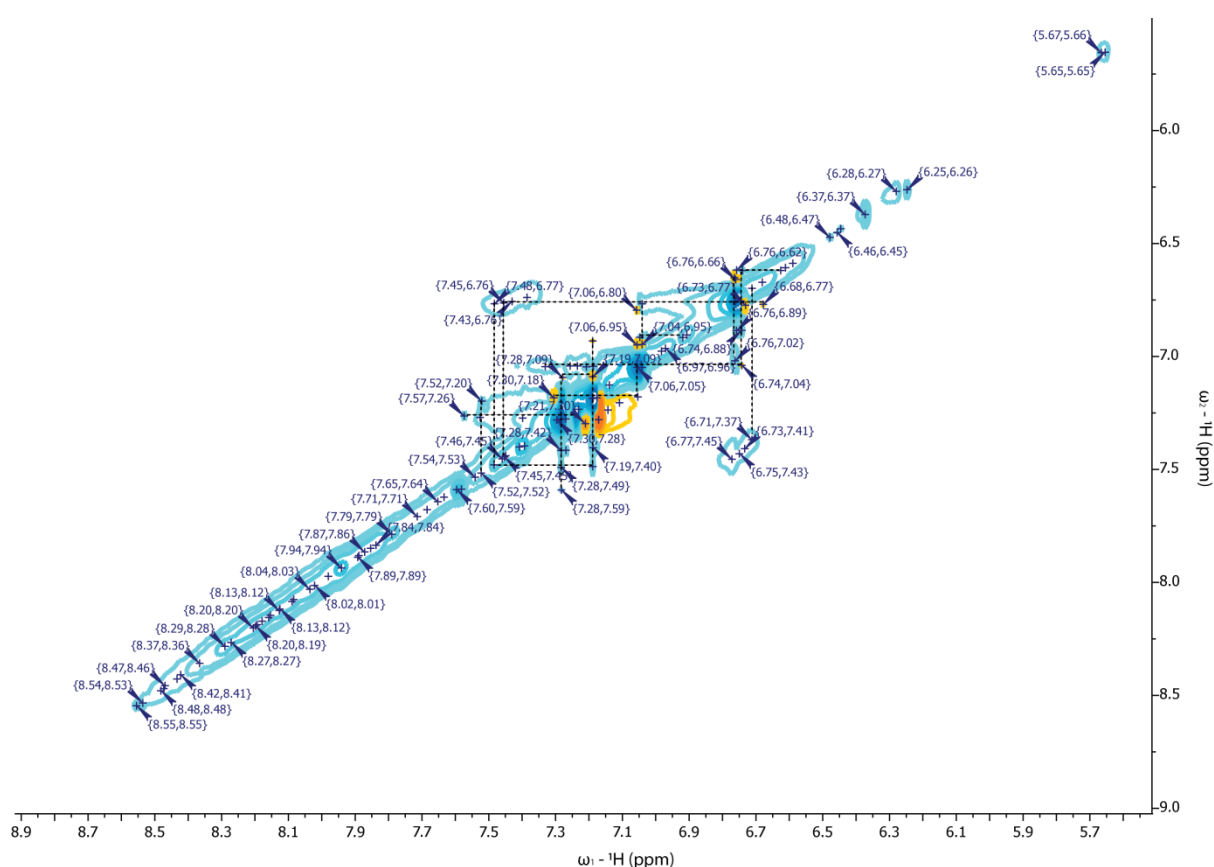


Figure 5.9: 2D  $^1H$ - $^1H$  TOCSY NMR spectra of *h*PF<sub>D</sub> subunit 6. NH to  $\alpha$ ,  $\beta$  &  $\gamma$  region of *h*PF<sub>D</sub> subunit 6.

Although 2D NMR experiments do not resolve NMR signals as well as 3D NMR experiments, they are often sufficient for the study of peptides up to approximately 35 kDa. Generally proteins and peptides greater than 35 kDa are of insufficient resolution in 2D NMR due to the greater number of resonances present in the spectra and the longer rotational correlation times resulting in broadened linewidths, although that was not seen here. *hPFD* subunits 5 and 6 are 22.8 and 20.1 kDa respectively; double the recommended size for 2D NMR, however it can be seen in Figure 5.4-5.8 that overlapping signals and broadened linewidths were not the difficulties present in the 2D NMR of the *hPFD* subunits. Instead, signals were limited and not all spin systems were present in the spectra, indicating flexibility, molecular dynamics or an unstructured complex. These points can also be applied to the incomplete NOE connectivity patterns observed in the NOESY spectra, in addition to signal overlap.

It has also been previously reported that the spectra of  $\alpha$ -helix rich proteins are less well dispersed than those from  $\beta$ -sheet-rich proteins due to the broader difference of chemical environments found in  $\beta$ -sheet structures [391]. The predominantly  $\alpha$ -helical structure of the *hPFD* subunits is likely to have induced overlapping signals to result in the low number of spin systems observed in the 2D spectra. It was also determined in Section 4.2.1 that the *hPFD* subunits have some regions of disorder. Although good NMR spectra have been obtained for both folded and completely unfolded proteins with well resolved resonances, proteins that are partially folded have been seen to often give rise to very poor quality spectra [391]. It was also reported that such partially folded proteins can exhibit substantial secondary structures in far-UV CD spectra, which is what has been seen here for the *hPFD* subunits. The poor quality NMR spectra can indicate the existence of a molten globule state and although the *hPFD* subunits may have well-formed secondary structures, the tertiary structure may not be as well-defined. Poor quality NMR spectra may also suggest a dynamic protein that is able to form oligomers.

Oligomerisation issues have been previously reported to obstruct NMR studies and resonance line broadening and subsequent poor sensitivity has been observed due to the dynamic properties of proteins fluctuating between multiple conformations at an intermediate rate on the chemical shift time scale [391], [395], [396]. It is possible that the poor overall signal-to-noise seen here for the *hPFD* subunits may be due to oligomerisation as it has been previously seen in prokaryotic PFD studies that homo-oligomeric assemblies occur for the PFD subunits [397].

Homo-oligomerisation of the *hPFD* subunits were further investigated in Section 6.2.2 and 6.2.3. The NMR results here are consistent with the dimers and trimers reported in Chapter 6 and this ultimately implies molecular dynamics are at play for the chaperone. The continuous interaction and

interconversion of the protein between different conformations and/or oligomeric states on the  $\mu$ s timescale can cause line broadening of signals, which can be seen for all 2D and 3D spectra of the hPFD subunits [391].

Although it was suggested that the size of the proteins, the rate of molecular tumbling, the conceivable molten globular state of the subunits and the possible dynamics of oligomerisation may have affected the NMR spectra, it should also be noted that precipitation was seen in the samples following both 2D and 3D NMR experiments. Therefore, in addition to these likely scenarios, protein aggregation did also occur and the hPFD subunits were unstable in NMR acquisition conditions.

It is possible that the limitations listed here may have impacted the number of PFD studies through NMR methods to elucidate the structure of the intact hexameric complex or its subunits. However, as previously mentioned, there is one published paper utilising the high resolution technique to determine the interactions of the distal tentacles of *ph*PFD to CPN [390]. Kato *et al.* was able to monitor the interaction between archaeal PFD and CPN through the selective carbon labelling of the terminal regions of the PFD coiled tentacles. Analyses were conducted using 1D  $^{13}\text{C}$  NMR on the partially labelled complex and, in comparison to the experiments conducted in this Chapter, there are many differences between the successful analysis and our data obtained. As Kato *et al.* focused on the terminal regions of the chaperone, fewer resonances were available for detection and any possible flexibility or molecular dynamics at play are possibility minimised. It was demonstrated on the basis of the  $^{13}\text{C}$  NMR data, that the N-terminal portions of both  $\alpha$  and  $\beta$  subunits of *ph*PFD possess significant degree of freedom in internal motions. The simple 1D NMR and the overall reduced amount of resonances results in a substantially smaller acquisition time, in comparison to the substantial amount of acquisition time needed for the analyses in this Chapter. This ultimately means that the 1D  $^{13}\text{C}$  NMR analysis of the *ph*PFD is likely to have no or low amounts of protein aggregation.

Some proteins that yield rather poor NMR spectra because of aggregation or low solubility might give excellent crystals [398], [399]. The results in this Chapter conclude that hPFD subunits 5 and 6 are not suitable for NMR spectroscopy, and by extension, eukaryotic PFD may be too large and unstable to be studied through NMR. Previously published reports utilising crystallography to determine *mt*PFD and *ph*PFD structures suggest that this is a more feasible approach to studying hPFD [84], [397]. Unfortunately, crystallography was not explored in this thesis and low resolution techniques were utilised instead.

## 5.3: Experimental

### 5.3.1: $^{15}\text{N}$ and $^{13}\text{C}$ Labelled *hPFD* Subunit Expression

All chemicals and materials were purchased from Sigma Aldrich (Missouri, United States) unless otherwise stated. All equipment was autoclave sterilised prior to use and all stock solutions were syringe filtered (0.2  $\mu\text{m}$ , GE Healthcare, Little Chalfont, United Kingdom).

A stock solution of 5xM9 minimal media (1 L) was prepared with  $\text{Na}_2\text{HPO}_4$  (33.9 g),  $\text{KH}_2\text{PO}_4$  (15 g),  $\text{NaCl}$  (2.5 g) and made up to 1 L with MQ  $\text{H}_2\text{O}$ . 1xM9 minimal media was prepared with 5xM9 (200 mL), 0.1 M  $\text{CaCl}_2$  (1 mL), 1M  $\text{MgCl}_2$  (2 mL), 20 %  $\text{C}^{13}$  labelled glucose (w/v, 20 mL), 8.5 % yeast nitrogen base without amino acids or ammonium sulfate (w/v, 20 mL), 10 %  $\text{N}^{15}$  labelled  $\text{NH}_4\text{Cl}$  (w/v, 15 mL) and made up to 1 L with MQ  $\text{H}_2\text{O}$ . 1xM9 minimal media was divided in half (2x500 mL) and into unbunged 2 L conical flasks.

DNA encoded plasmids were transformed and overexpressed in overnight cultures as detailed in Section 3.3.3. Following an overnight overexpression in LB broth (50 mL) and ampicillin (50  $\mu\text{L}$ ) at 37  $^\circ\text{C}$ , the supernatant was removed by centrifugation (5000  $g$ ) and decanted. The pellet was resuspended in 1xM9 minimal media (50 mL) and then split evenly into 2 50 mL falcon tubes. Each resuspension of cells were inoculated into 1xM9 minimal media (500 mL) into unbunged 2 L conical flasks and ampicillin (500  $\mu\text{L}$ ) and grown with shaking (180 rpm) at 30  $^\circ\text{C}$  for approximately 5 hours until the  $\text{OD}_{600}$  reached 0.4-0.6. IPTG (250  $\mu\text{L}$ , 1M) was added to the medium and the culture continued to grow with shaking overnight. Cells were then harvested by centrifugation (5000  $g$ ) for 10 minutes. Protein extraction and purification followed the same routine as the cells in Section 3.3.3.

### 5.3.2: NMR Sample Preparation

Following protein expression, extraction and purification, the sample was prepared for NMR experiments through buffer exchange into 20 mM MES, 150 mM  $\text{NaCl}$ , 25 mM  $\text{MgCl}_2$  at pH 6, and placed into a 3 mm NMR tube (Shigemi, Pennsylvania, USA). 10%  $^2\text{H}_2\text{O}$  (v/v) was added and thoroughly mixed with a long needle and syringe (Sigma Aldrich, Missouri, USA).

### 5.3.3: NMR Parameters

NMR spectra were recorded using a Varian Inova-600 NMR spectrometer, with a  $^1\text{H}$  frequency of 600 MHz, a  $^{13}\text{C}$  frequency of 150 MHz and a  $^{15}\text{N}$  frequency of 60 MHz. Experiments were carried out at 25

°C and referenced to water (4.82 ppm). HSQC experiments were recorded for the assignment of  $^{15}\text{N}$  and  $^{13}\text{C}$  resonances using correlations with the attached  $^1\text{H}$ . These experiments consisted of 8 time-averaged scans over a total of 512  $t_1$  increments. 2044 data points were collected over a spectral width of 12019.2 Hz in the  $^1\text{H}$  dimension (F2). The  $^1\text{H}$  frequency domain was referenced to DSS at 0.0 ppm, whilst the heteronuclear dimensions were referenced indirectly [400]. The standard HCCH-TOSCY, CBCA(CO)NH and HNCO [401] pulse sequences from the VnmrJ library were used. 64 ( $t_1$ ) and 37 ( $t_2$ ) increments, each consisting of 8 transients were acquired over 2044 data points for the spectra. The carrier frequencies for  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  were set to 4.773, 47.362 and 118.861 ppm, respectively and spectral widths of 12019.2 Hz ( $^1\text{H}$ ), 12062.0 Hz ( $^{13}\text{C}$ ) and 1944.3 Hz ( $^{15}\text{N}$ ) were used.

The water signal in both TOCSY and NOESY experiments was suppressed by presaturation. This was achieved by application of a low power RF pulse at the frequency of the water resonance during the 1.1 second relaxation delay between scans. The water signal in the DQF-COSY spectra was suppressed using gradient methods [402]. TOCSY, DQF-COSY and NOESY experiments were acquired in the phase-sensitive mode, with time proportional phase incrementations in  $t_1$ . For each experiment, 8 to 16 time-averaged scans were acquired per increment, with a total of 200  $t_1$  increments for TOCSY and NOESY spectra and 512  $t_1$  increments for DQF-COSY spectra. The FID in  $t_2$  consisted of 2044 data points over a spectral width of 9592.3 Hz. NOESY spectra were acquired with a mixing time of 150 ms and the TOCSY pulse sequence included a spin-lock of 80 ms.

The spectra were processed using VNMR software (VNMRJ, Version 2.1, Revision B) and viewed by MestReNova (MestreLab Research, Version 6.0.2-5475)

# Chapter 6:

## Ion Mobility-Mass Spectrometry of Mammalian Prefoldin and the Human Prefoldin Subunits 5 and 6

---

### 6.1: Introduction

The combination of IM and MS increases the specificity and selectivity of an analysis compared to MS alone through the orthogonal separation of IM and allows the simultaneous determination of CSS and mass. The coupling of these techniques with soft ionisation methods such as nanospray or ESI permits the elucidation of protein-protein interactions, protein-ligand interactions and low resolution 'native' like structures through the preservation of weak non-covalent bonds. Although the addition of IM has been a more recent development to native complex analyses, soft ESI-MS studies have been prevalent since 1991, with one of the first published studies detailing the interaction between a cytoplasmic receptor and rapamycin as detected by native ESI-MS [403]. Since then, the exponential increase in mass and complexity range combined with dissociation techniques available in both solution and gas phase has permitted MS to define complex stoichiometry, subunit connectivity and ligand specificity in unprecedented detail. Systems analysed through 'native' ESI-MS range from metal peptides, ligand protein assemblies and drug binding [182], [403], [404]. Structural elucidation of proteasome, monoclonal antibodies, membrane protein complexes, intact ribosomes, and large capsid viruses have also benefitted from the development of ESI-MS [404]–[406]. The addition of IM to this well established technique further refines the structural data obtained for macromolecule complexes.

As previously described in Section 2.2.5, IM has the ability to deduce complex shape and size and separate molecules based on different structural conformations and folding states. As the addition of IM separates ions based on their ability to traverse a chamber filled with inert neutrals under the influence of a constant electric field; correlated multiply-charged complex ions have ATDs that adhere to a pattern and this enables easy spectral deconvolution of the MS data. This can also be used to resolve ambiguity present in MS spectra of isoforms [128].

Recent applications of native IM-MS have highlighted its ability to define protein complex stoichiometry, topology and the existence of *in vitro* sub-complexes including the structural arrangement of these sub-complexes [407]–[411]. Protein complex models constructed through native IM-MS are typically represented as spheroids until further structural information is input. As protein complexes are dynamic entities, often consisting of stable core complexes decorated by peripheral subunits, the addition of partial denaturation to the analysis can potentially reveal subunit vicinities, binding cooperativity and estimations of binding affinities. Domain structure or protein folding information garnered from other structure elucidating techniques or from further IM-MS experiments can be integrated with the IM-MS data to refine the shape of the complex. There are several IM-MS experiments that can elucidate subunit stoichiometry and these assays typically induce unfolding and protein dissociation. It should be noted ‘native’ IM-MS conditions does not necessarily equate to all detected ions representing the native protein conformation. Possible quaternary remodelling and unfolding can still occur even with parameters set to ‘native’ like conditions [412].

Gas phase activation detailed in Section 2.2.5.5, can be optimised to access multiple levels of multiprotein organisation. Parameter optimisation is critical for gas phase unfolding as an intact protein will undergo unfolding or ejection of a highly charged monomer from the stripped oligomer. Denaturation of the subunits by CID will give unfolded conformations and further gas phase activation can induce fragmentation of the protein backbone [411], [413]–[415]. Although gas phase dissociations have the ability to create protein contact maps, there are several drawbacks of the technique. It should be noted that gas phase denaturation techniques lack solvation and as such largely do not reflect the same dissociation pathway as in solution. It has also been observed that gas phase activation can induce simultaneous unfolding with the dissociated intermediate and substrate structure [94]. As solution effects are of primary biochemical interest, complementary solution phase methodologies are typically employed in addition to gas phase activation experiments to determine complex stoichiometry and sub-complex topology.

Solution phase denaturation methods are able to generate orthogonal compositions of the complex prior to IM-MS analysis as opposed to gas phase activation methods which can only be engaged within the instrument. These solution methodologies are designed to perturb protein interfaces and disrupt non-covalent interactions whilst retaining IM-MS resolution. Typically, addition of small amounts of organic solvents, acids and/or bases and salts to the sample solution elicit the formation of topologically informative sub-complexes. Organic solvents such as alcohols are used to induce

unfolding of the protein complex by disrupting the hydrogen bonding prevalent in secondary to quaternary structure. Depending on the type of alcohol used, the intramolecular hydrogen bonds are able to form new bonds between the introduced alcohol molecule and the protein side chains [416]. Increasing the number of alcohol molecules to the sample solution increases the propensity for hydrogen bonding disruption and therefore protein complex unfolding. However, different alcohols can have different effects depending on the protein and in some cases alcohols can assist in hydrogen bonding and structured conformers due to the small changes in dielectric constant and hydration of ionic groups [128], [417]. As a consequence, a relatively exhaustive trial-and-error search of potential solution conditions for optimal disruption is usually necessary for protein dissociation by organic solvents. Unfortunately, the addition of acids and/or bases also requires optimisation for ideal disruption. This process can be relatively exhaustive, however the data acquired from these experiments is still desirable as solution phase disruption methods are able to dissociate and break the intact complex at the weakest interface to give insights to stoichiometry connectivities and interactions.

As explained in Section 4.2.2 acids and bases are able to disrupt hydrogen bonds and ionic bonds which occur in salt bridges. The addition of excess  $H^+$  or  $HO^-$  is able to interact with these side chain groups in a double replacement action. Likewise, chaotropic agents and salts are able to disrupt salt bridges in a similar manner. Chaotropic agents are typically utilised in dissociation experiments to manipulate the ionic strength of the solution to disrupt the weakest bonds in the complex and these are often non-covalent bonds at protein binding interfaces [418]. While the phenomenon of solution phase denaturation methods by chaotropic agents, ionic strength and solution pH have been thoroughly explored with MS in the elucidation of complex stoichiometry [238], [410], [419]–[422], real-time thermal denaturation experiments analysed by IM-MS has not been thoroughly exploited in the determination of protein complex structure.

Increasing the temperature of the solution during IM-MS ionisation can reveal the conformation of stable intermediates at different temperatures and dissociation patterns of the substrates. These melting experiments disrupt hydrogen bonds and non-polar hydrophobic interactions through the increase of kinetic energy, causing the molecules to vibrate so rapidly and violently that bonds have the capacity to break. Although solution phase thermal aggregation coupled to IM-MS has not been widely explored, there have been some published papers coupling thermal dissociation methods to MS, and these experiments have successfully delineated solution phase equilibria of macromolecules [94], [423]. These experiments are a good representation of possible cellular conditions, particularly in times of elevated levels of thermal stress. As high intracellular



temperatures are closely related to disease [424], determining protein complex conformations at elevated temperatures is desirable. It is well known that a number of biologically important complexes undergo changes in structure and activity in response to thermal stimuli, such as the sHSPs. Therefore, following the thermally induced structural changes of PFD by IM-MS may give an insight to its behaviour at elevated temperatures.

Following native ESI-MS and disruption experiments, computational analysis of the IM-MS data is essential to construct viable protein network maps and topological information. Several levels are required to deduce a refined structure and this begins with the analysis of the MS spectra. This first stage focuses on the deconvolution of the MS spectra of the intact complex, the constituent subunits and the sub-complexes generated from denaturing assays. Interpretation of the MS data to obtain accurate mass measurements of the components within the spectra is a critical and challenging stage especially when discrepancies in mass calculations arise due to the attachment of water, salts, and/or buffer molecules to the complex during mass spectrometric analysis. These desolvation problems have been seen in previous studies [410]. Fortunately, MS deconvolution software has been developed to identify attachment of these small molecules and to calculate the masses of the nanospray ESI-MS datasets of protein complexes. The development of automated systems allows easy interpretation of the MS spectra, especially in convoluted spectra containing multiple overlapping species in  $m/z$  [211], [425]. There is also software dedicated to the compilation and construction of protein complex contact diagrams [410], [426].

Following MS data analysis, interpretation of the IM data is conducted in 2 different stages. The first stage involves the conversion of the drift times into CCS values and the second stages compares the averaged CCS value to the computationally calculated model protein architectures. As this thesis utilises the non-linear TWIM, a calibration curve is constructed using standard proteins with known CCS values to convert IM drift times to CCS values. This plot enables the generation of the CCS values of the sample protein complex from the ATD. Unfortunately, accuracy of the experimentally calculated CCS values of the sample complex is heavily independent on the goodness of fit of the calibration curve to the TWIM experimental data. Contributing factors to the CCS calculation error includes the standards used and the reproducibility of the IM measurements. Fortunately, the increasing number of available standards has greatly improved the error associated with CCS [409], [427] to the extent where modern TWIM extrapolated CCS values have an uncertainty margin of 8-10 % in comparison to CCS measurements produced by other IM analysers [413].

Although this particular method of CCS calculation considers the peak position of the ATD through the intensity-weighted centre of the distribution, it does not acknowledge the width of the ATD and

therefore neglects the Gaussian shape distribution, which could potentially be significant on the conformation of the protein. The ATD peak width is thought to be the combination of ion diffusion and conformational ensembles [428]. Ion diffusion typically occurs through the net movement of the ion during IM-MS analysis and the ATD peak width of a single isomeric species is principally related to ion diffusion. However, peak widths relating to conformational ensembles give insight into the complex assembly and are therefore necessary commodities for molecular modelling. A more recent approach to analysing IM data that integrates the width of the ATD due to conformational ensembles is the extraction of the drift time range at full width at half-maximum (FWHM) of the distribution [428]–[431]. Peaks in the experimental ATD are fitted by Gaussian curves and at FWHM, drift time ranges are extracted and converted into CCS values, resulting in CCS distributions (CCSDs). The presentation of CCSDs allows the visual depiction of possible molecular shapes that better reflect the ion conformational ensemble of macromolecules.

As previously mentioned in Section 2.2.5.3, comparisons of the CCS generated *in silico* from model structures to the experimental CCS values calculated are desirable to determine accuracy of the IM experiments or to propose a new model. Recently published CCS values of complexes and their assemblies in the gas phase were found to exhibit a broad correlation to known structural data obtained from EM, NMR spectroscopy and/or X-ray crystallography [231], [232]. The estimation of PA is a fast and efficient *in silico* measurement of CCS and is described in detail in Section 2.2.5.3. Although computational calculations are inexpensive and are able to rapidly produce supporting coarse-grained complex representations, scaling of the estimate is often required due to the underestimation of the algorithm applied. Recent studies comparing computational results to experimental results support the accuracy of scaled-PA estimates for the IM analysis of proteins and complexes [432], [433].

### 6.1.1: Aim

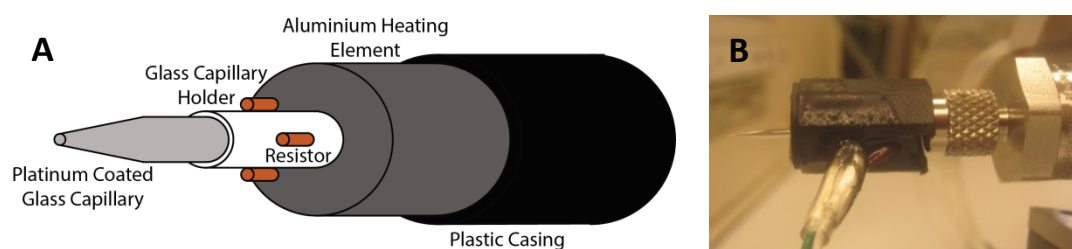
The aim of the work in this chapter was to determine higher order structural information for the individual *h*PFD subunits 5 and 6 and the intact *b*PFD complex using IM-MS. A combination of thermal degradation and solution disruption experiments were analysed by IM-MS to determine the stoichiometry, topology and dynamics of the chaperone. As there is little structural information available for intact eukaryotic PFD, data obtained in this chapter were compared to existing yeast models of PFD. The data determined here provides significant insight into the overall quaternary structural features of the eukaryotic PFD complex.

## 6.2: Results and Discussion

IM-MS analysis of *b*PFD and the *h*PFD subunits were performed on a Waters HDMS Synapt integrated with a TWIM cell, infused with inert helium gas. The experiments conducted here to elucidate the assembly of *b*PFD consist of a combination of 'native' MS, CID and thermal denaturation. Although solution disruption by small addition of solvents, chaotropic agents and the adjustment of pH levels were explored, optimisation of the experiments proved to be difficult as purified *b*PFD samples were limited, and therefore trial-and-error experimental approaches were restricted. Solution phase dissociation assays by thermal aggregation were the primary means of non-covalent disruption explored in this Chapter.

### 6.2.1: Development of CTC

To enable thermal denaturation experiments, a capillary temperature controller (CTC) was manufactured as part of this study to introduce heat to a sample directly before IM-MS analysis to allow the observation of thermal disruption and denaturation in real time. Although it was primarily designed for nanospray on the Waters HDMS Synapt, further developments to the design enabled the instrument to be easily transportable and transferrable and was successfully used on other MS instruments (Micromass QToF2). The CTC was constructed using a glass sheath able to hold the platinum coated glass capillary, with negative temperature coefficient (NTC) thermistors adhered to the glass tube covered by an aluminium heating element and sealed in resin then encased with plastic as seen in the schematic diagram in Figure 6.1A.



**Figure 6.1: The capillary temperature controller. (A)** A schematic diagram of the CTC. The internal capillary holder is made of glass with thermal resistors attached. The holder & the resistors are coated in an aluminium conductive heating element & the outer layer is made of plastic & all the elements are sealed in resin. **(B)** A photograph of the CTC in use on the Synapt HDMS Mass Spectrometer. It is easily assembled by slipping it onto the platinum coated glass capillary used for nanospray.

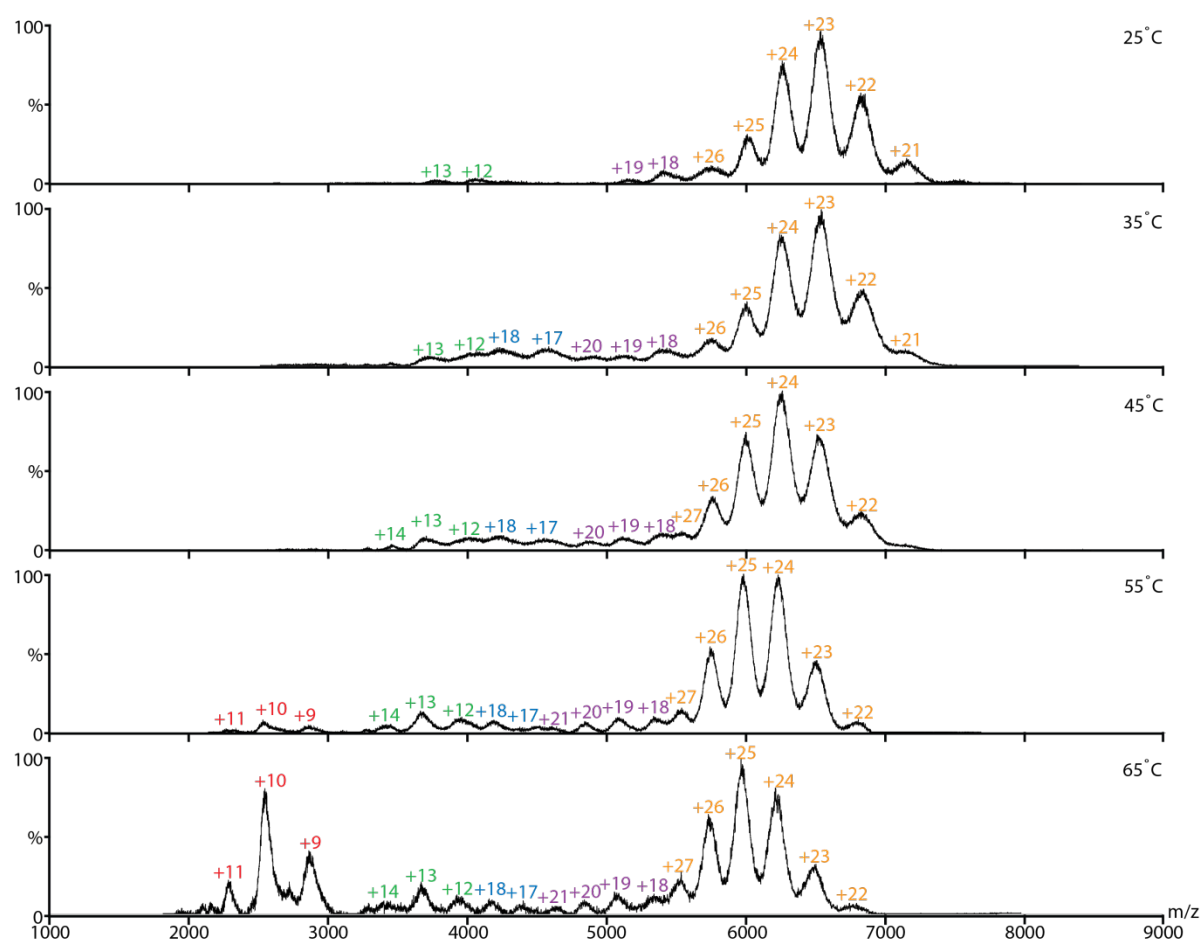
The aluminium element transfers heat to the sample solution through the glass capillary and glass capillary holder. The temperature is regulated and controlled by the NTC resistors. NTC resistors sense the output temperature and sends signals to the control box where the analogue data is converted to digital using a Pulse Width Modulation (PWM). This digital control simulates the voltage and changes the duty cycle; the ratio between the on and off time period. The pulse width is changed or modulated to achieve varying analogue values and this square wave signal switches the heating of the aluminium on and off. In this instance the PWM pattern is programmed fast enough that the set temperature remains constant throughout the experiment. The analogue output is converted to digital output using software available on Arduino. To adjust for the difference in read back from the NTC and the signal output, a control system (proportion integral derivative) was applied. Due to the length of the CTC, the needle must be extended longer than usual to allow the instrumentation to be slipped onto the nanospray capillary. As a precaution to ensure voltage across the tip of the glass capillary, it was coated in additional layers of platinum.

#### 6.2.1.1: Thermal Denaturation Experiment of IgG

To experimentally demonstrate the functioning mechanics of the CTC, the apparatus was attached to the nanospray source of the Waters Synapt and thermal dissociation assays were conducted on immunoglobulin G (IgG); a well characterised protein. IgG consists of 4 heterogenous chains, 2 light and 2 heavy, which have a mass of approximately 25 kDa and 50 kDa, respectively. The globular conformation of IgG resembles a Y shape, with the 2 heavy chains constructing the bulk outline of the Y and at the fork of the Y, a light chain is attached to a heavy chain. The forks have a predominant secondary structure of  $\beta$ -sheet folds and these are stabilised through inter-chain disulphide bridges [434], [435]. To investigate the thermal solution dissociation of IgG, the sample was buffer exchanged into a solution of AA at pH 7.4 and IM-MS parameters were set to 'native' conditions and were constant throughout the experiment. The temperature of the CTC was allowed to equilibrate for 3 minutes prior to IM-MS acquisition and data was obtained for temperatures of 25, 35, 45, 55 and 65 °C. The extracted MS spectra are illustrated in Figure 6.2.

It can be seen in Figure 6.2 that at room temperature, a Gaussian charge state distribution can be observed, highlighted in yellow correlating to the mass of intact IgG (148414 Da), with the most abundant peak centred around charge state  $[M+23H]^{23+}$ . Although mild samples conditions were used, trace amounts of in solution dissociation of the molecule is observed at  $[M+18H]^{18+}$ ,  $[M+19H]^{19+}$  (highlighted purple) and  $[M+12H]^{12+}$ ,  $[M+13H]^{13+}$  (highlighted green) for dimers consisting of 2 heavy chains and 2 light chains, respectively. Figure 6.2 reveals that an increase in

temperature to 35 °C induces further complex dissociation, with ions at  $[M+17H]^{17+}$  and  $[M+18H]^{18+}$  (highlighted blue) corresponding to the mass of a heavy chain and a light chain. At 45 °C, the energy applied has induced partial unfolding of the intact complex and a shift of the charge state distribution to higher  $m/z$ . The centred charge state is now seen at  $[M+24H]^{24+}$  and progressively moves to  $[M+25H]^{25+}$  as the temperature increases to 65 °C. The spectrum at 55 °C shows thermal dissociation of the monomeric light chain at  $[M+9H]^{9+}$ ,  $[M+10H]^{10+}$  and  $[M+11H]^{11+}$  (highlighted red) and a further temperature increase to 65 °C reveals a dramatic increase of ion abundance for the light chain monomer.



**Figure 6.2:** MS spectra of IgG at different temperatures (25–65 °C, 10  $\mu$ M IgG in 100 mM AA) utilising the CTC coupled to the Synapt HDMS Mass Spectrometer. The highlighted charge states correspond to (yellow) intact IgG, (purple) dimer of 2 heavy chains, (blue) dimer of 1 heavy & 1 light chain, (green) dimer of 2 light chains & (red) monomer of 1 light chain.

The  $m/z$  values of Figure 6.2 were extracted, deconvoluted and correlated to the corresponding IgG species. It can be deduced from the MS data and the identified correlating sub-complexes that the thermal dissociation of intact IgG energetically favours the dissociation of the light chains, however,

the corresponding dissociated heavy chains are not observed in the spectrum. The mass of a single heavy chain is equal to the mass of 2 light chains and it is highly likely that the ions corresponding to these complexes overlap. The charge states of  $[M+12H]^{12+}$ - $[M+14H]^{14+}$  (highlighted in green) are seen to increase in abundance following temperature increases. It is possible that the dimers of 2 light chains seen here dissociate to monomers and the corresponding heavy chain monomers replace the ions seen at  $[M+12H]^{12+}$ - $[M+14H]^{14+}$ .

Although previously published structural studies of IgG show the light chains are not bound in the IgG complex [434], [435], the results here show ions correlating to a dimer species of 2 light chains. This ultimately reveals that solution phase dissociated light chains can interact with other light chains and this has also been previously demonstrated to occur [436]. It is interesting to see that ions relating to monomeric light chains are not observed in the spectrum until temperatures of 55 °C and higher.

Previously published thermal dissociation studies of IgG have determined the complex to be of relatively high thermal stability, which ultimately explains the observation of the intact complex throughout the thermal dissociation assay conducted here. Complete dissociation and aggregation of IgG has been recorded at temperatures above 60 °C [437], [438].

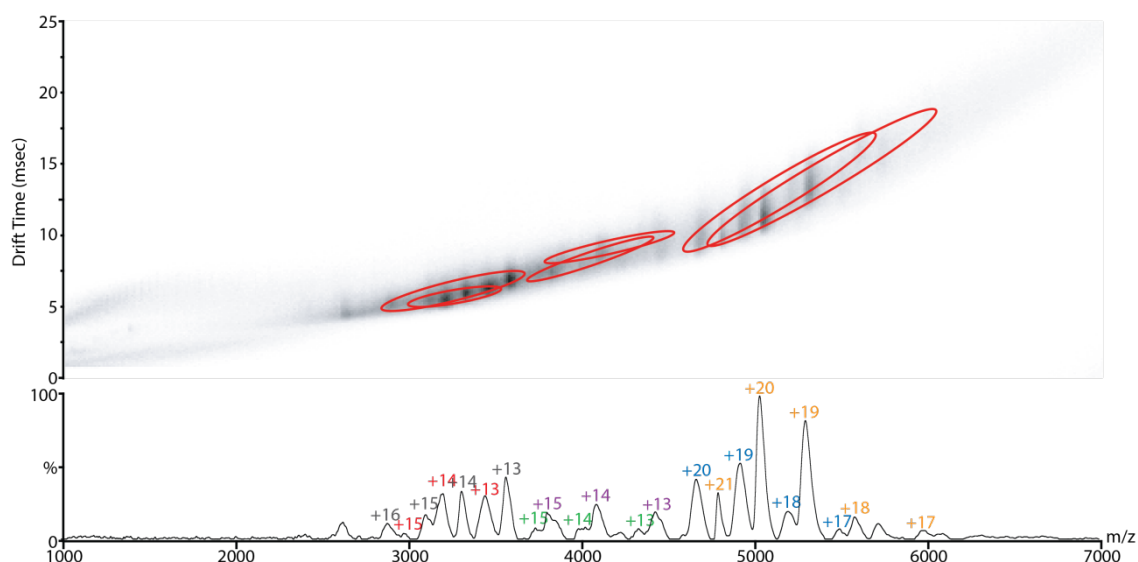
The experiments here were able to give an insight into the IgG complex topology and in solution interactions. It can therefore be concluded that the CTC is an effective and efficient instrumentation for the direct analysis of thermally dissociating complexes, with the capability of determining unfolding pathways and complex topology.

## 6.2.2: IM-MS Analysis of *b*PFD

### 6.2.2.1: 'Native' IM-MS Analysis of *b*PFD

Purified *b*PFD was buffer exchanged into an aqueous AA solution and IM-MS spectra were recorded. The IM-MS spectrum of *b*PFD (10 µM, in 100 mM AA, pH 7.4) is shown in Figure 6.3. Charge state distributions are indicated in the spectrum, with the most abundant peak seen at charge state  $[M+20H]^{20+}$ . Deconvolution of the MS spectrum reveals that this peak corresponds to intact *b*PFD (Figure 6.3, charge states highlighted in yellow) with a mass of approximately 100 kDa. It is also observed that sub-complexes are present in the sample. Although mild sample conditions were used to maintain native protein structures (100 mM AA, pH 7.5, room temperature), dissociation of the

complex is observed. It is therefore assumed that the *b*PFD complex is able to form sub-complexes at equilibrium under native like solution conditions without external denaturation.



**Figure 6.3:** IM-MS spectrum of *b*PFD (10  $\mu$ M *b*PFD, 100 mM AA, pH 7.4, Waters HDMS Synapt, TWIM, N<sub>2</sub>). The lower panel shows the conventional mass spectrum, whilst the top panel shows the IM drift plot indicating the drift times of individual ions. The highlighted charge states correspond to (yellow) intact *b*PFD, (blue) endoplamin, (purple) trimer consisting of *b*PFD subunits 2, 3 & 5, (green) trimer of subunits 2, 3 & 4, (grey) trimer of subunits 1, 5 & 6 & (red) trimer of subunits 1, 4 & 6.

Figure 6.3 reveals 6 major complexes identified by IM-MS in the purified *b*PFD sample. As previously mentioned, the charge state distribution of  $[M+20H]^{20+}$  labelled in yellow has an extracted mass of 100431 Da, corresponding to the calculated molecular mass of intact *b*PFD. It is seen here in the MS spectrum that several of the peaks associated with the intact chaperone have the highest relative abundance, therefore indicating the intact complex is the predominant macromolecular assembly within this sample.

Figure 6.4 displays the extracted ATDs for the charge states of intact *b*PFD measured at a wave height of 9 V. The drift times extrapolated from the ATDs correspond to the peak of the Gaussian like distribution. The drift plots of the extracted intact *b*PFD ions show peaks with increasing ATD widths following the increase in drift time. This is a characteristic associated with ion diffusion. To minimise errors in the conversion process of drift time to CCS, the CCS values were averaged over the recorded IM-MS spectra, which were each performed such that the IM wave heights for IM were varied between 9-11 V whilst maintaining the wave velocity at 350 ms<sup>-1</sup>. Although it has been previously stated that the conversion of IM data to CSS values through the extraction of the drift

time at the peak of the ATD neglects possible conformational ensembles, the FWHM method is applied at a later stage.

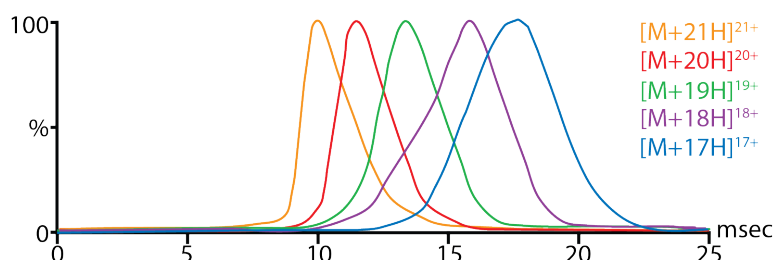


Figure 6.4: Extracted ATD of the ions corresponding to *b*PFD charge states  $[M+21H]^{21+}$ ,  $[M+20H]^{20+}$ ,  $[M+19H]^{19+}$ ,  $[M+18H]^{18+}$ , &  $[M+17H]^{17+}$  at a wave height of 9 V.

Calibrant proteins of known CCS were analysed by IM-MS utilising the same parameters for the native *b*PFD experiments (varying wave heights of 9-11 V and wave velocity of  $350 \text{ ms}^{-1}$ ) in order to construct a calibration plot for CCS measurement [426]. The extracted drift times of *b*PFD were then converted to CCS values using the calibration curve, as summarised in Table 6.1.

Of the 6 protein complexes identified in the spectrum (Figure 6.3), 4 of the complexes, with highlighted charge states in grey, red, purple and green are found to correlate to sub-complex trimers of *b*PFD. Deconvolution of these charge state distributions reveal masses of 46134 Da, 44079 Da, 54297 Da and 56352 Da, respectively with their respective components summarised in Table 6.1. Unfortunately, a high abundance of the molecular HSP endoplasmin (refer to Appendix C) is also observed in the spectrum. The charge state distribution highlighted in blue ( $5434 \text{ m/z } [M+17H]^{17+}$ ,  $5193 \text{ m/z } [M+18H]^{18+}$ ,  $4919 \text{ m/z } [M+19H]^{19+}$ ,  $4673 \text{ m/z } [M+20H]^{20+}$ ) has an extracted mass that corresponds to endoplasmin (Table 6.1; 92369 Da). This reinforces the proteomic analysis in Section 3.3.2 and further verifies contamination of the isolated *b*PFD sample.

In a similar manner to the ions corresponding to intact *b*PFD, the drift times extracted from the peak centre of the ATDs for the sub-complexes were extracted from the spectra and the drift times were converted to CCS using a standard calibration protocol (refer to Section 6.3.3) and are displayed in Table 6.1. Errors associated with the CCS determined from the calibrants, the calibration curve and the relative precision of replicate measurements is estimated to be approximately 8-10 % [358].



**Table 6.1: The  $m/z$  and CCS values derived from the spectrum of intact *b*PFD complex & the dissociating sub-complexes observed in an aqueous solution of 50 mM AA, analysed by TWIM Waters synapt ( $N_2$ ).**

Calculated Mass Charge State	$m/z$	CCS ( $\text{\AA}^2$ )
100431 Da ( <i>b</i> PFD)		Ave: 6146
+17	5909	5968
+18	5581	6127
+19	5287	6128
+20	5023	6144
+21	4783	6147
+22	4566	6362
46134 Da ( <i>b</i> PFD sub 1,5,6)		Ave: 3499
+13	3550	3312
+14	3296	3413
+15	3077	3556
+16	2884	3711
44079 Da ( <i>b</i> PFD sub 1,4,6)		Ave: 3337
+13	3392	3203
+14	3150	3308
+15	2940	3500
54297 Da ( <i>b</i> PFD sub 2,3,4)		Ave: 3807
+13	4178	3660
+14	3879	3816
+15	3621	3943
56352 Da ( <i>b</i> PFD sub 2,3,5)		Ave: 3840
+13	4336	3695
+14	4026	3857
+15	3758	3967
93466 Da (endoplasmin)		Ave: 5680
+17	5498	5556
+18	5193	5659
+19	4919	5715
+20	4673	5796

Averages are calculated over the charge state distribution observed in the spectrum.

IM-MS conditions used in this analysis were controlled to retain native protein structures. It is therefore expected that the 4 sub-complex populations reflect a solution phase dissociation of *b*PFD. It is proposed that the intact hexamer undergoes some minor dissociation in solution before IM-MS analysis. These dissociations are not due to gaseous interactions within the Synapt, as typical gas phase dissociations eject a single monomer of the stripped oligomer, inducing simultaneous dissociation and unfolding of the macromolecule, which is not observed here [223], [238], [422], [423], [439]–[441].

Previous structural studies of eukaryotic PFD propose that the  $\alpha$  like subunits 3 and 5 construct the core component of PFD and are therefore within the centre of the jelly fish shaped complex [89], [91], [93], [95], [98]. The suggested subunit arrangement positions the  $\alpha$  subunits 3 and 5 opposite each other surrounded by  $\beta$  subunits (Figure 6.5A). The ‘native’ mass spectrometric analysis of *b*PFD revealed dissociated trimers of the complex, suggesting possible internal interactions and subunit topology. It is presumed that the subunits within the respective dissociated trimer must have strong

non-covalent interactions with the ability to retain this assembly throughout the analysis. Therefore, observation of the MS data of the dissociated trimers 1+5+6, 2+3+4, 1+4+6, 2+3+5 along with the previously proposed subunit assembly within the PFD complex reveals possible underlying substructures. Figure 6.5 illustrates the *b*PFD subunit arrangement previously proposed, and the possible subunit arrangements according to the MS analysis.

Observation of trimers consisting of subunit 1+5+6 and 2+3+4 from previously published subunit assemblies of eukaryotic PFD (Figure 6.5A), indicates several potential scenarios. As these subunits are opposite to each other, it is highly likely that dissociation has occurred through the middle of the intact *b*PFD complex, ultimately suggesting that the bonds holding these trimers in the intact *b*PFD complex are not as strong as the bonds holding the subunits within their respective dissociated trimer. Figure 6.5B illustrates this scenario with the breaking of bonds between subunit 2 and subunit 1, and subunit 4 to subunit 6, as indicated by the red line. Alternatively, the dissociation of these trimers may suggest undiscovered binding interactions between the respective trimers as illustrated in 6.5B by the purple lines for trimer 1+5+6 and by the green lines for trimer 2+3+4. However, there are some published papers stipulating a strong interaction occurring between the archaeal PFD  $\alpha$  subunits 3 and 5, which conflict with the dissociation of these trimers [89], [93], [98]. Nevertheless, it has been previously indicated in this thesis that prokaryotic PFD studies do not necessarily translate to eukaryotic PFD and this particular interaction may not be present in *b*PFD.

The dissociation of the trimers 1+4+6 and 2+3+5 observed in the mass spectrum are more difficult to explain. It is proposed here that the sub-complexes 1+4+6 and 2+3+5 may have an underlying interaction with the respective subunits in their trimer, forming 2 substructures within the jelly fish like shape. A closer look at subunits 3 and 5 on the previously deduced EM model of eukaryotic PFD (Figure 1.6B), illustrates that these subunits have more of an internal arrangement in comparison to subunits 1, 4 and 6. However, the results here indicate that the trimer 2+3+5 has an overall larger CCS value. It is possible that trimer 1+4+6 constructs the inner core or scaffold of the eukaryotic PFD while trimer 2+3+5 is overlaid, therefore having a more external arrangement in the PFD complex assembly, resulting in trimer 2+3+5 being more exposed, allowing the dissociation of the trimers. The proposed interactions are illustrated in Figure 6.5C.

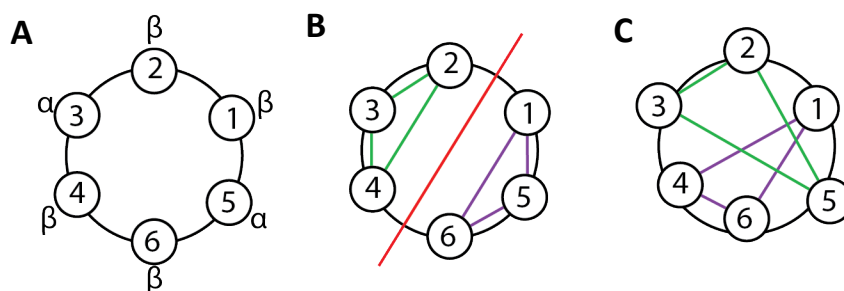


Figure 6.5: (A) Schematic diagram of the previously proposed *bPFD* subunit arrangement. (B) The complex is able to dissociate as indicated by the red line to give trimers consisting of subunits 1+5+6 & 2+3+4. (C) This subunit arrangement may have internal underlying interactions as indicated by the purple & green lines that allow the dissociation of trimers 2+3+5 & 1+4+6.

Typical IM-MS experiments utilise the calculated CCS restraints to guide the development of a model of the intact complex, generally integrating previous structural information derived from NMR, crystallography or molecular dynamic simulations for comparison. As is expected, the calculated results here reveal the dissociated trimers have CCS values smaller than the intact complex. Figure 6.6 illustrates a schematic diagram of the CCS value of intact *bPFD* and the sub-complexes.

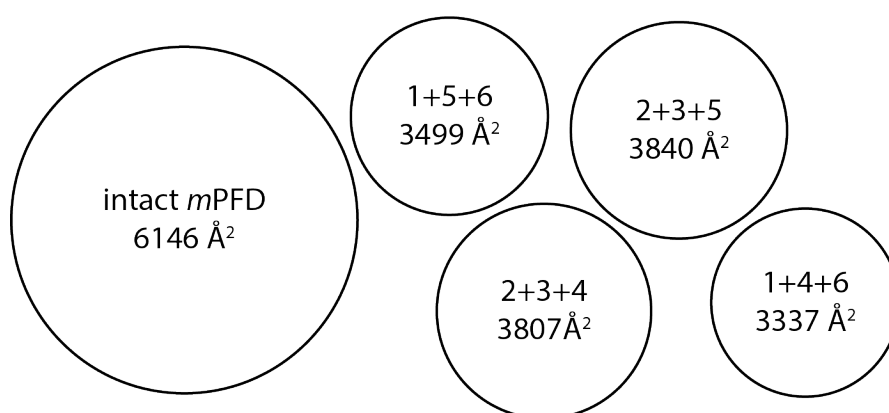


Figure 6.6: Representative diagram of the calculated CCS of intact *bPFD* & the trimers of subunits 1+5+6, 2+3+4, 2+3+5 & 1+4+6.

A comparison between the measured CCS and the computationally derived CCS of the complex model structure is desirable. As previously mentioned, very little structural information of *bPFD* is available, ultimately obstructing the comparison of the derived CCS *bPFD* to eukaryotic PFD models. Fortunately there are bacterial PFD models with well-defined structures available in the PDB acquired through x-ray crystallography. The BLAST global alignment of *phPFD* and *mtPFD* to *bPFD* previously compared in Section 3.2.1 shows that there is little global sequence similarity, however,

domains of PFD were previously determined to be conserved across homologues, suggesting that function and structure across evolutionary lineages may have been conserved. Therefore, the computationally derived CCS models of *ph*PFD and *mt*PFD can be used as a rough estimation for comparison to experimental *b*PFD CCS.

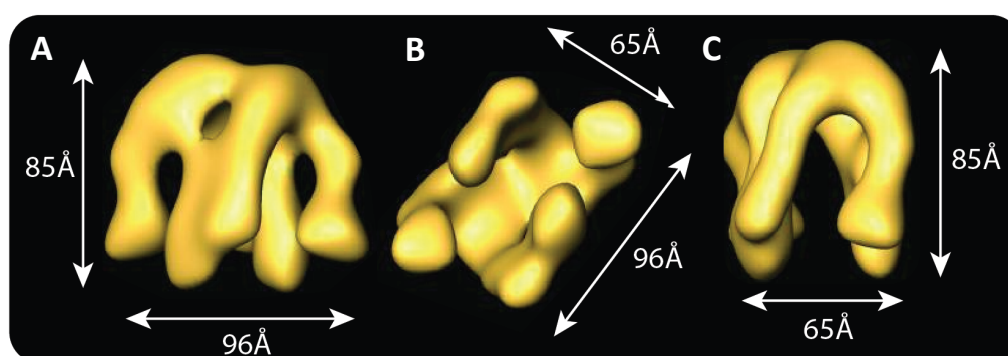
To compare the CCS of *b*PFD to *ph*PFD and *mt*PFD, the Leeds method was employed [229], [230], which utilises the projection algorithm (PA). This is a good initial estimation and is explained in detail elsewhere [442]. The PA methodology neglects multiple collisions between the proteins, ions and buffer gas and is therefore generally an underestimation of the experimental CCS of proteins [233], [372]. However, it has been shown that it is correlated with the experimental CCS for protein complexes ( $R^2 > 0.99$ ) [432]. The scaled PA is used here by multiplying the PA CCS by a factor of 1.14 as previously demonstrated [432], [433]. The CCS of the PFD species can therefore be compared here and are listed in Table 6.2.

**Table 6.2: Computational & experimental CCS values of different species of PFD**

Species of PFD	Mass (kDa)	PDB	CCS ( $\text{\AA}^2$ ) derived from IM-MS TWIM. $\text{N}_2$ .	CCS ( $\text{\AA}^2$ ) derived from Leeds method. Helium.	Scaled PA CCS ( $\text{\AA}^2$ )	Reference
<i>b</i> PFD	100.1	-	6146 $\pm$ 614	-	-	-
<i>ph</i> PFD	88.0	2ZDI	-	4002 $\pm$ 59	4562 $\pm$ 67	[84]
<i>mt</i> PFD	79.8	1FXK	-	3973 $\pm$ 60	4529 $\pm$ 68	[93]

The experimentally derived CCS of intact *b*PFD of 6146  $\text{\AA}^2$  can be compared to the theoretically calculated CCS values of *ph*PFD and *mt*PFD. In terms of calculated CCS, *ph*PFD and *mt*PFD are smaller than *b*PFD by approximately 26 %. These size comparisons are unsurprising, as the prokaryotic PFDs are significantly smaller than intact *b*PFD in terms of mass by 12.1 % and 20.3 %, *ph*PFD and *mt*PFD respectively. Furthermore, the overall primary structures of these species are shorter in length and have replicate subunits unlike the heterohexameric *b*PFD. However, the comparison here can give insight into the different chaperoning mechanisms of the PFD species and the substrate size capacity as published studies stipulate eukaryotic PFD is able to encapsulate the substrate for correct refolding [91], [95], [96]. The larger experimental CCS of the *b*PFD suggests a larger cavity inside the eukaryotic chaperonin. This helps reinforce the studies that suggest eukaryotic PFD aids in the refold of misfolding substrates through the simultaneous distal tip interaction and partial encapsulation, rather than only through the distal tips, which is suggested to occur for prokaryotic PFD.

The experimentally derived CCS of intact *b*PFD can also be compared to the size of the negatively stained TEM images. Figure 6.7 illustrates the reconstructed models of *h*PFD derived from TEM and the approximate lengths of the *h*PFD models. Rough CCS estimates can be determined from these models by calculating the lengths to determine the area of the models. The area calculated for the front of *h*PFD (Figure 6.7A) is approximately 8125 Å<sup>2</sup>, while the bottom of the molecule is estimated around 5672 Å<sup>2</sup> and the side of the complex is calculated to be approximately 5529 Å<sup>2</sup>. An average of these areas gives a rough CCS estimation of 6442 Å<sup>2</sup>, which is within the range of our experimentally derived CCS of intact *b*PFD of 5532-6760 Å<sup>2</sup>.



**Figure 6.7:** The 3D reconstruction of *h*PFD from electron micrographs including measured lengths. (A) Side view of *h*PFD, (B) bottom view of *h*PFD & (C) side view of *h*PFD. Adapted from [95].

The comparison of the experimentally derived CCS average of *b*PFD to the crystallography models of *ph*PFD and *mt*PFD and the TEM data of *h*PFD reveals a relatively similar CCS estimation. Subsequently, there is high confidence in the calculated CCS from the IM data.

#### 6.2.2.2: CID IM-MS Analysis of *b*PFD

In an effort to extract further complex assembly information from the *b*PFD sample, tandem MS analysis was applied. Tandem MS, following isolation of ions at *m/z* 5023 corresponding to *b*PFD ([M+20H]<sup>20+</sup>) at a voltage of 200 V is seen in Figure 6.8.

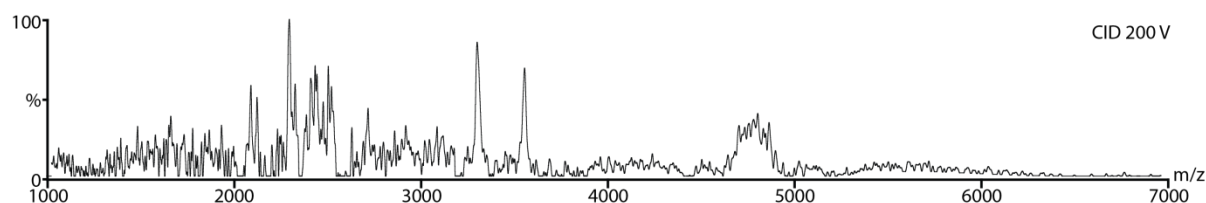


Figure 6.8: Tandem MS spectrum of selected ion 5023  $m/z$ . A CID voltage of 200 V was applied. No ions corresponding to stripped oligomers are observed in the spectrum. Tandem MS analysis was performed on at least 3 separate occasions, displaying the same peaks observed here.

Following CID, the intact *b*PFD complex was found to dissociate (Figure 6.8). Although many peaks can be seen between 1000-3000  $m/z$  and trace amounts of ions observed around 4800 and 5250  $m/z$ ; the low ion count, similar ion mobilities and overall convoluted spectrum did not allow the deconvolution of the spectrum at these  $m/z$  ratios. However, 2 defined peaks at 3296 and 3550  $m/z$  can be clearly seen in the spectrum. These peaks correspond to  $[M+14H]^{14+}$  and  $[M+13H]^{13+}$  of the trimer consisting of subunits 1+5+6, as previously calculated in Section 6.2.2.1. Although the amount of ions for the trimer peaks was low, the ATD was still extracted and the calculated CCS average of the CID induced trimer was 4474 Å<sup>2</sup>. This ultimately shows that an increase in CID voltages results in subsequent *b*PFD and subunit dissociation and unfolding, discussed in Section 6.2.2.1. As the trimer of *b*PFD subunit 1, 5 and 6 is observed in Figure 6.8, this reveals the strong interactions of these subunits to each other with unfolding of the trimer evident in the calculated CCS average of 4474 Å<sup>2</sup> as opposed to the 'native' IM-MS analysed trimer calculated to be 3499 Å<sup>2</sup>.

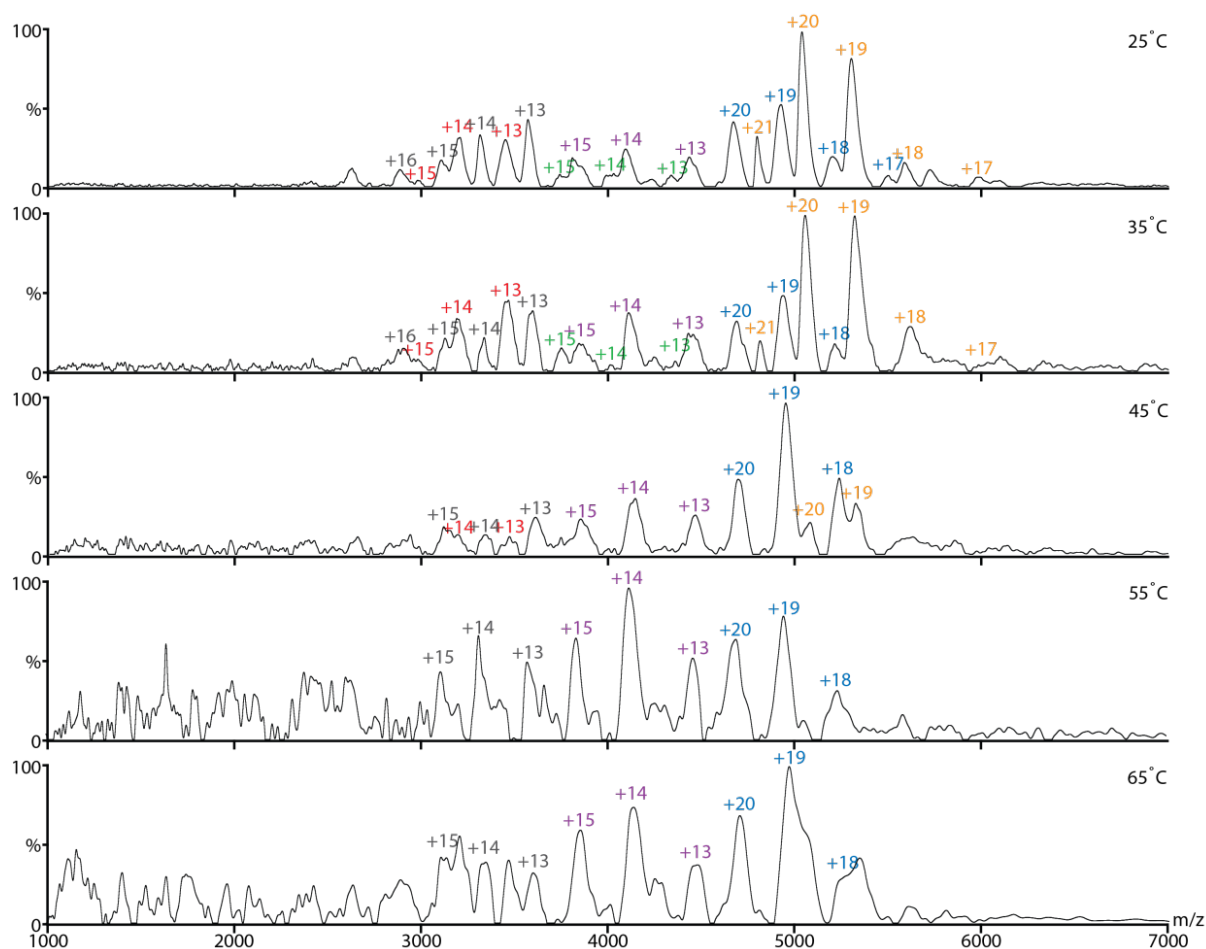
Although the CID data presented here reveals trimer dissociation of the intact *b*PFD, the results are not consistent with normal CID pathways (Section 6.2.2.1). Fortunately, the deviation from the expected results has been observed elsewhere and it is possible that CID parameters and the unique protein structure has attributed to these outcomes. A study of intact *mt*PFD conducted by Robinson *et al.* utilising CID was able to demonstrate the expected monomer and oligomer products [94]. *mt*PFD concentration was at 15 µM in an aqueous solution of 10 mM AA at pH 8, analysed by MS with CID voltage set to 50-150 V, pressures of  $8.7 \times 10^{-6}$  Pa (source) and  $7.1 \times 10^{-8}$  Pa (analyser) and a capillary voltage of 1650 V. Disruption of the *mt*PFD hexamer to the pentameric species and monomer  $\beta$  subunit was clearly observed in the spectrum [94]. However, a 1/25 dilution of the sample in an aqueous solution of 10 mM AA at pH 8, analysed by MS with CID voltage set to 150 V, pressures of  $1.9 \times 10^{-5}$  Pa (source) and  $7.1 \times 10^{-8}$  Pa (analyser) and a capillary voltage of 1750 V reveals a spectrum with a high abundance of trimers and tetramers. Relative high abundances of the *mt*PFD dimer were also formed as a dissociation product of intact *mt*PFD under conditions that

strongly destabilise the complex [94]. This particular study also investigated the aqueous thermal dissociation of *mtPFD*.

It can therefore be concluded that the parameters utilised for the tandem MS analysis of *bPFD* were under conditions that strongly destabilised the complex, resulting in possible monomers that could not be resolved (1000-3000 *m/z*) and a trimer consisting of subunits 1+5+6.

#### 6.2.2.3: Thermal Denaturation IM-MS Analysis of *bPFD*

Thermal denaturation was applied to the sample using the CTC developed. As seen in Figure 6.9, increasing the temperature of the *bPFD* complex causes a shift of the charge state distributions to the lower *m/z* region of the spectrum as well as an overall decrease in sensitivity as the complex is thermally dissociated. Figure 6.9 illustrates that an increase in temperature of the CTC from room temperature to 35 °C marginally decreases the intensities of the peaks corresponding to the contaminant endoplasmin. However, all other peaks in the spectrum appear to be the same. A further increase to 45 °C reveals the peaks corresponding to intact *bPFD* have drastically decreased in relative intensity, and endoplasmin is now the predominant species in the thermally treated sample. MS analysis at this temperature reveals peaks corresponding to trimer *bPFD* 2, 3 and 5 of 4336, 4026 and 3758 *m/z* are no longer observed in the spectrum. A temperature increase to 55 °C shows that the intact *bPFD* complex and the dissociated trimer 1, 4, 6 are no longer present in the spectrum, which suggests further dissociations or aggregation. Peaks corresponding to the *bPFD* trimers of 1, 5, 6 and 2, 3, 4 are still present in the spectrum and are also present in the MS analysis at 65 °C. Unfortunately the contaminant endoplasmin is present throughout the thermal IM-MS assay. This is unsurprising as it is a heat shock protein and is upregulated and stable in thermal stress conditions [66].



**Figure 6.9:** MS spectrum of *bPFD* at different temperatures (25–65 °C, 10  $\mu$ M *bPFD* in 100 mM AA) utilising the CTC coupled to the Synapt HDMS mass spectrometer. The highlighted charge states correspond to (yellow) intact *bPFD*, (blue) endoplamin, (purple) trimer consisting of *bPFD* subunits 2, 3 & 5, (green) trimer of subunits 2, 3 & 4, (grey) trimer of subunits 1, 5 & 6 & (red) trimer of subunits 1, 4 & 6.

It is evident in Figure 6.9 that although some small dissociations of the intact *bPFD* was observed in the ‘native’ IM-MS analysis at room temperature, increasing the temperature of the CTC resulted in further dissociation of the hexamer, primarily into its trimers of subunits 1+5+6 and subunits 2+3+4. This ultimately means that the binding interactions of the subunits in the respective trimers have strong non-covalent interfaces. The increase in noise following the incremental increases in CTC temperature suggests aggregation and protein precipitation. Although it is unclear whether smaller species are present in the spectra due to the loss in sensitivity with the increase in solution temperature, further dissociations of the *bPFD* trimers cannot be ruled out.

Following MS analysis of the thermal dissociation assays of *bPFD*, the FWHMs of the ATDs of the ions were extracted and the drift times were converted to CCS utilising the standard calibration curve [426]. Figure 6.10 reveals the reconstructed CCSDs of the ions corresponding to intact *bPFD* analysed



at temperatures 25-45 °C. The spot on the CCSD marks the CCS value derived from the drift time at the peak of the ATD.

It can be seen on Figure 6.10 that the charge states at 25 °C have the lowest CCS values in comparison to CCS of 35 and 45 °C. At charge  $[M+17H]^{17+}$ , the CCS reflects the smallest protein conformer. The ions analysed at 25 °C of higher charge states  $[M+18H]^{18+}$ - $[M+21H]^{21+}$  have similar CCS values at approximately  $6150 \text{ \AA}^2$ , revealing a similar conformation. At charge state  $[M+20H]^{22+}$  an increase in CCS to approximately  $6350 \text{ \AA}^2$  is observed. In comparison, the analysis of *b*PFD at 35 °C illustrates a similar trend to the ions observed at 25 °C, with a slight increase in CCS values at approximately  $50 \text{ \AA}^2$  and higher CCS values at charge states  $[M+21H]^{21+}$  and  $[M+22H]^{22+}$  by approximately  $150 \text{ \AA}^2$ , possibly revealing an unfolding intermediate at this temperature. Further increases in temperature to 45 °C show 2 peaks corresponding to intact eukaryotic PFD, with significant increases in CCS values to approximately  $6770 \text{ \AA}^2$ , suggesting a high proportion of protein unfolding. Unfortunately, at higher temperatures, no peaks were seen related to intact *b*PFD, revealing complete dissociation of the macromolecule or possible aggregation. Further analysis of the data and CCSDs of the *b*PFD ions show a variation in the width of the CCSDs, with charge states  $[M+21H]^{21+}$  and  $[M+22H]^{22+}$  illustrating notably longer distributions than the more refined charge states  $[M+17H]^{17+}$ - $[M+20H]^{20+}$ . This reveals an increase in conformational ensembles for the ions corresponding to the higher charge states  $[M+21H]^{21+}$  and  $[M+22H]^{22+}$ , consistent with coulombic unfolding.

It should be noted that the variation in CCS values between charge states is a typical occurrence. Higher charged ions generally have a more unfolded structure, as there is an increase in surface area enabling the attachment of more protons during ESI [412]. The similar calculated CCS values and CCSDs of charge states  $[M+18H]^{18+}$ - $[M+20H]^{20+}$  of *b*PFD reveal a stable conformation. The thermal denaturation of *b*PFD by CTC and analysed by IM-MS was repeated to demonstrate the reproducibility of the results obtained. The repeat was found to have similar spectra to those seen in Figure 6.9 and the CCSDs calculated differed at most by  $\pm 10 \%$ .

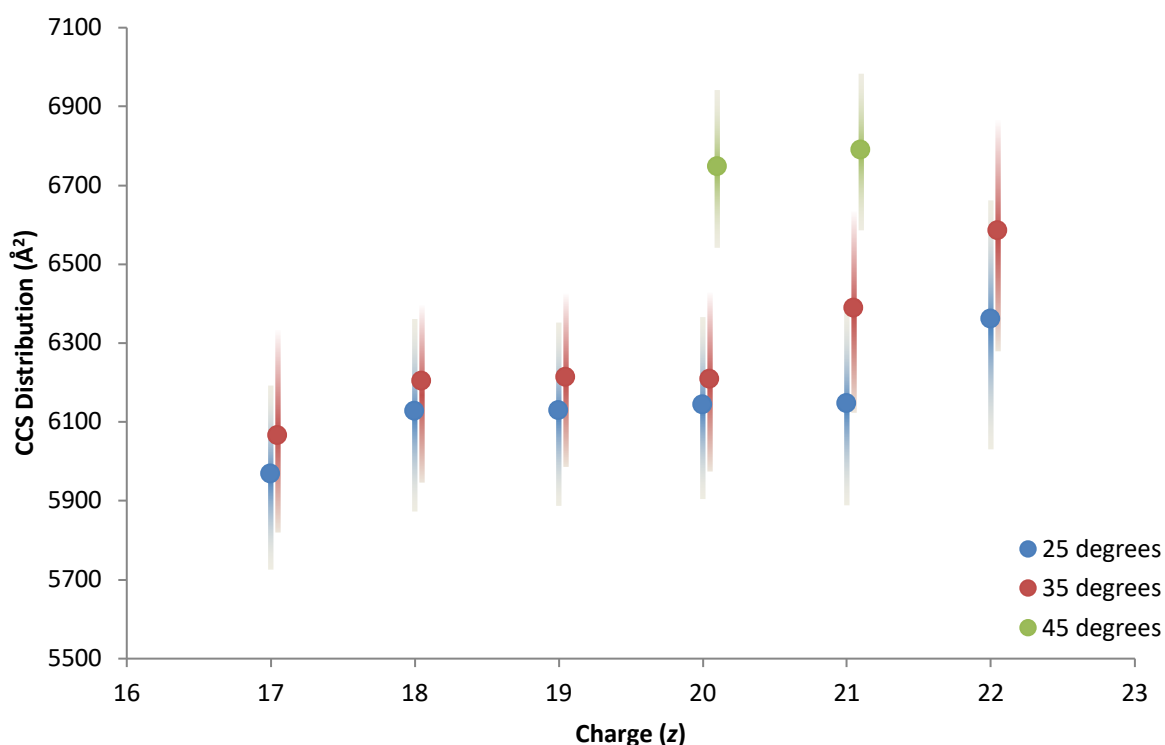


Figure 6.10: The CCSDs derived from the IM-MS thermal dissociation experiments of intact *bPFD* complex at temperatures 25 °C (blue), 35 °C (red) & 45 °C (green).

Following extraction of the CCS values and CCSDs of intact *bPFD*, the same method was applied to the dissociated trimers, and the CCSDs of the charge states  $[M+13H]^{13+}$ - $[M+15H]^{15+}$  are plotted as a function of temperature in Figure 6.11. Figure 6.11A illustrates that the ions corresponding to charge state  $[M+13]^{13+}$  for 4 trimers are seen trend in a similar manner. A steady increase in calculated CCS values is seen to occur from 25-45 °C, suggesting partial unfolding of the trimers. At temperatures of 45-55 °C, the CCS values seem to remain relatively similar, indicating a stable conformation. Further thermal aggregation to 65 °C induces further protein unfolding as the calculated CCS increases. It is noted that ions corresponding to trimers 2+3+4 (highlighted in green) and 1+4+6 (highlighted in red) in Figure 6.11 are seen to increase in size following temperature changes and have no peaks observed at higher temperatures. This ultimately reveals that these trimers are sensitive to thermal stimuli, possibly resulting in further dissociations or aggregation at these high temperatures.

In comparison, ions relating to the trimers 1+5+6 (highlighted in blue) and 2+3+5 (highlighted in purple) are observed throughout the thermal dissociation assay, indicating high resistance to thermal stimuli. Figure 6.11B illustrates that the ions of charge state  $[M+14]^{14+}$  have very different trends with the increase of temperature. Ions corresponding to trimer 1+5+6 at this charge state are

seen to retain a relatively similar CCS values throughout the assay, whilst trimer 2+3+5 is seen to increase in size throughout the analysis of 25-35 °C, and remaining stable at temperatures 35-45 °C. Further increases in CCS value is observed for the temperatures of 45-55 °C and the last acquisition of 65 °C is seen to retain this calculated CCS. Interestingly, the trends observed in Figure 6.11A for all the trimers are also present in Figure 6.11C, albeit at slightly higher CCS peak values.

The violin plots of Figure 6.11, enable a visual comparison of the peak CCS values and CCSDs of the ions correlating to the *b*PFD sub-complexes to simultaneously reveal the conformational ensembles and unfolding intermediates of the trimers. The CCS values plotted as a function of temperature aids in the determination of the unfolding structure following the thermal stimulus and all ions correlating to the *b*PFD complex and the dissociating sub-complexes are seen to exhibit increasing drift times over the course of the thermal aggregation assay, which equates to increasing CCS values synonymous with the unfolding of protein structure [443]–[445]. The CCSDs observed in these plots reveal that the distribution increases for the ions analysed at high temperatures of 55 °C and above. Large CCSD widths are associated with a high number of different conformational ensembles and intrinsic polymorphism, which is likely to have occurred here for the ions analysed at 55 °C and 65 °C. Travelling wave IM instruments have been reported to have a resolving power of 40 therefore corresponding to a resolution of 0.025 or 2.5 % [446], [447]. The peak widths observed in Figure 6.11 for the sub-complexes are seen to range from 0.070 (7.0 %) to 0.111 (11.1 %), which are considerably greater than that expected CCSDs based on the resolution of the instrument. This ultimately shows that the large widths of the distributions are attributed to the structure of the complexes and therefore supports the idea of conformational flexibility.

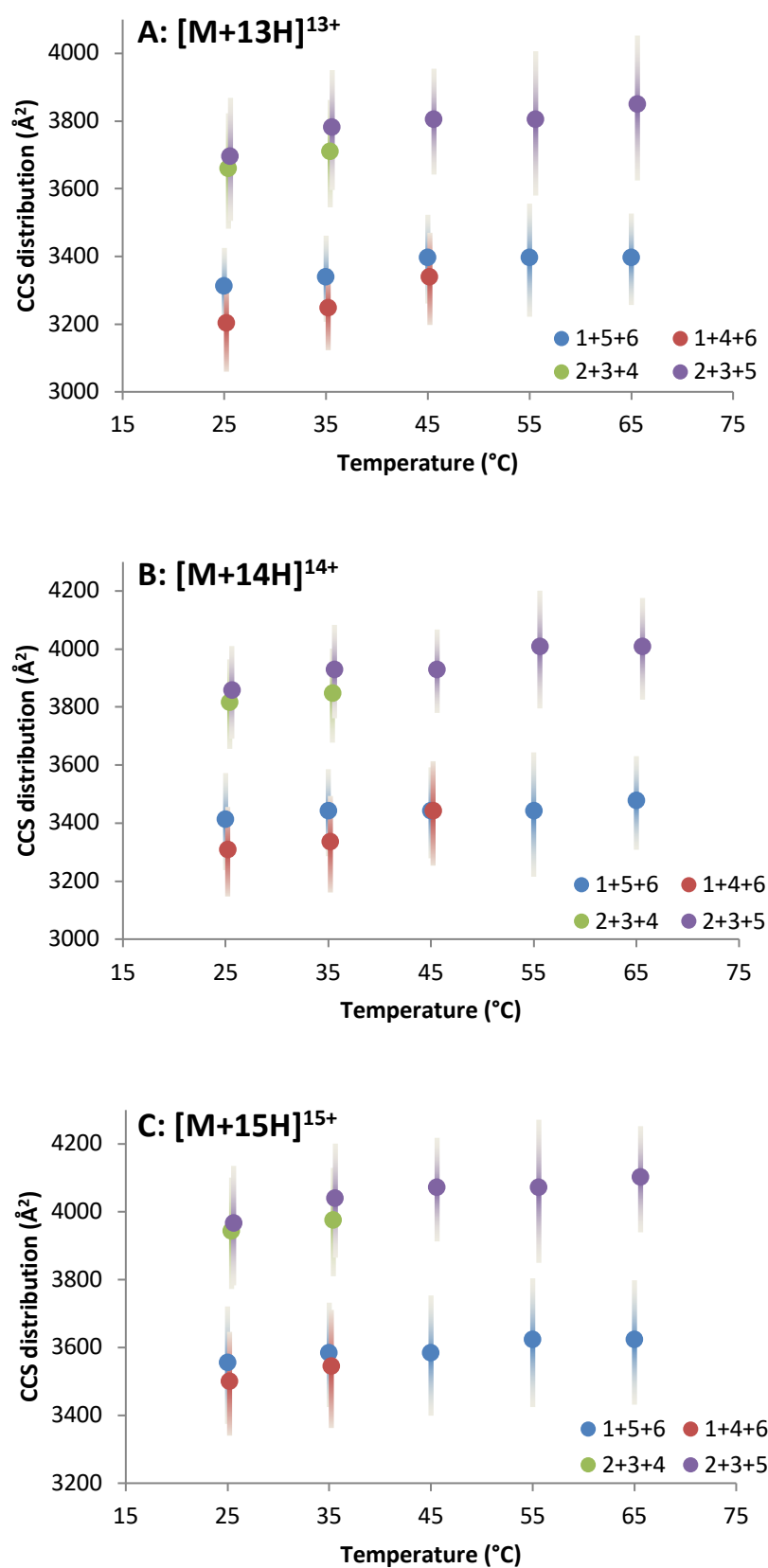


Figure 6.11: The CCSD derived from the IM-MS thermal dissociation experiments of the *b*PFD sub-complexes; trimer of subunits 1, 5 & 6 (blue), trimer of subunits 1, 4 & 6 (red), trimer of subunits 2, 3 & 4 (green) and trimer of subunits 2, 3 & 5 (purple). The product ions (A)  $[M+13H]^{13+}$ , (B)  $[M+14H]^{14+}$  & (C)  $[M+15H]^{15+}$  are plotted as a function of temperature.

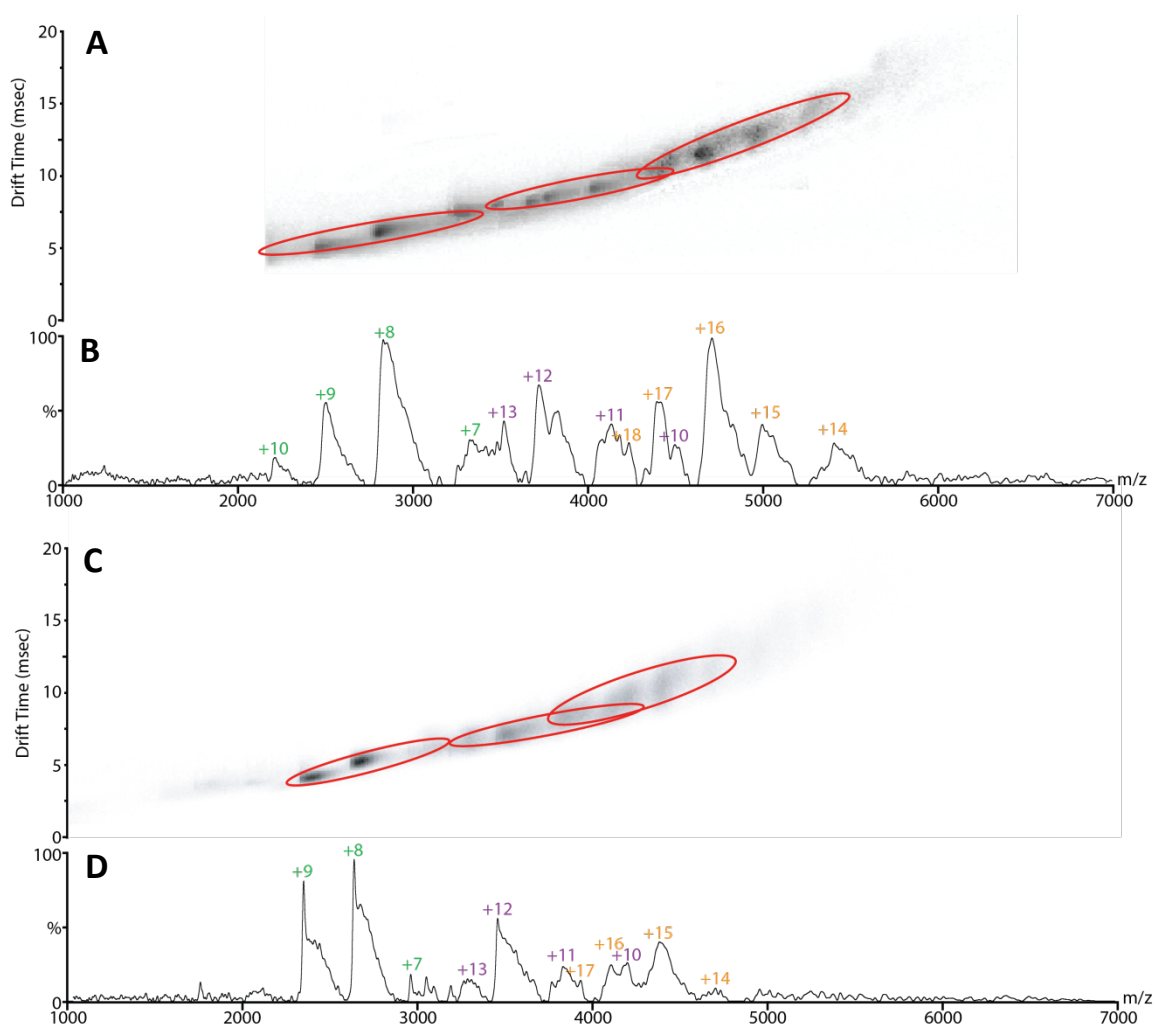
As previously mentioned in Section 6.2.2.2, the thermal solution phase and gas phase dissociation of prokaryotic PFD has previously been thoroughly investigated [94]. The MS spectrum of *mtPFD* under mild 'native' conditions reveals small dissociations of the complex, similar to the data obtained here seen in Figure 6.2. The intact *mtPFD* complex was found to be a relatively stable macromolecule despite high gas-phase activation and thermal aggregation. Ions correlating to the intact prokaryotic molecule are seen in the spectra following MS analysis at temperatures of 70 °C and CID activation of 150 V [94]. This was not seen for the analysis of *bPFD*, which was observed to completely dissociate at temperatures above 45 °C and with CID at voltages of 200 V. The difference in complex dissociation between the species of PFD is unsurprising as *mtPFD* is a hyperthermophile that grows optimally at 65 °C [94]. Given the structural integrity of *mtPFD* in conditions of elevated kinetics, complete dissociation was observed at temperatures above 85 °C and a thorough examination of the MS spectra revealed products of dimers and trimers, however the predominant product seen was the monomeric  $\beta$  subunit of *mtPFD*. As monomer subunit dissociation was not observed for *bPFD*, it is suggested that there is a high degree of structural cooperativity within the eukaryotic PFD, particularly in the sub-complex trimers seen in the spectra. It is therefore proposed that the folding and biological activity of *bPFD* differ from *mtPFD*.

To summarise, the IM-MS analysis of the *bPFD* sample in 'native' conditions revealed small amounts of complex dissociation into its trimer sub-complexes. Following thermal aggregation, dissociation of the complex was seen to increase and resolution is lost. Complete dissociation of the intact *bPFD* complex is observed at a temperature of 55 °C. At this temperature trimer 2+3+5 is also seen to disappear from the analysis followed by trimer 1+4+6 at 65 °C. Trimers 1+5+6 and 2+3+4 are present to the end of the IM-MS analysis. The CCS values calculated are observed to increase with the increase in thermal activation, a common characteristic of unfolding macromolecules. Unfortunately, further assays were unable to be conducted past 65 °C as difficulty was present in sample ionisation and obtaining an even spray at higher temperatures. The data seen here for solution-phase thermal activation suggests that between 25-35 °C there seems to be an equilibrium in solution between the intact *bPFD* and the dissociated trimers. Increasing the kinetic energy available causes this equilibrium to shift towards unfolding of the complex and further dissociations, particularly to trimers 1+5+6 and trimer 2+3+4.

## 6.2.3: IM-MS Analysis of *h*PFD Subunits 5 and 6

### 6.2.3.1: 'Native' IM-MS Analysis of *h*PFD Subunits 5 and 6

*h*PFD subunits 5 and 6 were buffer exchanged into a solution of AA at a concentration of 10  $\mu$ M and analysed by a Waters Synapt HDMS. The positive ion nanospray IM-MS spectrum of the *h*PFD subunits are shown in Figure 6.12. Above the conventional mass spectrum is a drift plot, showing the drift time of the ions present in the sample. It is seen in Figure 6.12 that 3 separate Gaussian state distributions can be clearly seen in each plot. A deconvolution of the MS spectrum reveals that the complexes observed here correlate to a monomer, dimer and trimer of each *h*PFD subunit.



**Figure 6.12:** IM-MS spectrum of (A&B) *h*PFD subunit 5, (C&D) *h*PFD subunit 6 (10  $\mu$ M, 100 mM AA, pH 7.4). (A&C) The IM drift plot of the *h*PFD subunits. Circled are different species of the subunits. (B&D) The conventional mass spectra of the *h*PFD subunits. The highlighted charge states correspond to (yellow) trimer, (purple) dimer & (green) monomer.

Figure 6.12B reveals that the relative abundance of monomeric *h*PFD subunit 5 is similar to trimeric *h*PFD subunit 5, suggesting a significant population for both species and stability of the oligomeric species. The dimers and trimers present in the spectrum are likely to be formed in solution due to a self-chaperoning or self-aggregation mechanism. This ultimately reveals the solution dynamics of *h*PFD subunit 5 and its ability to interact with itself to form complexes, as previously suggested in Chapter 5. In comparison, the predominant species in the MS analysis seen in Figure 6.12D is the monomeric *h*PFD subunit 6, followed by the dimer and then the trimer species. It is apparent that eukaryotic PFD subunit 6 has the ability to interact with itself to form complexes; however, in ‘native’ conditions at equilibria, the  $\beta$  subunit has a preference for the monomeric species.

It must be noted that the native MS acquisition of the *h*PFD subunits reveal spectra with peaks that are relatively broad. Unfortunately, the broad MS peaks are due to solvent adducts from the salts and detergent required to extract and solubilise the proteins from the *E. coli* cells. Extensive buffer exchanges were applied to the sample to remove the majority of salts and detergent, however complete removal could not be done without compromising the native state of the proteins.

The drift time of the complexes were extracted from the drift plot (Figure 6.12A&C) at the peak of the ATD and were converted to CCS values using a standard calibration protocol [426] and are displayed in Table 6.3.

**Table 6.3: The mass & calculated CCS values derived from the spectrum of the individually expressed *h*PFD subunit 5 analysed by TWIM Waters synapt (N<sub>2</sub>).**

Charge State	<i>h</i> PFD Subunit 5		<i>h</i> PFD Subunit 6	
	m/z	CCS (Å <sup>2</sup> )	m/z	CCS (Å <sup>2</sup> )
Monomer	22827 Da	Ave: 1823	20215 Da	Ave: 1688
+7	3261	1665	2980	1533
+8	2853	1794	2608	1680
+9	2536	1865	2318	1852
+10	2283	1969	2086	-
Dimer	45654 Da	Ave: 3039	40430 Da	Ave: 3029
+10	4565	2859	4044	-
+11	4150	2965	3676	2921
+12	3805	3085	3370	3015
+13	3512	3248	3111	3153
Trimer	68481 Da	Ave: 4670	60645 Da	Ave: 4368
+14	4892	4490	4691	4084
+15	4565	4605	4378	4182
+16	4280	4730	4104	4305
+17	4028	4858	3863	4385
+18	3805	5055	3648	-

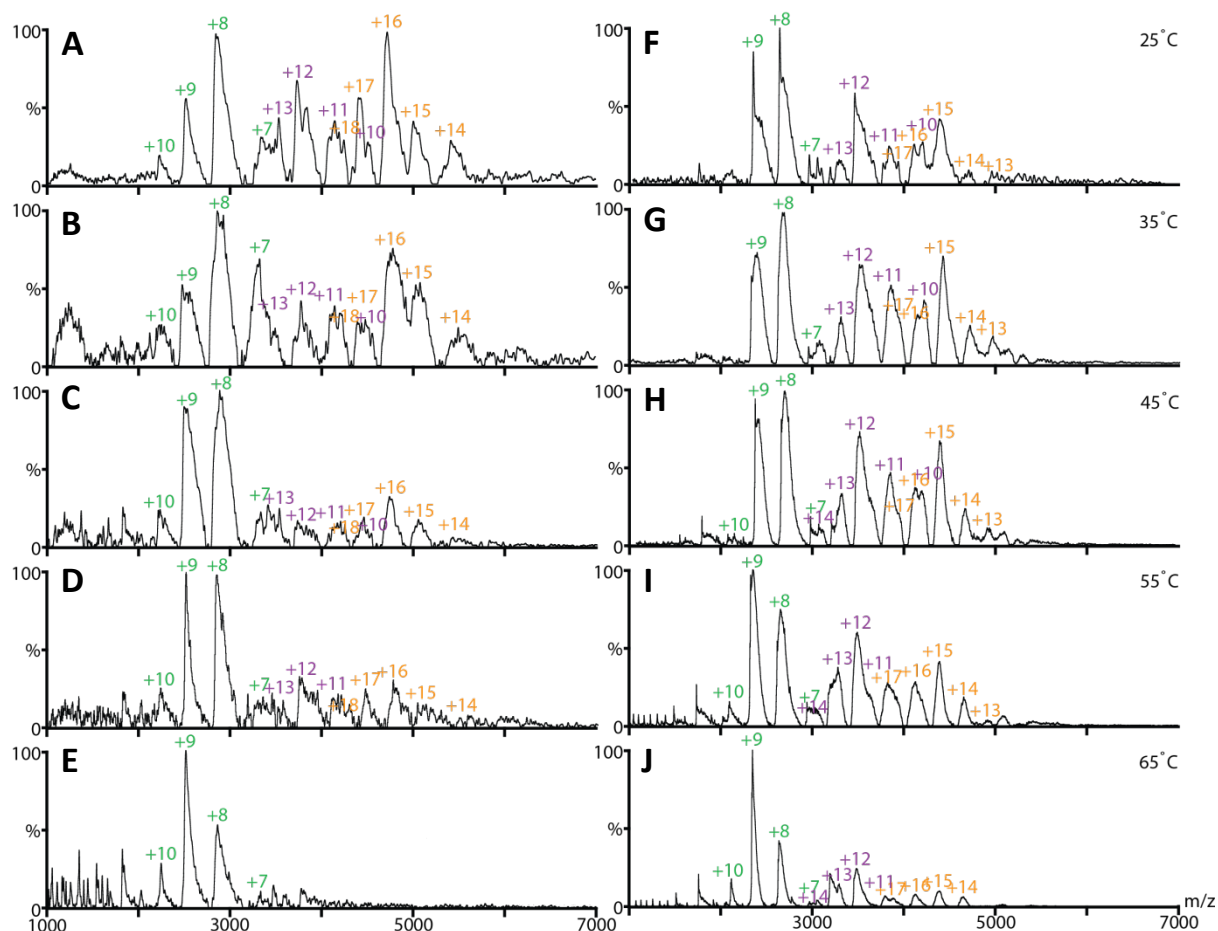
Averages are calculated over the charge state distribution.

Unsurprisingly, Table 6.3 reveals the calculated monomeric CCS for *h*PFD subunit 6 is significantly smaller than subunit 5, due to the respective amino acid sequence lengths. However, it is interesting to see that the average CCS values calculated for the dimers reveal a similar CCS. This suggests that the oligomerisation interactions of *h*PFD subunit 5 are more compact than for *h*PFD subunit 6 and can give insight to the binding mechanisms of  $\alpha$  and  $\beta$  subunits of eukaryotic PFD. It is suggested that the coiled coils of *h*PFD subunit 5 encase and intertwine itself to form a more compact structure, while *h*PFD subunit 6 may be interacting with itself through contacts at the distal tips, resulting in a slightly larger conformation.

#### **6.2.3.2: Thermal Denaturation IM-MS Analysis of *h*PFD Subunits 5 and 6**

To examine the thermal stability of the *h*PFD subunits by IM-MS, the CTC developed was utilised and spectra were acquired for the temperatures 25-65 °C. IM-MS parameters were set to 'native' conditions, to ensure all spectra acquired were reflective of the solution thermal dissociation conformers. Figure 6.13 illustrates the mass spectra of *h*PFD subunit 5 and *h*PFD subunit 6 analysed from 25-65 °C.





**Figure 6.13:** MS spectrum of (A-E) *hPFD* subunit 5 & (F-J) *hPFD* subunit 6 at different temperatures (25-65 °C, 10  $\mu$ M protein in 100 mM AA) utilising the CTC coupled to the Synapt HDMS Mass Spectrometer. The highlighted charge states correspond to (yellow) trimer, (purple) dimer & (green) monomer.

Figure 6.13 reveals an increase in the monomeric species for both *hPFD* subunits following the increase in temperature, consistent with dissociation observed in thermal denaturation of proteins [351], [363], [448], [449]. As it has already been established in Section 6.2.3.1 that *hPFD* subunit 5 exists predominantly as a trimer, the MS analysis at 35 °C shows that the abundance of ions correlating to the trimer (highlighted in orange, Figure 6.13B) has decreased at this temperature and the most abundant peak is now related to the monomeric molecule at charge state  $[M+8H]^{8+}$ , highlighted in green. Interestingly, the opposite is true for the *hPFD* subunit 6 (Figure 6.13B), and a marginal increase in relative abundance of the oligomeric species is observed at this temperature. Increases in temperature to 45 °C reveal little change to the MS spectrum for *hPFD* subunit 6 (Figure 6.13H). However, in Figure 6.13C for *hPFD* subunit 5, a further decrease in the relative abundance of the oligomeric species in comparison to the monomer is exhibited. Following further temperature increases to 55 °C (Figure 6.13D&I), a shift of the charge state distribution of the monomeric species is seen to move to the higher  $m/z$ , and the most predominant ion in both spectra correlate to charge

state  $[M+9H]^{9+}$ . A small decrease in the relative abundance of the oligomeric species for *h*PFD subunit 6 is also observed. The data obtained at 65 °C show further decreases in oligomeric ions for *h*PFD subunit 6 (Figure 6.13J), while there are only peaks relating to monomer and denatured *h*PFD subunit 5 present in the spectrum (Figure 6.13E), indicating complete dissociation of the *h*PFD subunit 5 dimer and trimer. Following the MS analysis of the thermal dissociation experiment, the ATD of the ions were extracted and the peak CCS value and CCSDs were derived and are plotted in Figure 6.14.

Figure 6.14 reveals that the ions corresponding to the monomer of *h*PFD subunits 5 and 6 illustrate different trends. Figure 6.14A shows that the ions relating to charge state  $[M+6H]^{6+}$ - $[M+9H]^{9+}$  of monomer *h*PFD subunit 5 are seen to change very little throughout the thermal aggregation assay. Although each ion exhibits a different CCS value, the increase in temperature is seen to affect the calculated CCS values by a maximum of 40 Å<sup>2</sup>. It should also be noted that at higher temperatures (55-65 °C), the ions corresponding to  $[M+6H]^{6+}$  and  $[M+7H]^{7+}$  were not observed. This is related to the charge state distribution shifting to higher charges following thermal activation that is synonymous with the unfolding of a macromolecule. The unfolding events can be clearly seen for the CCS value and CCSDs of charge state  $[M+10H]^{10+}$ , as the CCS value increases significantly from room temperature to 65 °C. In contrast, ions relating to monomer *h*PFD subunit 6 of charge states  $[M+7H]^{7+}$  and  $[M+8H]^{8+}$  are seen to have relatively small CCS increases.

Similar to the monomer ions of charge state  $[M+10H]^{10+}$  of *h*PFD subunit 5, the ions correlating to the dimeric form is seen to have very small changes with the temperature increase from 25-35 °C. However, the analysis at 45 °C results in slightly larger increases of the CCS values, which is continued at 55 °C. Further temperature increases revealed no ions correlating to dimers or trimers were observed at 65 °C, therefore indicating complete dissociation of the oligomeric species. Figure 6.14D illustrates that the ions for the trimer of *h*PFD subunit 5 behaved in a relatively similar manner to the dimer. The temperature change from 45-55 °C was seen to initiate the largest CCS value increase seen for the majority of the ions, and no trimer ions were observed in the analysis at 65 °C.

It is interesting to see that the trend seen for *h*PFD subunit 5 is observed for *h*PFD subunit 6 ions correlating to the dimer and trimer species. The ions illustrated in Figure 6.14E-F are seen to increase in size following the temperature increase from room to 35 °C. The ions are seen to remain at this size until the temperature reaches 65 °C, where another increase in CCS is seen. The exception of this trend is observed for the charge state of  $[M+14H]^{14+}$  of the dimer and charge state of  $[M+15H]^{15+}$  of the trimer, where the ions of  $[M+14H]^{14+}$  have a calculated CCS value that remains

constant throughout the assay and  $[M+15H]^{15+}$  of the trimer is seen to follow a steady increase in CCS from 25-45 °C and remains at the CCS of 4479 Å<sup>2</sup> until the end of the assay.

Interestingly, the relative CCSDs for the ions relating to the monomers of the *h*PFD subunits appear to be larger than the ions correlating to the dimers and trimers. As previously mentioned in Section 6.2.2.3, the resolution of traveling wave instruments is 0.025 (2.5 %), [446], [447]. All distribution widths observed here are greater than the resolution of the instrument and therefore reflect conformational assemblies and flexibility of the subunits. The normalised CCSDs relating to the monomer of *h*PFD subunit 5 are calculated to range from 0.051 (5.1 %) to 0.173 (17.3 %). It is highly likely that the large distributions observed here are a result of molecular dynamics of the self-chaperoning mechanism. In contrast, the CCSDs for the dimeric species is seen to range from 0.050 (5.0 %) to 0.088 (8.8 %), while the relative distributions seen for the trimers have a slightly larger range from 0.046 (4.6 %) to 0.110 (11.0 %). The ions corresponding to *h*PFD subunit 5 monomers are observed to have the biggest range in comparative CCSDs following the thermal dissociation assays. This suggests flexibility of the monomer conformation, indicating a higher number of conformational assemblies in comparison to the dimers and trimers. In comparison, the relative CCSDs of dimers appear to have a small range and this can be explained through the activation of the self-chaperoning mechanism equating to a more rigid structure. This reduced conformational flexibility is also seen for the trimeric species, however, to a smaller extent.

A similar trend to *h*PFD subunit 5 is observed for the calculated relative CCSDs of *h*PFD subunit 6. The ions corresponding to monomers have a normalised distribution of 0.059 (5.9 %) to 0.128 (12.8 %). This range is significantly smaller than the calculated range for *h*PFD subunit 5, indicating that monomeric subunit 6 is a less dynamic compound in comparison to monomer *h*PFD subunit 5. The relative distributions of the dimer are between 0.047 (4.7 %) and 0.108 (10.8 %) and for the trimer the lowest comparative CCSDs is at 0.063 (6.3 %) and the largest is calculated to be 0.117 (11.7 %). It is therefore proposed that the conformational assemblies and structural integrity of the *h*PFD subunit 6 species appear to remain relatively similar throughout the thermal dissociation assay.

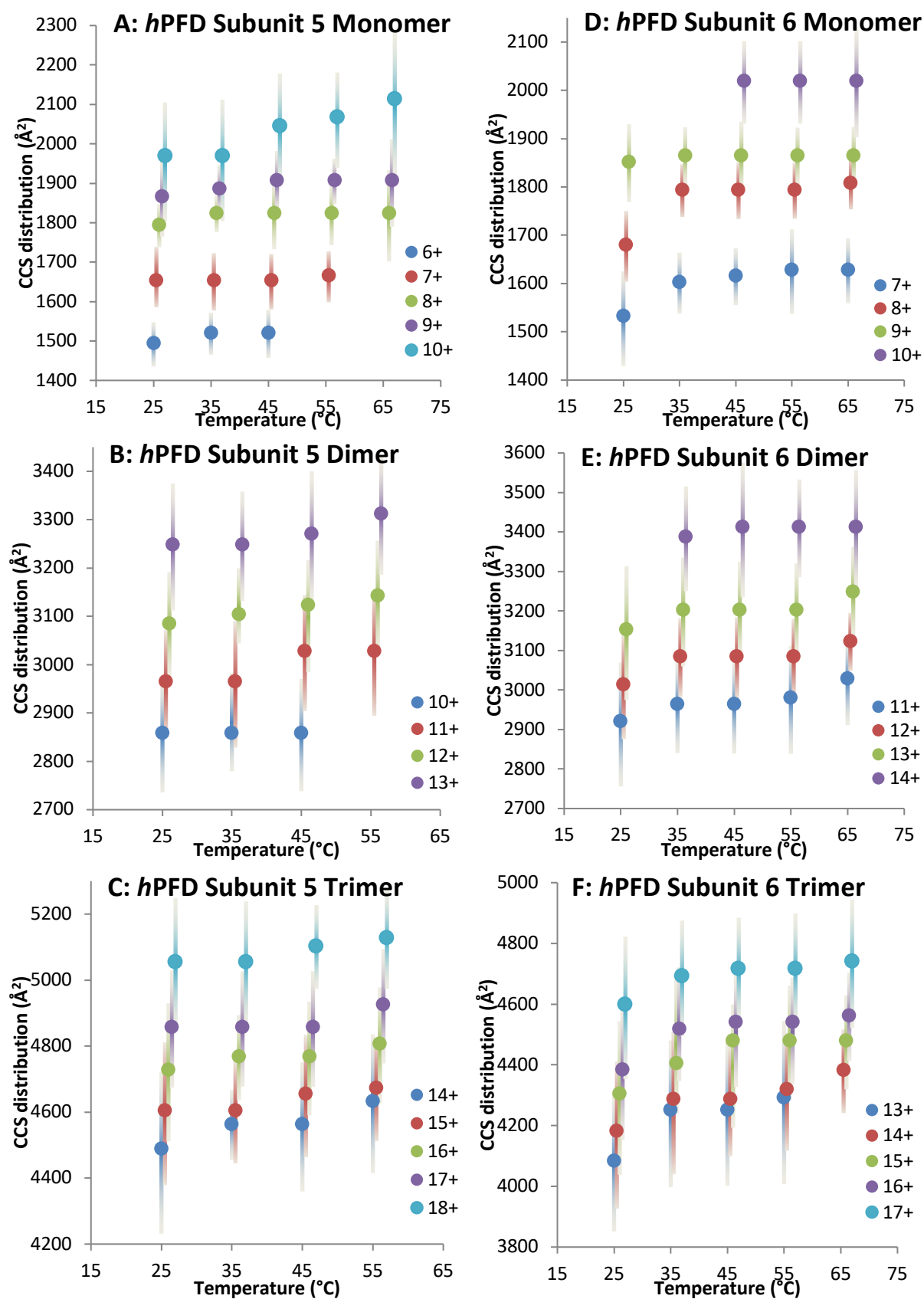


Figure 6.14: The CCSDs derived from the IM-MS thermal dissociation experiments of the (A-C) hPFD subunit 5 & (D-F) hPFD subunit 6. The product ions seen for the (A&D) monomer, (B&E) dimer & (C&F) trimer of the respective subunits are plotted as a function of temperature.

To summarise, the native IM-MS data obtained for the *h*PFD subunits reveal that the molecules free in solution at a temperature of 25 °C were able to interact to form homo-oligomers. Although it has been stipulated in previous studies that oligomerisation can occur due to the concentration of the sample, nanospray of proteins at these micro molar concentrations have been previously calculated to contain on average one molecule per droplet [450]. It is interesting to see that *h*PFD subunit 5 trimer is the predominant form found in the solution in native IM-MS analysis, while the folded monomeric form is the predominant species in the data obtained for *h*PFD subunit 6. This indicates that *h*PFD subunit 5 is more active in self-association, reinforcing the molecular dynamics seen for the subunits in Chapter 5. The results here can also be compared to previously published data of the prokaryotic PFD, with the *h*PFD subunits 5 and 6 representing their respective subunit class. The published MS analysis of the *mt*PFD  $\alpha$  subunit in low-energy conditions showed peaks correlating to monomers and dimers, with the dimers the predominant species in the spectrum [94]. Fändrich *et al.* suggested that the *mt*PFD  $\alpha$  subunit formation was due to the dimeric core seen in the intact hexameric PFD complex. The results shown here reveal the dominance of the trimeric form by *h*PFD  $\alpha$  subunit 5 which do not align with the published data for prokaryotic PFD. In contrast, the analysis of *mt*PFD  $\beta$  subunit was found to be monomeric under all MS conditions applied [94]. Although, trimers and dimers of *h*PFD subunit 6 were observed in the spectrum, the monomeric species was the predominant form, consistent with the data for the prokaryotic *mt*PFD  $\beta$  subunit. It is possible that the formation of the oligomers by the *h*PFD subunits may be due to a self-chaperoning mechanism, however it is also plausible that the trimers formed may have some underlying biological chaperone function, with the incorporation of additional subunits potentially increasing the molecular cavity resulting in an increase in substrate binding and therefore improved chaperone efficiency. The different propensities observed for trimer formation by *h*PFD subunits 5 and 6 could be a contributing factor to the fibril inhibition efficiencies reported for each subunit in Section 3.2.4.1.

It has been previously determined that eukaryotic PFD and its subunits are holdase chaperones, with biological activity likely to be achieved through a transient hold and release mechanism. The required chaperone to substrate ratios are typically determined through concentration dependant studies. However, the free forming dimers and trimers were not monitored in our studies and it is possible that the self-oligomerisation of the *h*PFD subunits may have affected the ThT assays conducted in Section 3.2.4.1. There are 2 possible scenarios arising from the self-oligomerisation of the *h*PFD subunit chaperones. The first considers the chaperone to substrate ratio remaining the same throughout the self-oligomerisation process despite the increase in chaperone size. As the *h*PFD subunit binds to itself, this affects the number of units available to interact with the misfolding

peptide and consequently detrimentally impacts the chaperoning efficiency. Alternatively, the *h*PFD subunits require the formation of oligomers to efficiently perform holdase activity, as seen in a previously published study of *mj*PFD, where the prokaryotic PFD  $\alpha$  subunit is required to form oligomers for chaperone activity [87]. Both possible scenarios lead to the overall diminished number of available units for chaperone holdase function.

These assumptions and possible scenarios can also be compared to the experiments conducted with *b*PFD, where self-oligomerisation is not seen. A trend is therefore proposed between the propensity for oligomerisation and chaperone activity. It is suggested that the increase in self-oligomerisation results in the decrease of macromolecule units available for holdase function and subsequently results in a decrease in chaperone activity. This correlates the IM-MS results seen here and the fibril inhibition efficiencies observed for *h*PFD subunit 5, *h*PFD subunit 6 and *b*PFD in Section 3.2.1.4.

Further observation of the *h*PFD subunit thermal aggregation data shows the difference in thermal dissociation between the eukaryotic PFD subunits 5 and 6. Increases in temperature were seen to dissociate the oligomers of both *h*PFD subunits, with steady decline in ions correlating to the trimers and dimers observed until complete dissociation at approximately 55 °C. Although both subunits were seen to behave similarly under thermal conditions, there were a few key differences observed in the IM-MS analysis that reveal the different thermal dissociation pathways of the *h*PFD subunits. Following increasing temperature, the IM analysis of *h*PFD subunit 5 was seen to exhibit minor changes in the calculated CCS values of folded monomer while the MS spectra revealed that the relative abundance of ions shifted towards the unfolded monomer. In comparison, the MS data collected for *h*PFD subunit 6 showed that the folded monomeric conformation was the predominant species throughout the entire experiment, while the IM analysis revealed that there were large CCS changes occurring to the molecule. This ultimately shows that the *h*PFD subunits undergo different thermal dissociation pathways and reinforces the resilience of *h*PFD subunit 6 to thermal degradation, which was also observed in Section 4.2.3.

Analysis of the thermal aggregation assay conducted with intact *b*PFD establishes that the complex dissociates into trimer sub-complexes. The data presented in this Chapter suggests the dissociated trimers have strong binding interactions and are stable in conditions of high kinetic energy, in particular the sub-complex consisting of eukaryotic PFD subunits 1, 5, and 6 was observed throughout the thermal dissociation experiment. It should be noted that trimers were also observed in the native IM-MS analysis of the intact *b*PFD complex. This may suggest that the protein readily dissociates into its respective trimers. Although the dissociation of eukaryotic PFD has not been reported before, sHSP chaperones have been previously reported to readily modulate quaternary

structures for biological function, such as dissociate into dimers and monomers during heat shock [451]–[453]. It is possible that the dissociation observed here for *b*PFD into its respective trimers may have the same implications, however further studies are required to investigate the validity of this.

In conclusion, the experiments in this Chapter were able to verify the stoichiometric arrangement of the subunits in *b*PFD and show possible internal interactions of the subunits. The propensity for trimer formation of the *h*PFD subunits 5 and 6 when free in solution combined with the results seen in Chapter 3 establish a possible trend between self-oligomerisation and fibril inhibition efficiencies. In addition, it was previously reported in Chapter 4 that the *h*PFD subunits, in particular subunit 6, were tolerable to high temperature changes and this was further validated here. The IM-MS data was also able to validate the molecular dynamics of the *h*PFD subunits as was previously suggested in Chapter 5.

## 6.3: Experimental

All materials were purchased from Sigma Aldrich (USA) unless otherwise stated.

### 6.3.1: Sample Preparation

*b*PF<sub>D</sub> was purified from bovine testis tissues as described in Chapter 3. The resulting intact *b*PF<sub>D</sub> complex solution was adjusted to 10  $\mu$ M with an aqueous solution (50 mM AA at pH 7.4 unless otherwise stated). The concentration of *b*PF<sub>D</sub> was determined spectrophotometrically by measuring the absorbance at 280 nm and using the calculated extinction coefficient for each individual subunit (Refer to Appendix E). The concentration of the contaminants was deducted from the overall calculated concentration.

The subunits of *h*PF<sub>D</sub> were expressed and purified from *E. coli* as described in Chapter 3. The subunit solutions were adjusted to a concentration of 10  $\mu$ M (50 mM AA at pH 7.4 unless otherwise stated). The concentration of *h*PF<sub>D</sub> was determined spectrophotometrically by measuring the absorbance at 280 nm, using a calculated extinction coefficient (Refer to Appendix E).

### 6.3.2: Ion-mobility Mass Spectrometry

'Native' like IM-MS spectra were acquired on a Synapt HDMS system (Waters, UK) [114], using nanospray in the positive ion mode. The sample was introduced using platinum-coated borosilicate capillary needles that were prepared in-house. The Synapt HDMS instrument parameters were optimised to remove the majority of adducts whilst preserving non-covalent interactions, and were typically as follows; capillary voltage, 1.6 kV; cone voltage, 60 eV; trap collision energy, 10 V; transfer collision energy, 10 V; source temperature, 50 °C; backing pressure, 5 mBar; IMS cell pressure (N<sub>2</sub>), 0.5 mBar; travelling wave velocity, 350 ms<sup>-1</sup>; travelling wave height, 9-11 V.

### 6.3.3: Data Analysis

Peaks in the experimental ATD are fitted by Gaussian curves and the CCS values were derived from the drift time corresponding to the peak of the ATD. The peak drift times were obtained from the Synapt HDMS were normalised for charge state and a nonlinear correction function was applied for calibrant ions such that their relative differences mirror those previously observed for the same ions as described in detail elsewhere [206], [231], [426]. CCSs of the reference samples were taken from the literature using values for ubiquitin, myoglobin, cytochrome c and serum amyloid P component [454].



CCSDs were derived from the FWHM of the ATD and the smallest and largest values of the range were converted into CCS values in a similar fashion described for the CCS values obtained from the peak ATD. CCSDs were plotted as violin plots.

CCSs of model protein structures were calculated using the Leeds Method (developed by Ashcroft and co-workers). Structural coordinates from previous studies using NMR and X-ray crystallography were obtained from the Protein Data Bank (PDB) with accession numbers indicated in the text.

#### **6.3.4: Thermal Denaturation Experiments**

Following the experiments of 'native' IM-MS, the CTC was attached to the glass capillary and the temperature was allowed to stabilise for 60 seconds before acquisition was initiated. A series of IM-MS acquisitions at different temperatures (over the range of 25-65 °C) were acquired on the Synapt HDMS system (Waters, UK) using identical experimental parameters to 'native' like conditions: capillary voltage, 1.6 kV; cone voltage, 60 eV; trap collision energy, 10 V; transfer collision energy, 10 V; source temperature, 50 °C; backing pressure, 5 mBar; IMS cell pressure (N<sub>2</sub>), 0.5 mBar; travelling wave velocity, 350 ms<sup>-1</sup>; travelling wave height, 9 V.

# Chapter 7:

## Chemical Cross-linking of the Human Prefoldin Subunits 5 and 6

---

### 7.1: Introduction

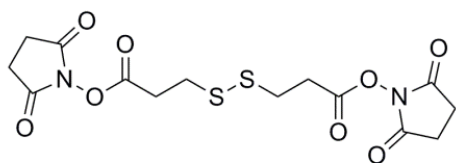
Chemical cross-linking (CXL) has the powerful ability to covalently link interacting molecules together. As a result of this property, CXL has developed over the last 2 decades to provide advances in structural biology [455]. The combination of CXL, proteolytic digest and MS analysis can provide distance constraints for structural modelling and bind transient interactions. CXL studies are now commonly employed to elucidate low resolution tertiary structures [262], [456]–[459] or to probe protein-protein interactions [113], [266], [267].

As described previously in Chapter 2, CXL bottom up studies typically share a common workflow, with minor deviations in the analysis protocol depending on the CXL reagent used and the particular instrumental analysis required [457], [460]–[463]. Although the aim of any cross-linking reaction is to form covalent bonds between proximal residues to elucidate native conformations or investigate protein-protein interactions, other reaction products are possible. The reactive groups of the cross-linking reagent may not bio-conjugate with an appropriate residue or may be deactivated prior to bio-conjugation through hydrolysis. Alternatively, a peptide may contain several CXL modifications. This therefore results in a complex mixture of products including unmodified proteolytic peptides, dead-end, intramolecular and intermolecular cross-linked peptides, or higher order adducts (Figure 2.10).

Although the plethora of products increases the complexity of data analysis, identification of each cross-linking adduct type can provide some structural information. For example, dead-end cross-linking adducts are able to give solvent accessibility of the native conformation, intramolecular cross-links may give information about local structural elements i.e. secondary and tertiary structure, whilst intermolecular cross-links provide distance information that may be used to define binding interfaces [261], [274], [458], [460], [461]. Together, this information can generate distance constraints within a protein or protein complex that can be employed to approximate the higher

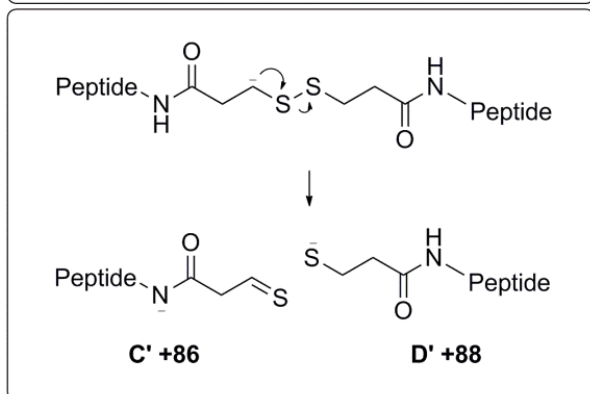
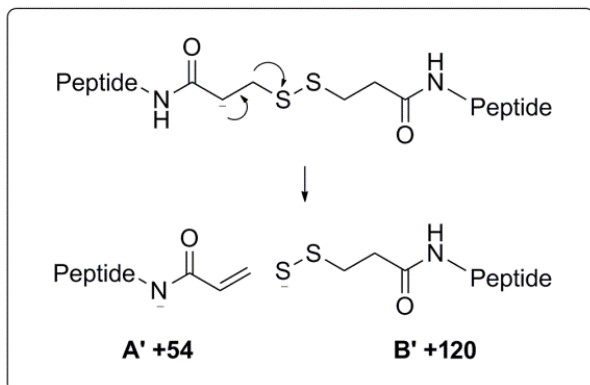
order structure of the complex. Analysis of the CXL data is often difficult and time consuming; here we present several approaches to facilitate the identification of cross-linked peptides, the first is the incorporation of a facile cleavable spacer arm within the CXL reagent.

CXL reagents designed to be readily cleaved by collision induced dissociation (CID) allow cross-linked peptides to be rapidly identified through their unique fragmentation patterns obtained by MS/MS. The facile CID fragmentation pathways of sulphur containing linkers in peptide and protein systems by both positive and negative ion MS are exploited in this chapter to rapidly identify cross-links. The popular commercially available homobifunctional cross-linking reagent dithiobis(succinimidyl propionate) (DSP), undergoes facile fragmentation of the disulphide bond upon CID in negative ion mode, giving rise to characteristic fragments of cross-linked peptides enabling easy identification of the cross-linking type and site. Scheme 7.1 shows the expected products of DSP fragmentation from intermolecular and dead end cross-linked peptides in negative ion MS/MS mode. Upon CID, peptide anions cross-linked with DSP will give rise to characteristic fragments of peptide +54 Da (A'), +120 Da (B'), +86 Da (C'), and +88 Da (D') [462], [463]. Although DSP products have been analysed by MS in both positive and negative ion mode [463], [464], it has been reported that positive ion MS analysis of DSP fragments are difficult to identify due simultaneous backbone fragmentations observed in the spectra [260], [465].

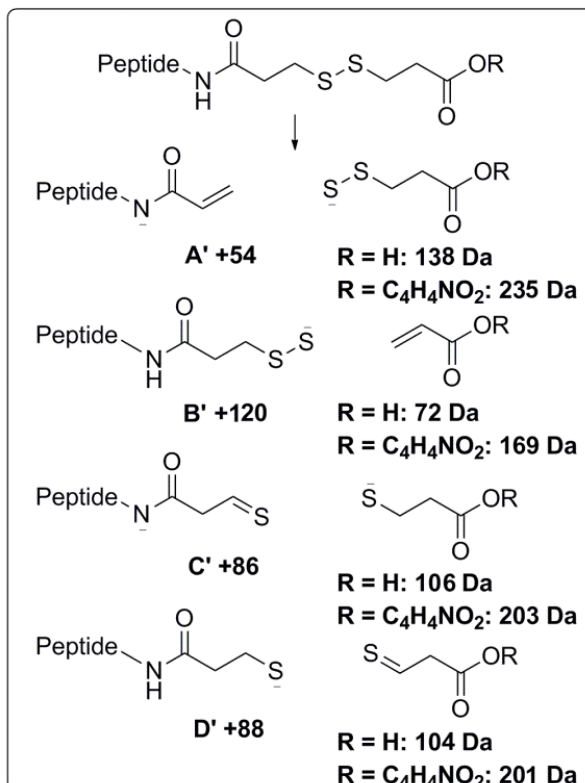


Dithiobis(succinimidyl propionate) (DSP)

Intermolecular cross-link

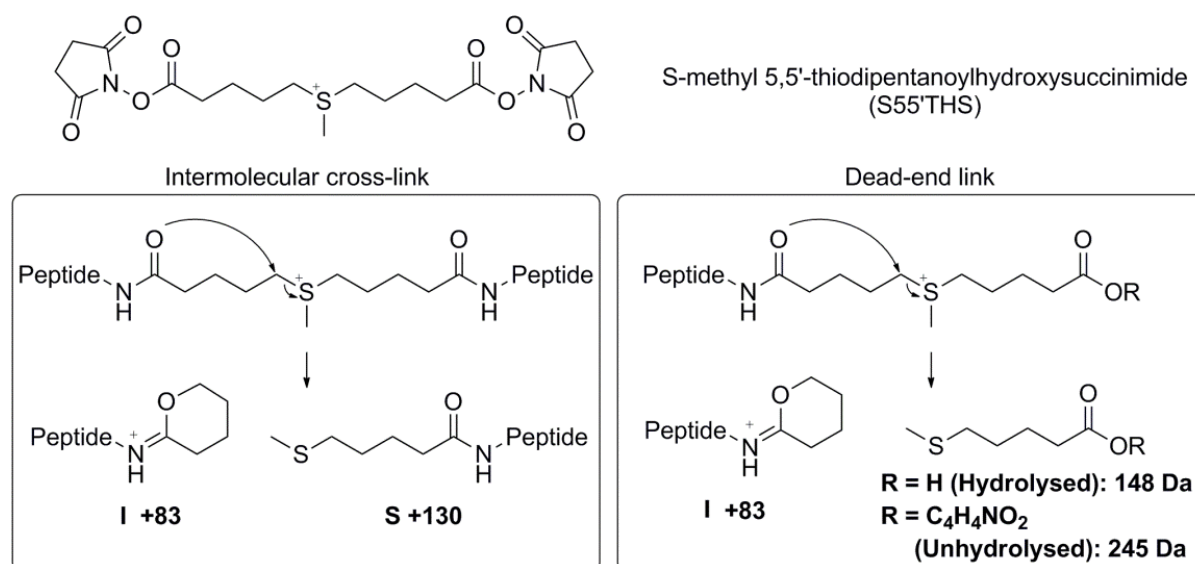


Dead-end link



**Scheme 7.1:** Expected fragmentation pathways for DSP intermolecular & dead end cross-linked peptides in negative ion MS/MS mode. DSP fragments about the disulfide bond are labelled A'-D' consistent with fragmentation pathways described by Calabrese *et. al* [463].

Lu *et. al* described the synthesis and application of a novel sulphonium ion-containing cross-linking reagent, S-methyl 5,5'-thiodipentanoylhydroxysuccinimide (S55'THS), designed to fragment exclusively via facile cleavage of the C-S bond directly adjacent to the positively charged sulphonium ion during positive ion MS/MS (Scheme 7.2) [466]. As seen in Scheme 7.2, CID results in a 6 membered heterocyclic iminotetrahydropyran (I) that increases the mass of the peptide by 83 Da, and a s-methylthiopentanoyl (S) fragment which correspondingly increases the peptide mass by 130 Da [466].



**Scheme 7.2:** Expected fragmentation pathways for S55'THS intermolecular & dead end cross-linked peptides in positive ion MS/MS mode. Fragments are labelled according to Lu *et al.* corresponding to formation of a 6-membered iminotetrahydropyran (I) & s-methylthiopentanoyl (S) group [466].

Both DSP and S55'THS are homobifunctional cross-linking reagents with two active NHS ester groups. As previously mentioned in Chapter 2, NHS esters principally target primary amines ( $-\text{NH}_2$ ) for bio-conjugation and in proteins, these are present either at the end of the amino acid polypeptide chain at the N-terminus or in the side chain of Lys [265]. S55'THS is able to cross-link peptides at distances up to approximately 20 Å (taken from C $\alpha$ -C $\alpha$  to allow for side chain flexibility), whereas DSP is able to cross-link up to a distance of 12 Å [265].

Traditional CXL workflows utilise software dedicated to the automated identification of cross-linked peptides to filter the plethora of products that are obtained in CXL experiments [273], [467]–[470]. Of these software packages, this work utilises the unique fragmentation identification feature of MeroX. MeroX was developed by Götze *et. al* to analyse LC-MS/MS spectra of facile CID cleavable CXL reagents [273]. The algorithms in this software generate all possible cross-linking combinations from the sequence database and subsequently match the theoretical cross-links to the acquired MS/MS spectra. Input of the relevant FASTA sequence file, the exported LC-MS/MS data in mgf format and the specificities of the cross-linker are therefore required for automated peptide cross-link identification. Scoring of the matched candidate and applying a mix-target decoy are also options available in MeroX to determine false identifications. Although MeroX is able to identify potential cross-links, ambiguity in the position of the cross-link can still be present. Previous LC-MS/MS analyses of cross-linked peptides have used 'pseudo MS3 experiments' to enable the sequencing of the intermolecular cross-linked peptides through MS/MS experiments on product ions formed by inducing capillary-skimmer dissociation in the ion source [128], [246]. These experiments

are useful for sequences with more than one Lys chain available for cross-linking as characteristic modifications of cross-linked Lys side chains allows for the identification of the exact location of the cross-linking site in protein systems of unknown structure. Unfortunately, spectra acquired at high CID can cause side chain fragmentation in addition to normal backbone fragmentation, complicating interpretation.

As previously explained in Chapter 2, orthogonal separation by IM allows for the simplification of MS spectra. Briefly, the selection of a drift time region combines the MS spectra of the detected ions with that selected mobility, therefore removing contributions of ions from outside the selected drift time. In systems without LC, IM can be utilised as an alternative separation method. This simplification of MS is exploited in our CXL analysis method when analysing with cross-linked products with a cleavable spacer arm. In addition CID is achieved following IM and prior to MS analysis, therefore resulting in the MS detected fragment ions having the same drift time as the precursor ions. This phenomenon, known as time aligned fragmentation, simplifies the spectra and due to efficient separation of size to charge, the same peptides with different CXL sites are able to be separated.

### 7.1.1: Aim

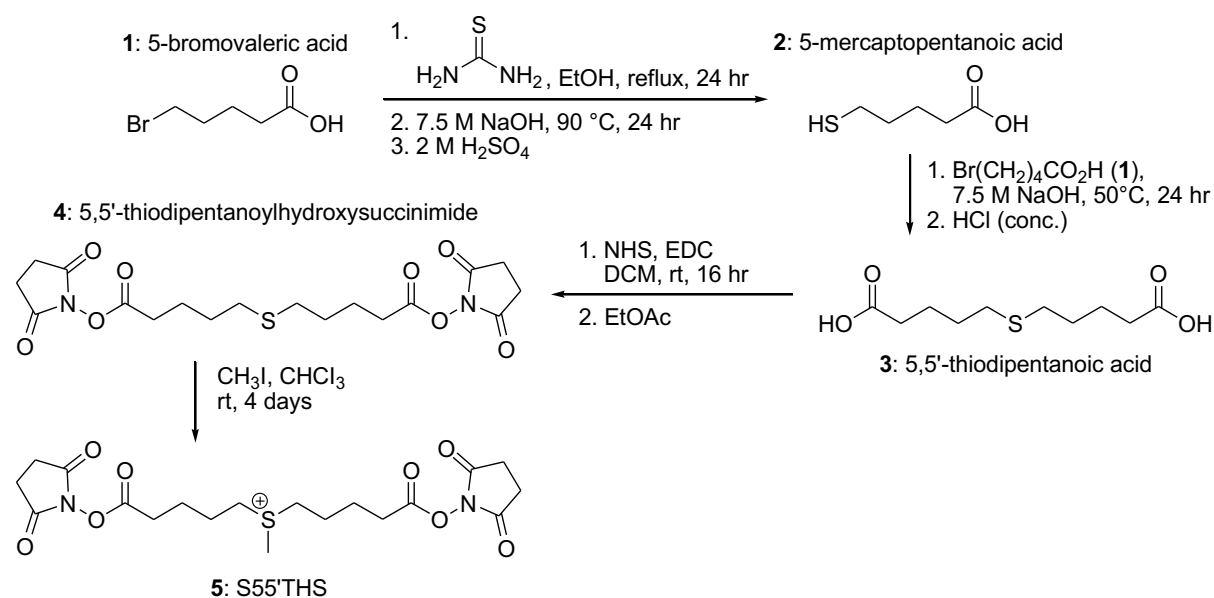
This chapter aims to compare the analysis of cross-linked protein samples by IM-MS, LC-MS/MS or LC-IM-MS/MS, utilising cleavable linkers, to inform the optimum workflow for protein CXL analysis.

To effectively utilise the CID spacer arm of the cross-linking reagents and the time aligned fragmentation to aid in cross-link identification, the diagnostic fragmentation pathways of the cross-linked products must be understood. The aim of this chapter is to develop a systematic CXL approach to identify protein cross-links in a site specific manner, and enable the construction of low resolution tertiary structures. The effectiveness of this method is first demonstrated with simple peptides and the structurally well characterised protein, ubiquitin, before attempting this method on *h*PF<sub>2</sub> subunits 5 and 6. An additional aim of this chapter is therefore to determine the binding interfaces of *h*PF<sub>2</sub> subunits 5 and 6 to A $\beta$  (1-40) through the newly developed CXL approach.

## 7.2: Results and Discussion

### 7.2.1: Synthesis of Cross-linker S55'THS

S55'THS was first synthesised based on the general method of Lu *et al.* [466] in a four step procedure starting with the cheap and readily available reagent 5-bromovaleric acid **1**. By this approach, 5-bromovaleric acid **1** was first converted to 5-mercaptopentanoic acid **2** under standard conditions (thiourea and ethanol). **2** was then converted to the dicarboxylic acid **3** with the addition of acid **1** in a basic aqueous solution at 50 °C via a nucleophilic substitution as seen in Scheme 7.3. The dicarboxylic acid **3** was purified over silica gel to give a white solid (92%).



Scheme 7.3: Four step synthesis of S55'THS cross-linking reagent adapted from Lu *et al.* [466].

Although Lu *et al.* utilised the insoluble coupling reagent N,N'-dicyclohexylcarbodiimide (DCC) to conjugate N-hydroxysuccinimide (NHS) to the dicarboxylic acid **3**, here it was found that the reagent DCC was challenging to solubilise in DCM, and the difficult removal of by-products compromised the yield of product **4**. Therefore the coupling reagent N-ethyl-N-(3-dimethylaminopropyl)carbodiimide (EDC) was used instead.

Experimentally, the mixture of EDC, NHS and dicarboxylic acid **3** were dissolved in DCM and stirred at room temperature for 16 hours. Isoarea and unreacted reagents were removed via silica gel filtration to afford a purified oily residue of cross-linker **4**. To obtain the desired fixed positive charge, methylation of sulphur was conducted with the addition of iodomethane dissolved in

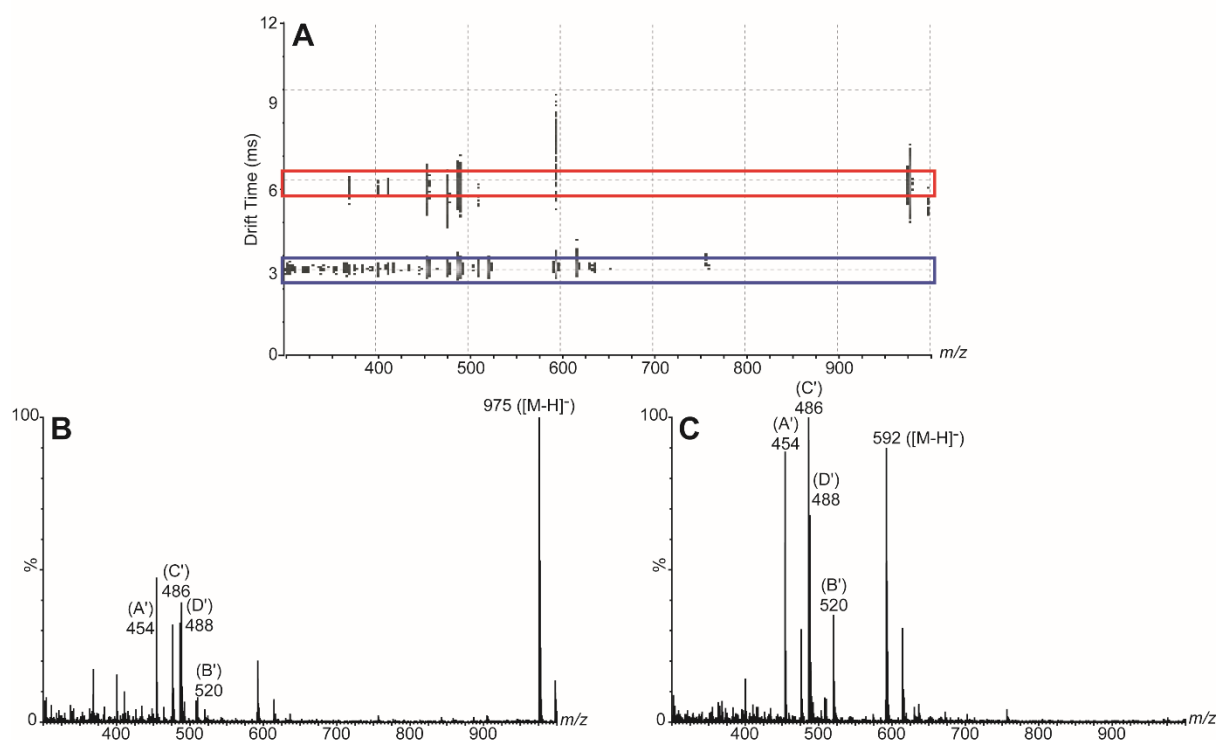
chloroform to give product **4** and this was stirred at room temperature for 4 days to afford product **5**, S55'THS (crude 78%). It was found that due to the moisture sensitive nature of S55'THS, it could not be purified without compromising the NHS reactive ends. Using crude S55'THS will likely increase the amount of by products in the CXL reaction; however it will not detract from the overall experiment. S55'THS was used in CXL experiments without further purification.

### 7.2.2: Cross-linking of Ac-AAKA

To experimentally demonstrate the efficient bio-conjugation of the reagents to primary amines and subsequent low-energy CID fragmentation of intermolecular cross-links, DSP and S55'THS were cross-linked to the model peptide Ac-AAKA ( $401.5 \text{ g.mol}^{-1}$ ) under conditions known to generate high yields of cross-linked adducts. Due to the acetylation of the N-terminus, this peptide contains a single free amine group on the side chain of the Lys. Subsequently, the cross-linking reagents may conjugate only at this amine and the possible products for this experiment include one unmodified peptide, one dead-end cross-link and one intermolecular cross-link product. Although this CXL experiment is not biologically relevant, it is able to validate NHS-ester containing cross-linking bio-conjugation and it is able to define the characteristic fragmentations of the resultant adducts [462], [466]. Following the cross-linking reaction, the organic solvent was removed and the sample was desalted. Subsequent analysis by IM-MS applying CID in the transfer region enables time aligned fragmentations. Intermolecular cross-linked peptides, dead-end products and unreacted peptide were present, as identified by the expected mass, and time aligned fragmentation data were recorded at low collision energy (Figure 7.1, 7.2).

Figure 7.1A shows the IM-MS/MS analysis of DSP cross-linking of Ac-AAKA. The intermolecular cross-links produced were identified based on  $m/z$  (975  $m/z$ ) at a corresponding drift time range of 5.8-6.1 msec (Figure 7.1A&B). The product ions of the DSP intermolecular cross-links are analogous to the natural Cys disulphide fragments described in Scheme 7.1, with all 4 possible characteristic product ions, A' to D' of 454, 486, 488 and 520  $m/z$  easily recognised, allowing rapid identification of the intermolecular cross-linked peptides. The singly charged dead-end cross-links can also be clearly seen in the spectrum, with the same CID product ions observed at a different drift time of 3.1-3.3 msec. The lower drift time is consistent with the smaller CCS of dead-end modified peptides (Figure 7.1A blue) compared with those that are intermolecular cross-linked (Figure 7.1 red).

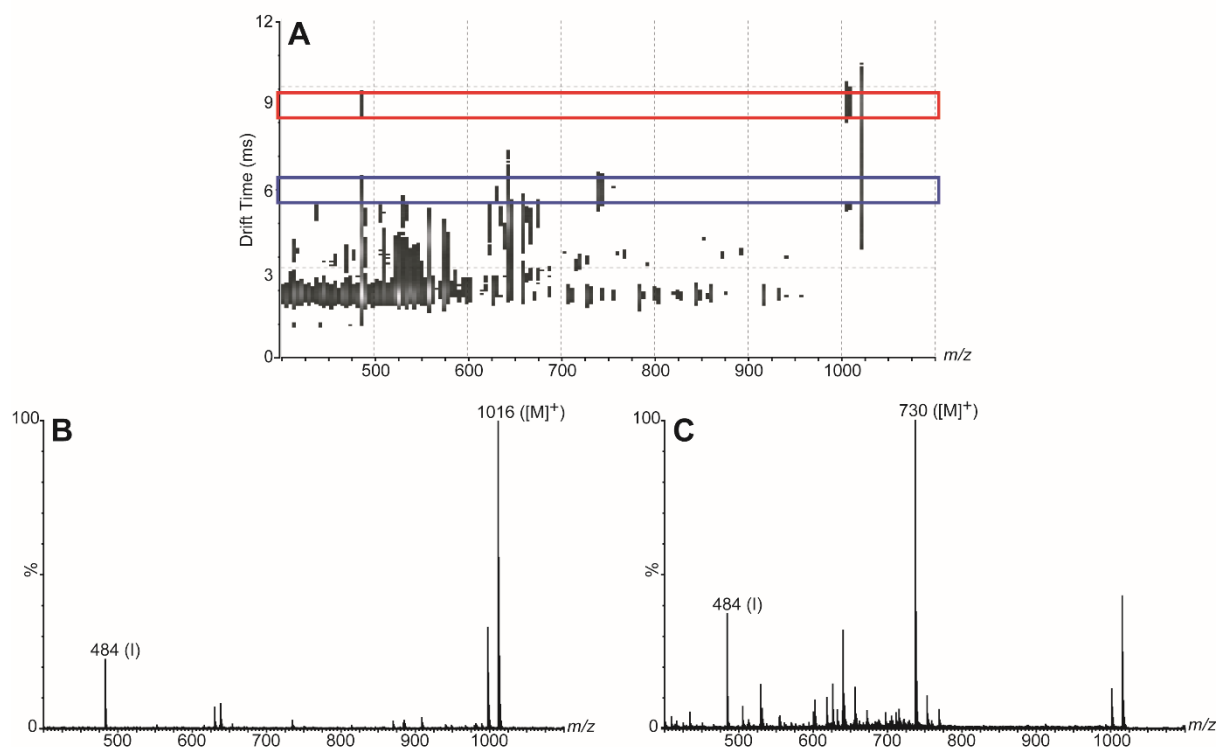




**Figure 7.1:** (A) Negative ion IM-MS/MS analysis of DSP cross-linked peptide Ac-AAKA using Waters Synapt. (B) Summation of the drift time 5.8-6.1 msec highlighted in red generates the simplified mass spectrum of the intermolecular cross-link (975  $m/z$ ). Here the intermolecular cross-linked peptide & its time aligned fragmentation products (the disulphide fragments labelled A'-D' at 454, 486, 488 & 520  $m/z$ ) are readily identified. (C) Summation of the drift time 3.1-3.3 msec highlighted in blue generates the simplified mass spectrum of the dead-end modified cross-link (592  $m/z$ ). The time aligned fragmentation products (the disulphide fragments labelled A'-D' at 454, 486, 488 & 520  $m/z$ ) are readily identified.

Positive ion MS analysis of DSP cross-links has been reported, and fragmentation of the reagent reveals 2 peaks with a 66 Da mass difference as a result of disulphide bond cleavage [465], [471]. However, these signals also occur in addition to the abundant backbone fragmentations which dominate the spectrum and are therefore difficult to observe. Consequently, our CXL methods utilising reagent DSP are only analysed by negative ion MS.

Figure 7.2A shows the positive ion IM-MS/MS of cross-linked Ac-AAKA using S55'THS and a range of products is observed. The singly charged intermolecular cross-linked peptides (1016  $m/z$ ) are observed to have an arrival time distribution (ATD) of 8.9-9.2 msec. Summation of this drift time range reveals the simplified spectrum shown in Figure 7.2B. The time aligned fragment product I (the 6-membered iminotetrahydropyran, 484  $m/z$ ) is readily identified. From Figure 7.2A, MS summation of the drift time 5.7-6.2 msec generates the spectrum in Figure 7.2C. The singly charged dead-end cross-link is also seen in Figure 7.2A and the MS spectrum clearly reveals the corresponding fragment product ion I is aligned to the precursor ion (730  $m/z$ ) by drift time.



**Figure 7.2:** (A) Positive ion IM-MS/MS analysis of S55'THS cross-linked peptide Ac-AAKA using Waters Synapt. (B) Summation of the drift time 8.9-9.2 msec highlighted in red generates the simplified mass spectrum of the intermolecular cross-link (1016  $m/z$ ). Here the intermolecular cross-linked peptide & its time aligned fragmentation product (the 6-membered iminotetrahydropyran labelled I, 484  $m/z$ ) are readily identified. (C) MS summation of the drift time 5.7-6.2 msec highlighted in blue generates the simplified mass spectrum of the dead-end modified cross-link (730  $m/z$ ). Here the modified peptide & its time aligned fragmentation product (the 6-membered iminotetrahydropyran labelled I, 484  $m/z$ ) are readily identified.

Fragmentation of S55'THS modified peptides seems to favour formation of the product ions containing iminotetrahydropyran (I) rather than s-methylthiopentanoyl (S) modifications. This is consistent with the analysis of Lu *et al.* who suggested that the two C-S bonds do not have identical cleavage efficiencies, hypothesised to be due to different “solvation” environments of the nucleophilic R-amide bond and  $\epsilon$ -amide bond responsible for the sulphonium cleavage reactions, which subsequently affects the nucleophilic reactivity of these groups [466].

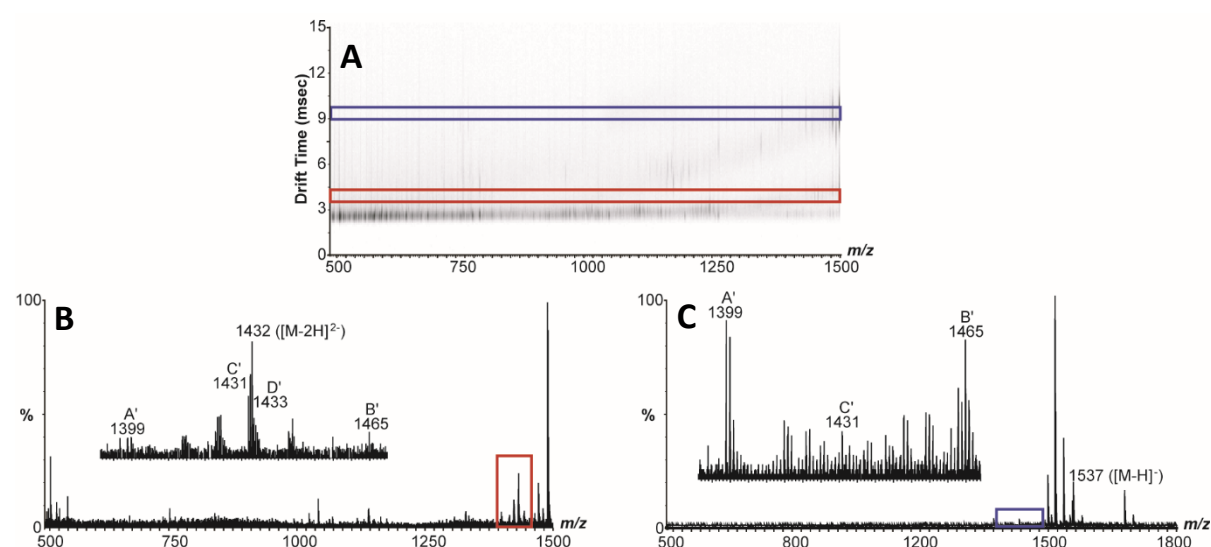
Analysis of Ac-AAKA cross-linked with either DSP or S55'THS demonstrates the fast and efficient covalent bonding of the cross-linking reagents and easy identification of the cross-links through the simplification of the spectra utilising IM. Although 3 products are possible in this CXL reaction (unmodified, dead-end, inter-molecular cross-links) Figures 7.1 and 7.2 reveal a range of species detected. This is likely due to multiply charged product ions, side chain fragmentations and back-bone cleavages obtained with CID. It is noted that the transfer CID energy required for fragmentation of negative ion DSP intermolecular links was higher than that for positive ion

fragmentation of intermolecular 5SS'-THS cross-links. Higher abundances of unwanted fragmentation by-products were also observed in the DSP spectra.

### 7.2.3: Cross-linking of Ubiquitin

To further validate the application of DSP and 5SS'-THS as IM-MS cleavable CXL reagents, the model protein ubiquitin was cross-linked, proteolytically digested, desalted and analysed by IM-MS using the appropriate ion mode. Predicted cross-linked peptides were generated using MeroX software and manually identified through the characteristic expected cross-linker fragments aligned in drift time.

DSP cross-linking of ubiquitin has been well investigated in both negative and positive ion MS [260], [275], [463] with partial sequence information obtained for some peptide fragments [128]. Figure 7.3A shows the negative ion IM-MS spectrum of ubiquitin treated with DSP. Although the data contains a plethora of products, the cross-links are identified by their  $m/z$ , and further confirmed by the characteristic time aligned fragmentations. The selected drift times in Figure 7.3A correspond to the intermolecular cross-link LIFAGK<sub>48</sub>QLEDGR-LIFAGK<sub>48</sub>QLEDGR and the dead-end cross-link LIFAGK<sub>48</sub>QLEDGR-DSP-OH. The difference in mobility of the cross-linked peptides allows the summation of the corresponding drift time range to reveal the time aligned fragmentation to the precursor cross-links.



**Figure 7.3:** (A) IM-MS/MS analysis of DSP cross-linked ubiquitin using a Waters Synapt. (B) Summation of the drift time region 3.7-4.1 msec highlighted in red generates the simplified mass spectrum of the intermolecular cross-link (LIFAGK<sub>48</sub>QLEDGR-LIFAGK<sub>48</sub>QLEDGR). (C) MS summation of the drift time 9.2-9.5 msec highlighted in blue generates the simplified mass spectrum of the dead-end (LIFAGK<sub>48</sub>QLEDGR-DSP-OH) cross-linking products. Here the time aligned fragmentation products (the disulfide fragments labelled A'-D') are indicated accordingly.

The intermolecular cross-linked peptide LIFAGK<sub>48</sub>QLEDGR-LIFAGK<sub>48</sub>QLEDGR is a symmetrical dimer link, with a calculated monoisotopic mass of 2865.5 Da and observed at 1432  $m/z$  ( $[M-2H]^2$ ) as seen in Figure 7.3B. The ATD of the ions is found to be 3.7-4.1 ms and the simplified MS extracted from this drift time range reveals the cross-linked fragments observed at 1399  $m/z$  and 1465  $m/z$  consistent with A' and B' fragmentations of LIFAGK<sub>48</sub>QLEDGR. Unfortunately, the fragments corresponding to C' and B' overlap the doubly charged precursor and are not readily observed. However, the drift time alignment of the A', B' fragments to the precursor ion confidently confirms the cross-link identity.

The 1537  $m/z$  precursor identified as the dead-end modified peptide of LIFAGK<sub>48</sub>QLEDGR has an ATD of 9.2-9.5 ms, as highlighted in Figure 7.3A. The corresponding simplified MS spectrum shows singly charged fragment ions at 1399  $m/z$ , 1431  $m/z$ , and 1465  $m/z$  (A'-C') (Figure 7.3C). These product ions are therefore drift time aligned to the dead-end hydrolysed LIFAGK<sub>48</sub>QLEDGR precursor ( $[M-H]^1$ ). Dead-end cross-linking products are generally readily identifiable due to the small mass losses resulting from cleavages of the labile disulphide bond. The fragmented by-products of 106 Da and 138 Da (Scheme 7.1) are also observed in the simplified MS (data not shown), further verifying the identity of the dead-end cross-link. The low abundance observed of the A'-C' fragmented ions is due to the minimal CID activation energy utilised to reduce competing fragmentation of side chains and backbone. Unfortunately in this case, several unwanted peaks corresponding to unmodified partially digested peptides overlap the drift time of intermolecular and dead end LIFAGK<sub>48</sub>QLEDGR cross-links and can also be observed in both spectra.

An intramolecular cross-linked peptide of AK<sub>29</sub>IQDK<sub>33</sub>EGIPPDQQR was also observed in the cross-linking experiment however, fragmentation of this peptide produces a complex MS/MS spectrum. Cleavage of just the cross-linker is unobservable as this pathway produces no change in mass and instead minimal energy dissociation induces backbone and sidechain fragmentation. CID activation shows a loss of water at  $m/z$  1349 and side chain cleavages occur at Asp<sub>12</sub> and Asp<sub>5</sub>, which are consistent with previous negative ion peptide CID studies (data not shown) [128], [463]. However, as cleavage of the cross-linker moiety is not observed, these modifications cannot be confirmed by our method. Nevertheless, limited information is offered by this type of bio-conjugation, as it provides distance constraints between residues relatively close in sequence and therefore can only define local secondary and tertiary structures.

The observed products and the corresponding fragment ions of the cross-linked products of ubiquitin using DSP are summarised in Table 7.1. In total, 1 intermolecular cross-link, 1 intramolecular cross-link, and 3 dead-end modifications are confidently identified.

**Table 7.1: Summary of observed cross-linking products for DSP treated ubiquitin with observed fragmentations reported.**

Peptide	Monoisotopic Mass (Da)	Observed Ions	Drift Time (ms)	Cα distance (Å)
<b>Intermolecular Cross-Links</b>				
LIFAGK <sub>48</sub> QLEDGR-LIFAGK <sub>48</sub> QLEDGR	2865.5	2 <sup>-</sup>	4.0	NA
Observed fragments:				
LIFAGK <sub>48</sub> QLEDGR (A')	1399.8	1 <sup>-</sup>		
LIFAGK <sub>48</sub> QLEDGR (B')	1465.9	1 <sup>-</sup>		
LIFAGK <sub>48</sub> QLEDGR (C')	1431.7	2 <sup>-</sup>		
LIFAGK <sub>48</sub> QLEDGR (D')	1433.7			
<b>Intramolecular Cross-Links</b>				
AK <sub>29</sub> IQDK <sub>33</sub> EGIPPDQQR	1895.9	1 <sup>-</sup> , 2 <sup>-</sup>	12.2, 2.8	
<b>Dead End Modifications</b>				
IQDK <sub>33</sub> EGIPPDQQR-DSP-THS-OH	1715.8	1 <sup>-</sup> , 2 <sup>-</sup>	10.7, 2.8	
Observed fragments:				
IQDK <sub>33</sub> EGIPPDQQR (C')	1608.7	1 <sup>-</sup> , 2 <sup>-</sup>		
IQDK <sub>33</sub> EGIPPDQQR (D')	1610.8	1 <sup>-</sup> , 2 <sup>-</sup>		
LIFAGK <sub>48</sub> QLEDGR-DSP-THS-OH	1538.7	1 <sup>-</sup> , 2 <sup>-</sup>	9.3, 2.7	
Observed fragments:				
LIFAGK <sub>48</sub> QLEDGR (A')	1399.8	1 <sup>-</sup>		
LIFAGK <sub>48</sub> QLEDGR (C'-H <sub>2</sub> O)	1414.7	1 <sup>-</sup>		
TLSDYNIQK <sub>63</sub> ESTLHLVLR-DSP-THS-OH	2322.1	-	-	
Observed fragments:				
TLSDYNIQK <sub>63</sub> ESTLHLVLR (C')	2215.1			
TLSDYNIQK <sub>63</sub> ESTLHLVLR (D')	2217.1			

\*For fragment ion annotations refer to Scheme 7.1.

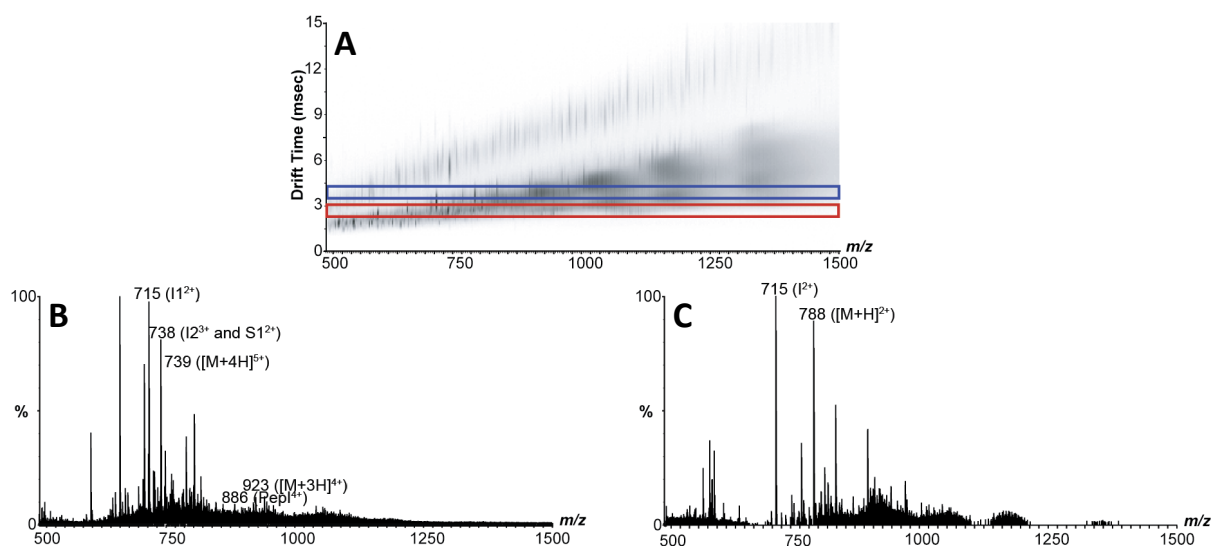
Disulphide cross-linking reagents that are analysed in negative ion mode MS are rarely utilised to determine low resolution 3D structures of proteins due to low MS sensitivity and competing low energy backbone and side chain dissociation. Although this is a unique feature of negative ion MS, such low energy backbone and sidechain cleavages may compete with cross-linker fragmentation and therefore potentially hamper the identification of the cross-linked peptide. It has been reported that the Asp side chain effectively loses water in negative ion mode MS and is generally the most abundant peak observed in negative ion spectra of peptides [128], [246], [472]. In addition to this water loss, negative ion sidechain fragmentations have also been reported for Glu, Gln and Asn (Asn and Gln lose NH<sub>3</sub> from the side-chain) [246]. The side chain fragmentation of Ser has also been reported to be more facile than backbone cleavage [473]–[475]. It is due to these previous reports of sidechain and backbone fragmentations dominating negative ion spectra that analysis of ubiquitin cross-linked with DSP were not conducted at higher CID energies. Other disadvantages of the

negative ion approach include loss in MS sensitivity and the sequencing of peptides is substantially less routine in this mode. It should also be noted that in a previous study where the negative ion cleavages of disulphide containing proteins were investigated, it was shown that all four possible product ions were not evident in all cases [463]. Although the application of negative ion based proteomics approaches provides a complementary method to the conventional positive ion studies of proteins and peptides, the disadvantages mentioned here result in difficult identification of cross-links and have led us to focus on positive ion MS for the investigation of the tertiary structures of *h*PFD subunit 5 and 6 (Section 6.2.5).

Although several studies have investigated the low resolution structure of ubiquitin using other cross-linking reagents in positive ion mode MS [275], the cross-linking of ubiquitin to a positive sulphonium ion-containing facile cleavable cross-linking reagent is novel. Figure 7.4A shows the positive ion IM-MS/MS spectrum of ubiquitin treated with S55'THS. Similar to the data for ubiquitin treated with DSP, the reaction mixture contains a range of products. Possible cross-links generated by MeroX were matched to  $m/z$  features in the spectra and the identity of the cross-linked peptides was further confirmed through the observation of the characteristic time aligned fragmentations (Scheme 7.2). The simplified MS spectrum of the intermolecular linked product and dead-end modified product containing the same tryptic peptide sequence (LIFAGK<sub>48</sub>QLEDGR) is selected as an example and shown in Figure 7.4. The selected drift time at 2.6-3.0 ms reveals the intermolecular cross-link of LIFAGK<sub>48</sub>QLEDGR-TLSDYNIQK<sub>63</sub>ESTLHLVLR observed at 923  $m/z$  ( $[M+3H]^{4+}$ ) with a calculated monoisotopic mass of 3687.9. The cross-link is further verified through the observation of the time aligned CID fragmentation products of 738  $m/z$ , corresponding to the iminotetrahydropyran, I fragment of the TLSDYNIQK<sub>63</sub>ESTLHLVLR peptide and 715  $m/z$ , corresponding to the *s*-methylthiopentanoyl group, S fragment of the LIFAGK<sub>48</sub>QLEDGR peptide. This same  $m/z$  ion can also be seen in the MS spectrum of the dead-end modified LIFAGK<sub>48</sub>QLEDGR (calculated monoisotopic mass of 1428.7) in Figure 7.4C. The MS spectrum here is a summation of the ions with a drift time of 3.7-4.2 ms. It should be noted that the mass of the dead-end cross-linked peptide LIFAGK<sub>48</sub>QLEDGRTLSDYNIQK<sub>63</sub>ESTLHLVLR-S55'THS-OH corresponds to the mass of intermolecular cross-link LIFAGK<sub>48</sub>QLEDGR-TLSDYNIQK<sub>63</sub>ESTLHLVLR (3687.9 Da). A fragment corresponding to I fragment of the dead-end modified peptide (PepI) was identified in the simplified spectrum, ultimately revealing the similar drift time of the ions associated with both the intermolecular and the dead-end cross-link.

Figure 7.4C reveals the extracted spectrum for the dead-end modified peptide LIFAGK<sub>48</sub>QLEDGR peptide ( $[M+H]^{2+}$ ) reacting at K<sub>48</sub> with one NHS group of S55'THS while the other has hydrolysed. The

715  $m/z$  fragment ion is drift time aligned with a precursor at 788  $m/z$ . The mass difference of 148 between precursor and fragment ion immediately identifies the peptide as dead-end modified, providing a rapid method for differentiation of cross-linking product type, while the detection of both parent and fragment ions gives confident identification of the modified peptide.



**Figure 7.4:** (A) IM-MS/MS analysis of S55'THS cross-linked ubiquitin using a Waters Synapt. Highlighted are selected drift time regions from which simplified mass spectra can be generated. (B) Summation of the drift time region 2.6-3.0 msec highlighted in red generates the simplified mass spectrum of the intermolecular cross-link LIFAGK<sub>48</sub>QLEDGR-TLSDYNIQK<sub>63</sub>ESTLHLVLR. Ions corresponding to the dead-end LIFAGK<sub>48</sub>QLEDGR-TLSDYNIQK<sub>63</sub>ESTLHLVLR-S55'THS-OH (Pepl) are also found in this spectrum. (C) Summation of the drift time region 3.7-4.2 msec highlighted in blue generates the simplified mass spectrum of dead-end LIFAGK<sub>48</sub>QLEDGR-S55'THS-OH. The cross-linker modified ions are labelled I & S for the iminotetrahydropyran & *s*-methylthiopentanoyl fragments respectively, while the cross-linked peptides are labelled M.

In total for S55'THS cross-linked ubiquitin analysed by IM-MS/MS, an intermolecular peptide cross-link, an intramolecular cross-link and 4 dead-end modifications were identified from the spectra, which are summarised in Table 7.2 along with the observed characteristic time aligned fragments. These results are consistent with the ability of S55'THS to cross-link primary amines on Lys residues at a distance of up to 20 Å between  $\alpha$ -carbons on the protein backbone, taking into account the length and flexibility of the lysine side chain, and with the known sequence and structure of ubiquitin. It is observed that the majority of modifications detected involve K<sub>48</sub>, suggesting this residue is highly accessible to cross-linking reactions.

**Table 7.2: Summary of observed cross-linking products for S55'-THS treated ubiquitin analysed by IM-MS/MS.**

Peptide	Monoisotopic Mass (Da)	Observed Ions	Drift Time (ms)	Cα distance (Å)
<b>Intermolecular Cross-Links</b>				
LIFAGK <sub>48</sub> QLEDGR-TLSDYNIQK <sub>63</sub> ESTLHLVLR	3687.9	4 <sup>+</sup> , 5 <sup>+</sup>	3.6, 2.9	17.9
Observed fragments:				
LIFAGK <sub>48</sub> QLEDGR (I)	1428.7	2 <sup>+</sup>		
LIFAGK <sub>48</sub> QLEDGR (S)	1475.7	2 <sup>+</sup>		
TLSDYNIQK <sub>63</sub> ESTLHLVLR (I)	2212.1	3 <sup>+</sup>		
TLSDYNIQK <sub>63</sub> ESTLHLVLR (S)	2258.1			
<b>Intramolecular Cross-Links</b>				
AK <sub>29</sub> IQDK <sub>33</sub> EGIPPDQQR	1934.9	2 <sup>+</sup> , 3 <sup>+</sup>	4.2, 1.8	
<b>Dead End Modifications</b>				
IQDK <sub>33</sub> EGIPPDQQR-S55'-THS-OH	1753.8	2 <sup>+</sup> , 3 <sup>+</sup>	3.4, 1.5	
Observed fragments:				
IQDK <sub>33</sub> EGIPPDQQR (I)	1605.8	2 <sup>+</sup> , 3 <sup>+</sup>		
LIFAGK <sub>48</sub> QLEDGR-S55'-THS-OH	1575.7	2 <sup>+</sup> , 3 <sup>+</sup>	3.6, 1.9	
Observed fragments:				
LIFAGK <sub>48</sub> QLEDGR (I)	1428.7	2 <sup>+</sup>		
TLSDYNIQK <sub>63</sub> ESTLHLVLR-S55'-THS-OH	2360.1	3 <sup>+</sup> , 4 <sup>+</sup>	2.7, 2.0	
Observed fragments:				
TLSDYNIQK <sub>63</sub> ESTLHLVLR (I)	2212.1	3 <sup>+</sup> , 4 <sup>+</sup>		
LIFAGK <sub>48</sub> QLEDGRTLSDYNIQK <sub>63</sub> ESTLHLVLR-S55'-THS-OH	3687.9	5 <sup>+</sup>	2.9	
Observed fragments:				
LIFAGK <sub>48</sub> QLEDGRTLSDYNIQK <sub>63</sub> ESTLHLVLR (I)	3539.8	4 <sup>+</sup>		

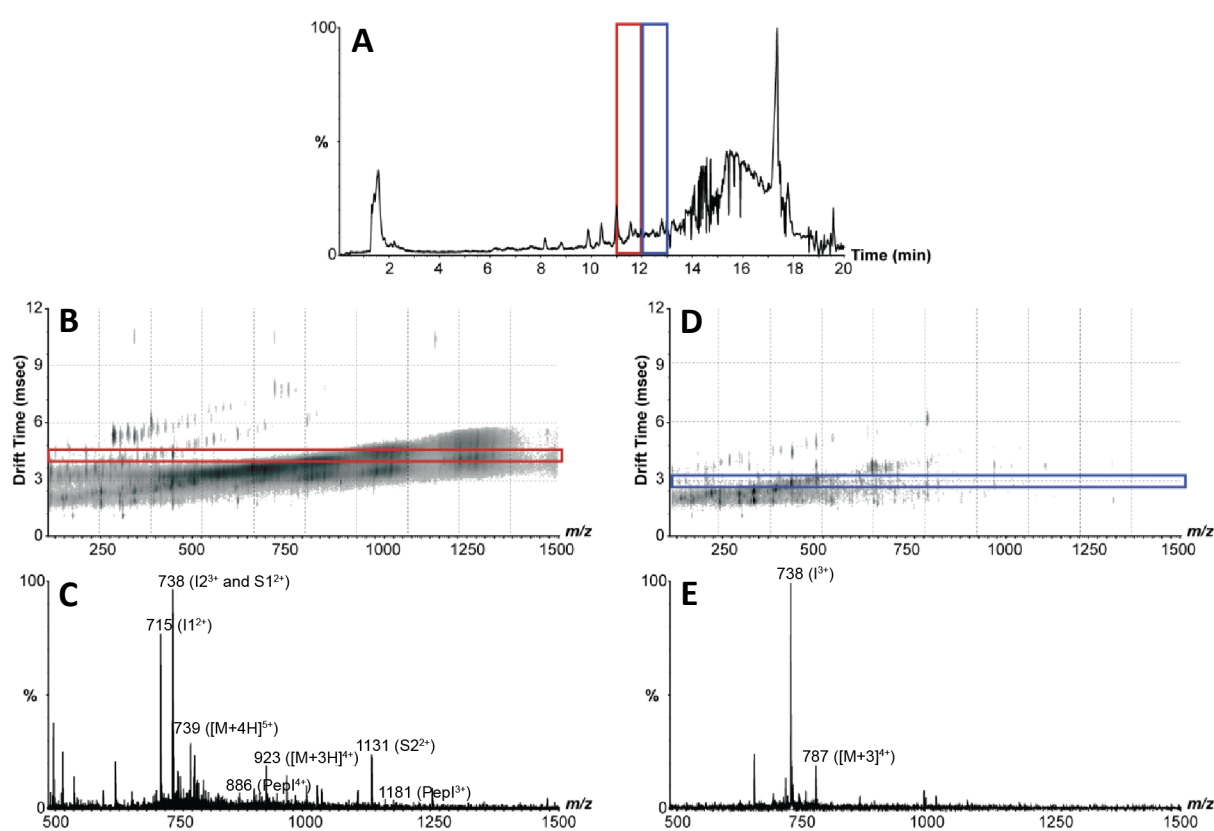
\*For fragment ion annotations refer to Scheme 7.2.

It should be noted there are signals observed in the MS spectrum corresponding to undigested, unmodified protein. These peaks do not obstruct the identification of cross-linked peptides, however they contribute to unwanted background ions and as a consequence, cross-linked peptides with low abundance may be overlooked. This can also mean that there are less modified peptides available for detection. The addition of LC separation was then added to the analysis workflow for S55'THS cross-linked ubiquitin to separate and further minimize the unwanted CXL products observed in the simplified MS spectra.

The products of ubiquitin crosslinked with S55'THS were analysed by LC-IM-MS and were thus orthogonally separated over 2 steps before MS detection. For comparison to the IM-MS analysis conducted previously of ubiquitin cross-linked with S55'THS, the same intermolecular cross-linked product (LIFAGK<sub>48</sub>QLEDGR-TLSDYNIQK<sub>63</sub>ESTLHLVLR) and the dead-end modified product (TLSDYNIQK<sub>63</sub>ESTLHLVLR-S55'THS-OH) are selected. The cross-linked peptides elute in different regions of the LC chromatogram, 10-11 min and 11-12 min bins respectively (Figure 7.5B&D) however, as the parameters for IM are consistent with the previous analysis, the same drift time is observed and the exported MS of the ions at 2.6-3.0 ms and 3.7-4.2 ms reveal the corresponding cross-linked peptides.



The ions previously observed in Figure 7.4B are seen here in Figure 7.5C, however an additional peak is detected at 1131  $m/z$  ( $[M+2H]^{2+}$ ) that corresponds to the TLSDYNIQK<sub>63</sub>ESTLHLVLR-S fragment with a calculated monoisotopic mass of 2259.1. Also, a comparison of Figure 7.4C to Figure 7.5E reveals the decrease in background ions, and the fragment ion of the TLSDYNIQK<sub>63</sub>ESTLHLVLR-I, is clearly observed. This ultimately demonstrates that the separation of the CXL products by both IM and LC increases MS sensitivity and significantly reduces the background ions to allow previously undetected ions to be observed, subsequently allowing confident identification of fragment ions, and verifying the effectiveness of this CXL analysis approach.



**Figure 7.5:** LC-IM-MS/MS analysis of S55'THS cross-linked ubiquitin using Waters Acquity System coupled to a Waters Synapt. Highlighted are selected regions of the (A) LC chromatogram which give rise to (B & D) drift time plots from which simplified mass spectra can be generated. Selection of drift time 3.5-3.7 msec reveals (C) intermolecular cross-link LIFAGK<sub>48</sub>QLEDGR-TLSDYNIQK<sub>63</sub>ESTLHLVLR. Ions corresponding to the dead-end LIFAGK<sub>48</sub>QLEDGRTLSDYNIQK<sub>63</sub>ESTLHLVLR-S55'THS-OH (Pepl) are also found in this spectrum. Selection of drift time 2.0-3.0 msec reveals (E) dead-end TLSDYNIQK<sub>63</sub>ESTLHLVLR-S55'THS-OH cross-link. The cross-linker modified ions are labelled I & S for the iminotetrahydropyran & s-methylthiopentanoyl fragments respectively, while the intermolecular cross-linked peptides are labelled 1 & 2, & precursor ion labelled M.

A close up of the drift plot at signals 739 ( $[M+4H]^{5+}$ ) and 923 ( $[M+3H]^{4+}$ ) reveal that each signal has only one ATD, this subsequently means that the ions corresponding to intermolecular cross-link

LIFAGK<sub>48</sub>QLEDGR-TLSDYNIQK<sub>63</sub>ESTLHLVLR and dead-end modified LIFAGK<sub>48</sub>QLEDGRTLSDYNIQK<sub>63</sub>ESTLHLVLR-S55'-THS-OH are not resolved in this analysis.

In total, for ubiquitin cross-linked with S55'-THS and analysed by LC-IM-MS, 3 intermolecular peptide cross-links, 1 intramolecular cross-link and 4 dead-end modifications were identified from the spectra, which are summarised in Table 7.3 along with the observed characteristic time aligned fragments. The intermolecular cross-linked peptides are consistent with previous CXL-MS studies of ubiquitin [463], [464], [476] and are also consistent with the ability of S55'-THS to cross-link primary amines on Lys residues at a distance of up to 20 Å between Cα on the protein backbone. Since all detected modified cross-linked peptides contain a single Lys, we are able to unambiguously assign the cross-link to residues 48-63.

**Table 7.3: Summary of observed cross-linking products for S55'-THS treated ubiquitin analysed by LC-IM-MS.**

Peptide	Monoisotopic Mass (Da)	Observed Ions	Retention Time (min)	Drift Time (ms)	Cα distance (Å)
<b>Intermolecular Cross-Links</b>					
LIFAGK <sub>48</sub> QLEDGR-TLSDYNIQK <sub>63</sub> ESTLHLVLR	3687.9	4 <sup>+</sup> , 5 <sup>+</sup>	13-14	3.6, 2.9	17.9
Observed fragments:					
LIFAGK <sub>48</sub> QLEDGR (I)	1428.7	2 <sup>+</sup>			
LIFAGK <sub>48</sub> QLEDGR (S)	1475.7	2 <sup>+</sup>			
TLSDYNIQK <sub>63</sub> ESTLHLVLR (I)	2212.1	2 <sup>+</sup> , 3 <sup>+</sup>			
TLSDYNIQK <sub>63</sub> ESTLHLVLR (S)	2258.1	2 <sup>+</sup>			
TITLEVEPSDTIENVK <sub>27</sub> AK- IQDK <sub>63</sub> EGIPPDQQR	3721.9	4 <sup>+</sup> , 5 <sup>+</sup>	14-15	3.5, 2.9	17.5
Observed fragments:					
TITLEVEPSDTIENVK <sub>27</sub> AK (I)	2069.1	3 <sup>+</sup>			
TITLEVEPSDTIENVK <sub>27</sub> AK (S)	2116.1	2 <sup>+</sup>			
IQDK <sub>63</sub> EGIPPDQQR (I)	1605.8	2 <sup>+</sup>			
IQDK <sub>63</sub> EGIPPDQQR (S)	1652.8	2 <sup>+</sup> , 3 <sup>+</sup> , 4 <sup>+</sup>			
TLSDYNIQK <sub>63</sub> ESTLHLVLR-TLTGK <sub>11</sub> TITLEVEPSDTIENVK	4629.5	4 <sup>+</sup> , 5 <sup>+</sup> , 6 <sup>+</sup>	14-15	4.8, 4.1, 3.7	19.2
Observed fragments:					
TLSDYNIQK <sub>63</sub> ESTLHLVLR (I)	2212.1	-			
TLSDYNIQK <sub>63</sub> ESTLHLVLR (S)	2259.1	3 <sup>+</sup> , 4 <sup>+</sup>			
TLTGK <sub>11</sub> TITLEVEPSDTIENVK (I)	2370.2	2 <sup>+</sup>			
TLTGK <sub>11</sub> TITLEVEPSDTIENVK (S)	2417.2	-			
<b>Intramolecular Cross-Links</b>					
AK <sub>29</sub> IQDK <sub>33</sub> EGIPPDQQR	1934.9	2 <sup>+</sup> , 3 <sup>+</sup>	8-9	4.2, 1.8	
<b>Dead End Modifications</b>					
IQDK <sub>63</sub> EGIPPDQQR-S55'-THS-OH	1753.8	2 <sup>+</sup> , 3 <sup>+</sup>	8-9	3.9, 3.5, 2.0	
Observed fragments:					
IQDK <sub>63</sub> EGIPPDQQR (I)	1605.8	2 <sup>+</sup>			
LIFAGK <sub>48</sub> QLEDGR-S55'-THS-OH	1575.7	2 <sup>+</sup> , 3 <sup>+</sup>	10-11	3.2, 2.0	
Observed fragments:					
LIFAGK <sub>48</sub> QLEDGR (I)	1428.7	2 <sup>+</sup> , 3 <sup>+</sup>			
TLSDYNIQK <sub>63</sub> ESTLHLVLR-S55'-THS-OH	2360.1	3 <sup>+</sup> , 4 <sup>+</sup>	11-12	3.0, 2.1	
Observed fragments:					
TLSDYNIQK <sub>63</sub> ESTLHLVLR (I)	2212.1	3 <sup>+</sup> , 4 <sup>+</sup>			
LIFAGK <sub>48</sub> QLEDGRTLSDYNIQK <sub>63</sub> ESTLHLVLR-S55'-THS-OH	3687.8	4 <sup>+</sup> , 5 <sup>+</sup>	13-14	3.6, 2.9	
Observed fragments:					
LIFAGK <sub>48</sub> QLEDGRTLSDYNIQK <sub>63</sub> ESTLHLVLR (I)	3539.8	3 <sup>+</sup> , 4 <sup>+</sup>			

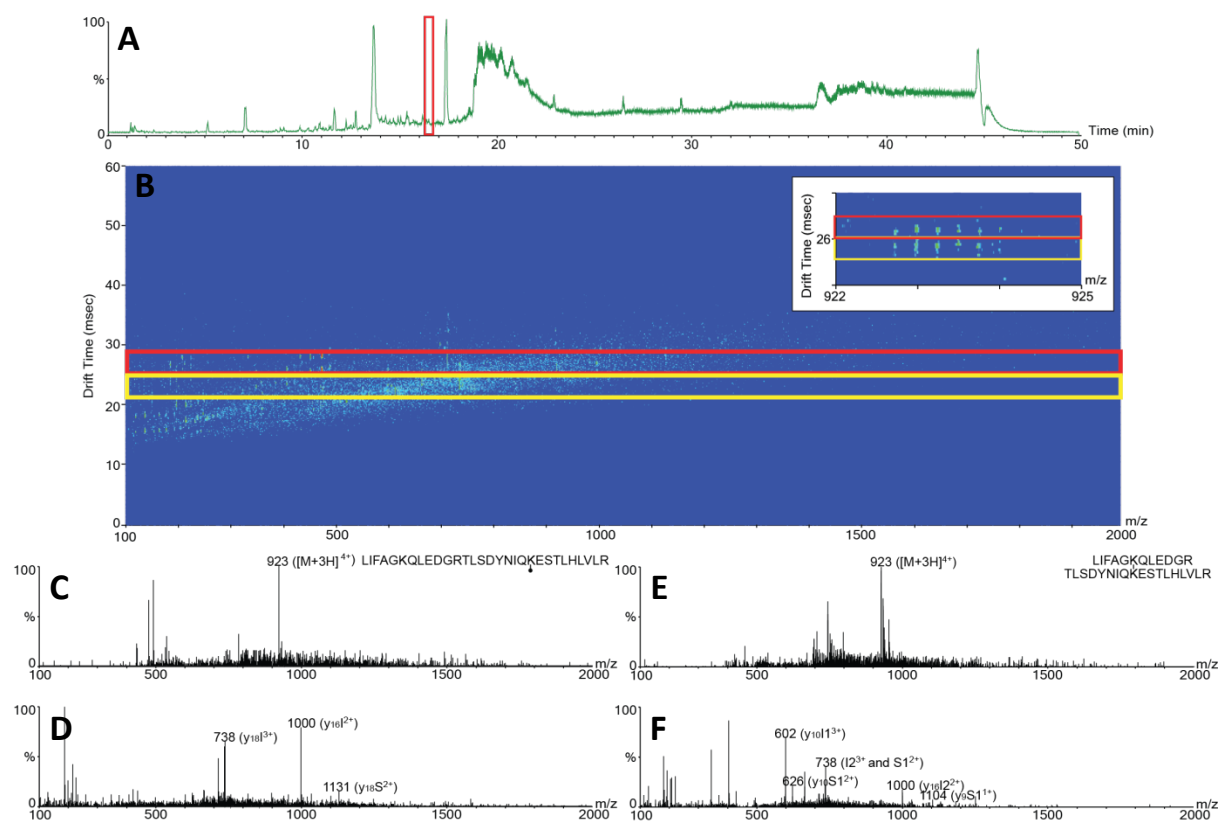
\*For fragment ion annotations refer to Scheme 7.2.

To evaluate the resolving power of the 2 currently commercially available LC-IM-MS systems, the same samples of S55'THS cross-linked ubiquitin were then run on an Agilent Infinity LC coupled to an Agilent 6560 IM-MS system. There were several differences in the experimental setup between the different instruments. For samples analysed on the Waters HDMS Synapt, constant CID voltage of 42 V was used throughout the entirety of the run. Although this was the optimal voltage observed for cross-linked Ac-AAKA, as both product and fragment ions could be seen in the spectrum, it is not necessarily the optimal voltage for larger peptides. It was also decided that a tandem MS approach would enable faster and easier cross-linker identification, especially through the automated software MeroX. Therefore, samples analysed by the Agilent Infinity LC-system coupled to the Agilent 6560 IM-MS were subjected to tandem MS. Unfortunately, the input of data into MeroX is limited to LC-MS/MS files, and IM files were not recognised. As a consequence a LC-MS/MS acquisition was run, with the parameters set to allow the selection of the 10 most abundant ions in each MS scan to be fragmented through a ramped voltage. This enabled a CID voltage applied to each ion so that the cross-linker spacer arm fragmentation could occur as well as and backbone cleavages. As a result, samples were analysed first by positive mode LC-MS/MS with an applied ramp CID voltage. The LC-MS/MS data was exported as a mgf file and processed by MeroX to identify cross-linked peptides, and to determine the optimal CID voltage. Samples were then analysed by LC-IM-MS/MS and the ions identified by MeroX were then manually cross referenced to the LC-IM-MS/MS data.

S55'THS treated ubiquitin was subjected to analysis using the Agilent Infinity LC-system and Agilent LC-IM-MS system. All the same ubiquitin cross-links identified in this analysis were identified in the Waters Acquity coupled to Waters HDMS Synapt analysis. It was previously determined that ions corresponding to intermolecular cross-link LIFAGK<sub>48</sub>QLEDGR-TLSDYNIQK<sub>63</sub>ESTLHLVLR and dead-end modified LIFAGK<sub>48</sub>QLEDGRTLSDYNIQK<sub>63</sub>ESTLHLVLR-S55'THS-OH (739 & 923 *m/z*) overlapped and could not be resolved. During the manual cross reference of the IM-MS data obtained using the Agilent 6560, the ions at 739 ([M+4H]<sup>5+</sup>) and 923 ([M+3H]<sup>4+</sup>) *m/z* were seen to elute from the column at 16.3-16.5 minutes as highlighted in Figure 7.6A. Summation of the ions at this retention time reveals the drift plot shown in Figure 7.6B. Although the ions at 739 *m/z* had one ATD, the ions at 923 *m/z* had 2 resolved ATD distributions, at 26.2 and 25.9 msec (Figure 7.6B inset). Extraction of the ions at these separate drift times reveals different mass spectra seen in Figure 7.6. C, D, E and F. The summation of the ions at the drift time 26.1-26.3 msec (Figure 7.6B red highlight) results in 2 different mass spectra. The first frame (Figure 7.6C) reveals the mass spectrum at low collision

energies and the intact cross-linked peptide 923 ( $[M+3H]^{4+}$ )  $m/z$  can be clearly seen in the figure. The second frame (Figure 7.6D) shows the mass spectrum with applied CID and the corresponding fragment ions to those seen in the first frame. As both spectra were extracted from the same drift time, the fragment ions 738, 1000 and 1131  $m/z$  observed in Figure 7.6D are time aligned to the precursor 923  $m/z$  and corresponds to fragments of the dead-end modified LIFAGK<sub>48</sub>QLEDGRTLSDYNIQK<sub>63</sub>ESTLHLVLR-S55'THS-OH. 738  $m/z$  corresponds to the backbone fragment at  $y_{18}$  in addition to the I spacer arm with a charge of 3+ ( $y_{18}I^{3+}$ ). 1000  $m/z$  corresponds to backbone fragmentation of  $y_{16}$  plus the fragmented spacer arm of I with an overall 2+ charge ( $y_{16}I^{2+}$ ) and 1131  $m/z$  corresponds to backbone  $y_{18}$  including the S cleaved spacer arm with an overall 2+ charge ( $y_{18}S^{2+}$ ). The ions observed here correlate to both backbone cleavage and a fragmented spacer arm, ultimately revealing that there are some backbone bonds that are energetically favoured for fragmentation over others.

In the same manner, extraction of the ions seen at drift time 25.8-26.0 msec (Figure 7.6B yellow highlight) reveal Figure 7.6E at low collisional energy and 7.6F at high collisional energy. From Figure 7.6F, the fragment ions can be correlated to the intermolecular cross-link LIFAGK<sub>48</sub>QLEDGRTLSDYNIQK<sub>63</sub>ESTLHLVLR. Similar to the peaks seen in Figure 7.6D, ions 602  $m/z$  ( $y_{10}I^{3+}$ ), 626  $m/z$  ( $y_{10}S^{3+}$ ), 1000  $m/z$  ( $y_{16}I^{2+}$ ), and 1104  $m/z$  ( $y_9S^{1+}$ ) correspond to the simultaneous cleavage of the backbone and spacer arm. However, a peak arising from spacer arm cleavage and intact LIFAGK<sub>48</sub>QLEDGR (S) and TLSYNIQK<sub>63</sub>ESTLHLVLR (I) is seen at 738  $m/z$  as previously observed in Figure 7.4B and 7.5C.

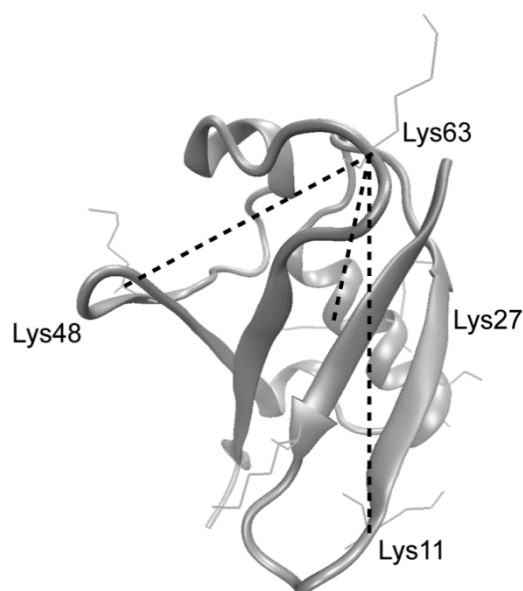


**Figure 7.6:** LC-IM-MS/MS analysis of S55'THS cross-linked ubiquitin using an Agilent Infinity LC-system coupled to an Agilent 6560 IM-MS system. Highlighted are selected regions of the (A) LC chromatogram which give rise to (B) drift time plot from which (C, D, E & F) extracted mass spectra can be generated by selection of drift time bins to identify (C & D) dead-end cross-link LIFAGK<sub>48</sub>QLEDGRTLSDYNIQK<sub>63</sub>ESTLHLVLR-S55'THS-OH & intermolecular cross-link LIFAGK<sub>48</sub>QLEDGRTLSDYNIQK<sub>63</sub>ESTLHLVLR. The cross-linker modified ions are labelled I & S for the iminotetrahydropyran & s-methylthiopentanol fragments respectively, while the intermolecular cross-linked peptides are labelled 1 & 2, & precursor ion labelled M.

The time aligned fragmentations seen here demonstrate the powerful orthogonal separation of IM and reveal that the Agilent 6560 is able to resolve ions at a higher resolution than the Waters Synapt. Since all the ubiquitin detected modified cross-linked peptides contain a single Lys, we are able to unambiguously assign the cross-links. However, in circumstances where large biological complexes have numerous Lys and therefore several possible sites of bio-conjugation, ambiguity in the location of the modified Lys is present. In these cases, the tandem MS system implemented on the Agilent 6560 is likely to induce backbone cleavages similar to those seen in Figure 7.6 and will allow site-specific cross-linking information. The remaining experiments were therefore conducted on the Agilent Infinity LC-system and Agilent 6560 IM-MS instruments.

The intermolecular linked peptides of ubiquitin are displayed on the protein crystal structure in Figure 7.7. This ultimately demonstrates the complementary results of high resolution crystallography and low resolution CXL techniques. It is seen from Table 7.3 that the majority of

modifications detected involve Lys63, and to a lesser extent Lys48 suggesting that these residues are highly accessible to cross-linking reactions.



**Figure 7.7: Identified intra-protein cross-links displayed on an X-ray crystal structure of ubiquitin (PDB 1UBQ) [16]. Only Lys side chains are shown.**

It is apparent from Figure 7.4, 7.5 and 7.6. that, even with orthogonal LC and IM separation, it is possible to observe a number of peptides in the extracted mass spectrum with unresolved elution and drift times from the complex tryptic sample. Nevertheless, the combination of parent ion mass detection with time aligned multiple fragment identification gives high confidence to the identification of cross-linked products from complex samples. In addition, the advantages of this approach, including the rapid time scales of IM separation, the ability to locate the modified Lys and the reduced need for accurate mass measurement with detection of multiple characteristic ions and simple differentiation of inter-, intra- and dead-end linked peptides makes the approach highly effective in protein structure determination.

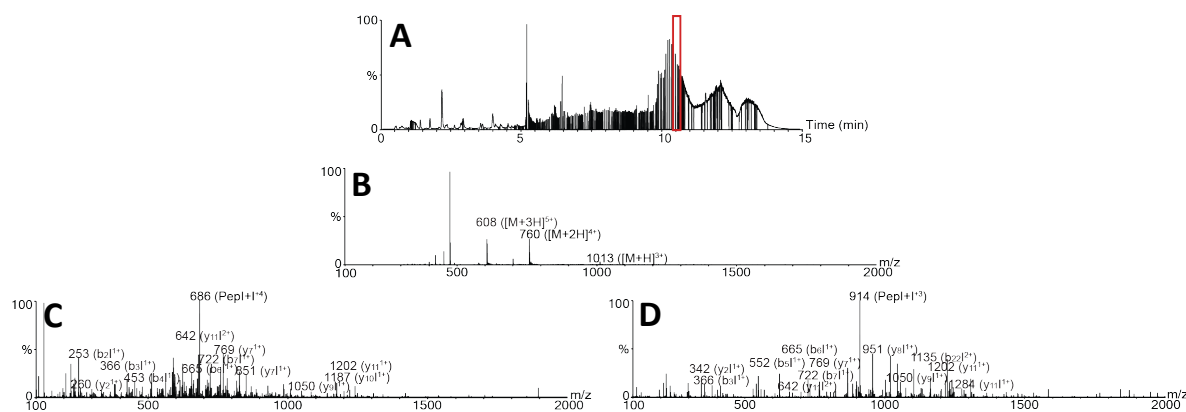
#### **7.2.4: Cross-linking of Caerin 1.10**

In order to further illustrate the resolving power of IM for CXL analysis, the peptide caerin 1.10 was treated with S55'THS. Caerin 1.10 (GLLSVLGSAK<sub>11</sub>HVLPHVVPVIAEK<sub>24</sub>L) is an amphibian peptide known for its antimicrobial activity and it was chosen for CXL analysis simply due to its availability and its 3 sites with the ability to bio-conjugate; at the N-terminus and 2 Lys side chains at amino acid

11 and 24 [477]. S55'THS was added to Caerin 1.10 at a 3:1 ratio of reagent: peptide and incubated at room temperature for 1 hour. The solvent was removed and the samples were reconstituted in water and subjected to analysis.

LC-MS/MS data of cross-linked caerin 1.10 was analysed using MeroX to identify sites of bio-conjugation. Only a single dead-end cross-link ( $561.9\ m/z\ ([M+4H]^{5+})$ ) was identified by MeroX, however, ambiguity in the location of the single dead-end modification is apparent (data not shown). There is difficulty in determining the specific site of the cross-linker modification, due to the fragments corresponding to both models of the doubly modified CXL peptide. It is therefore concluded that LC-MS/MS is unable to resolve isomeric peptides for caerin 1.10 with differences in cross-linking location in this case.

It was also noted that only 1 cross-link was identified by MeroX, despite the ions correlating to higher order cross-links seen in the spectra. The low number of possible cross-links generated by MeroX reveals the software is unable to efficiently identify higher order cross-links, as the algorithm implemented does not include the possibility of multiple cross-links [273]. These were therefore manually calculated for the peptide caerin 1.10 and identified through correlations within the spectra. Figure 7.8A shows the combined selected retention time region of 10.3-10.5 mins that gives an extracted mass spectrum at low collisional energy (Figure 7.8B) to reveal a doubly dead-end cross-link ( $759.9\ m/z\ ([M+2H]^{4+})$  and  $1012.9\ m/z\ ([M+H]^{3+})$ ). The MS/MS fragmentation of  $759.9\ m/z\ ([M+2H]^{4+})$  (Figure 7.8C) and  $1012.9\ m/z\ ([M+H]^{3+})$  (Figure 7.8D) exposes the locations of the cross-linkers at either the N-terminus and K<sub>11</sub> or at the N-terminus and K<sub>24</sub>. In both fragmented frames (Figure 7.8C&D) the highest abundant fragment in the spectra corresponds to the mass of caerin 1.10 with 2 iminotetrahydropyran modifications at  $685.7\ m/z\ (Pepl+I^{+4})$  and  $913.9\ m/z\ (Pepl+I^{+3})$  as indicated. Backbone fragmentation is also observed in these frames and both ions can be sequenced from this data. The spectra reveal cross-linker modifications at the N-terminus and K<sub>11</sub> and at the N-terminus and K<sub>24</sub> as seen in Figure 7.8C&D. It is noted that the abundance of the fragments relating to the cross-linking at the N-terminus and K<sub>11</sub> are higher than N-terminus and K<sub>24</sub> for both spectra. A manual search of the LC-MS/MS data revealed the doubly modified peptides are not resolved by LC.



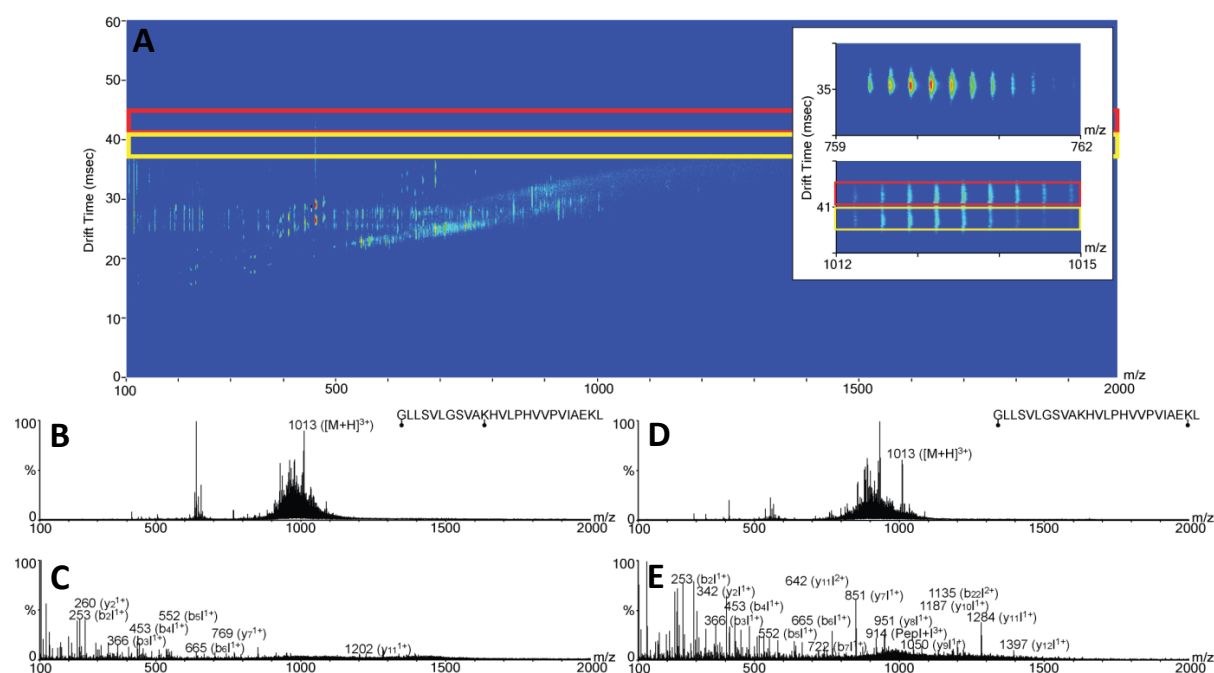
**Figure 7.8: LC-MS/MS analysis of S55'THS cross-linked peptide caerin 1.10 using an Agilent 1290 Infinity LC system coupled to an Agilent 6560 IM-MS system. Highlighted is the selected regions of the (A) LC chromatogram which gives rise to (B) the simplified mass spectra at low collisional energy & MS/MS fragmented product ions of (C) 760  $m/z$  & (D) 1013  $m/z$ .**

To demonstrate the resolving power of IM, cross-linked caerin 1.10 was next analysed by IM-MS/MS. The ions correlating to the doubly dead-end modified peptides (759.5 and 1012.9  $m/z$ ) were searched in the spectrum and were found in to be separated by IM as shown in Figure 7.9A. Figure 7.9A inset reveals the single ATD of 759.5  $m/z$  ( $[M+2H]^{4+}$ ) at 35.5 msec and the 2 separate ATD of 1012.9  $m/z$  ( $[M+H]^{3+}$ ) at 42.2 msec and 39.1 msec. The 2 separate ATD of the 1012.9  $m/z$  ions subsequently exposes the 2 separate structural isomers of the double dead-end cross-link.

It should be noted that the combination of the ions at drift time 35.0-35.3 msec, revealed a simplified mass spectrum with a high abundance of the ion seen at 759.5  $m/z$  ( $[M+2H]^{4+}$ ). However, the time aligned CID spectrum showed ions that correlated to cross-linker modifications at both the N-terminus and K<sub>11</sub> and at the N-terminus and K<sub>24</sub> (data not shown). This is due to the overlapping ATD of the highly charged ions.

In the same manner to cross-linked ubiquitin, MS/MS analysis by Agilent 6560, the selection of the drift times 41.2-43.3 and 39.0-39.2 msec, extracted the simplified mass spectra without CID (Figure 7.9B&D) and with an applied CID voltage (Figure 7.9C&E). The CID frame reveals the time aligned fragments to the isomers of 1012.9  $m/z$ . As simultaneous spacer arm and backbone cleavages are seen in the spectra, sequencing of the cross-linked peptide and the location of the cross-linker modifications can be identified.





**Figure 7.9: (A) IM-MS/MS analysis of S55'THS cross-linked peptide caerin 1.10 using an Agilent 6560 IM-MS system. Inset: Ions 760 m/z & 1013 m/z corresponding to 2x dead-end cross-linked peptides. (B&D) Summation of the respective drift times generates the simplified low collisional energy mass spectra. (C&E) Full sequence MS/MS coverage is extracted for the cross-linked peptide at high collisional energy.**

Figure 7.9C reveals ions relating to either backbone cleavages or simultaneous backbone cleavages and cross-linker spacer arm cleavages. The fragment ions seen in the spectrum and the identified sequences are summarised in Table 7.4, where it can be seen that the majority of ions correspond to backbone b ions attached to an iminotetrahydropyran (I) and are singly charged. This ultimately reveals that the amine at the N-terminus is covalently modified by S55'THS. Table 7.4 also reveals that the y ion cleavages of the peptide do not have bio-conjugations attached. As the y ions observed reach up to  $y_{11}$  without an I or S attachment, it can therefore be deduced that Lys at 24 has not been chemically modified. From these results, although full coverage of the amino acid sequence is not seen for the doubly-modified cross-linked peptide, it can be summarised that the location of the cross-linker is at the N-terminus and  $K_{11}$ .

**Table 7.4: Fragment ions observed in the MS spectrum of caerin 1.10 treated with S55'THS extracted at the drift time 42.2 msec. The modified cross-linkers are located at the N-terminus & at K<sub>11</sub>.**

<i>m/z</i>	Corresponding Peptide
1012.9	([M+H] <sup>3+</sup> )
253.1	I+GL (b <sub>2</sub> l <sup>1+</sup> )
260.2	I+GLL (b <sub>3</sub> l <sup>1+</sup> )
453.2	I+GLLS (b <sub>4</sub> l <sup>1+</sup> )
552.3	I+GLLSV (b <sub>5</sub> l <sup>1+</sup> )
665.4	I+GLLSVL (b <sub>6</sub> l <sup>1+</sup> )
366.2	KL (y <sub>2</sub> <sup>1+</sup> )
769.5	PVIAEKL (y <sub>7</sub> <sup>1+</sup> )
1201.7	PHVVPVIAEKL (y <sub>11</sub> <sup>1+</sup> )

\*For fragment ion annotations refer to Scheme 7.2.

The time aligned fragment ions observed in Figure 7.9E were extracted and summarised in Table 7.5, which illustrates a similar number of backbone b and y ions are present in the spectrum and each backbone cleavage observed has the addition of an iminotetrahydropyran. This subsequently reveals that the identified ions all have simultaneous backbone and cross-linker spacer arm cleavages. The fragment ions 253.1, 453.2, 552.3, 665.4 *m/z* are also seen in Table 7.4 and correspond to singly charged ions with an iminotetrahydropyran attached. It can therefore be deduced in the same manner as previous, that a cross-linking modification occurred at the N-terminus. Table 7.5 reveals 3 more ions correlating to backbone b cleavage with an attached iminotetrahydropyran, and therefore this site is confidently assigned. A doubly charged fragment ion at 1135.2 *m/z* is seen to correspond to the simultaneous backbone cleavage of GLLSVLGSAKHVLPVVPVIA and an iminotetrahydropyran spacer arm fragment (b<sub>22</sub>l<sup>2+</sup>), ultimately revealing that there is no CXL modification on K<sub>11</sub>. The observation of this ion suggests that the second site of CXL bio-conjugation is at K<sub>24</sub>. The remaining ions in Table 7.5 agree with this site of modification. The backbone y ions all have an attached fragmented spacer arm and therefore from the data presented here, the bio-conjugation is confidently assigned to the N-terminus and K<sub>24</sub>.

**Table 7.5: Fragment ions observed in the MS spectrum of caerin 1.10 treated with S55'THS extracted at the drift time 39.1 msec. The modified cross-linkers are located at the N-terminus & at K<sub>24</sub>.**

<i>m/z</i>	Corresponding Peptide
1012.9	([M+H] <sup>3+</sup> )
253.1	I+GL (b <sub>2</sub> l <sup>1+</sup> )
366.2	I+GLL (b <sub>3</sub> l <sup>1+</sup> )
453.2	I+GLLS (b <sub>4</sub> l <sup>1+</sup> )
552.3	I+GLLSV (b <sub>5</sub> l <sup>1+</sup> )
665.4	I+GLLSVL (b <sub>6</sub> l <sup>1+</sup> )
722.4	I+GLLSVLG (b <sub>7</sub> l <sup>1+</sup> )
1135.2	I+GLLSVLGSAKHVLPVVPVIA (b <sub>22</sub> l <sup>2+</sup> )
342.2	I+KL (y <sub>2</sub> l <sup>1+</sup> ),
851.5	I+PVIAEKL (y <sub>7</sub> l <sup>1+</sup> )
950.6	I+VPVIAEKL (y <sub>8</sub> l <sup>1+</sup> ),
1049.6	I+VVPVIAEKL (y <sub>9</sub> l <sup>1+</sup> )
1186.7	I+HVVPVIAEKL (y <sub>10</sub> l <sup>1+</sup> )
1283.7	I+PHVVPVIAEKL (y <sub>11</sub> l <sup>1+</sup> )
642.4	I+PHVVPVIAEKL (y <sub>11</sub> l <sup>2+</sup> )
1396.8	I+LPHVVPVIAEKL (y <sub>12</sub> l <sup>1+</sup> )
913.9	(Pepl+) <sup>3+</sup>

\*For fragment ion annotations refer to Scheme 7.2.

It can therefore be confidently concluded from this IM-MS analysis that the 1012.9 *m/z* ions with the drift time of 42.2 msec give the doubly dead-end cross-linked caerin with the modifications located at the N-terminus and K<sub>11</sub>, while the 1012.9 *m/z* ion with the drift time of 39.1 msec are the doubly dead end cross-linked peptide with the modifications located at the N-terminus and K<sub>24</sub>. A quick comparison of Table 7.4 and Table 7.5 reveal more fragment ions are presented in Table 7.5, therefore the spectrum seen in Figure 7.9E reveals higher sequence coverage of the cross-linked peptide at drift time 39.1 msec. Although the precursor ion for both doubly dead-end cross-linked caerin corresponds to the same *m/z*, it is possible that the isomers respond differently to the same CID voltage applied. It is suggested that the caerin peptide modified at the N-terminus and K<sub>11</sub> is more susceptible to CID than modified caerin at the N-terminus and K<sub>24</sub>, resulting in fragmentations of single residues that ultimately lead to the overall decreased amount of ions relating to backbone b and y ions.

In conclusion, the powerful orthogonal separation of IM is demonstrated here in the separation of these dead-end modified isomers. Although analysis by LC-MS/MS revealed both modified peptides, IM-MS/MS was able to resolve the doubly cross-linked isomers for confident location of the modifications. It is likely that analysis of S55'THS treated caerin 1.10 by LC-IM-MS/MS would further simplify the spectra, however, due to the easy sequencing and identification of the site specific cross-linking of the spectra available, no further analysis on this sample was needed. The combination of IM and MS/MS is therefore especially useful for the analysis of large protein

complexes with abundant Lys available for CXL that can lead to structural isomers, complicating cross-linker identification, and possibly leading to erroneous structural restraints.

### 7.2.5: Cross-linking of *h*PFD Subunits 5 and 6

To determine the possible structure and conformation of the individual *h*PFD subunits 5 and 6, CXL was employed. The synthesised *h*PFD subunits 5 and 6 were independently prepared in solutions of PBS and treated with S55'THS at a 50-fold concentration, which was calculated to be the optimum lysine to reagent ratio of 1:3 [478]. Following overnight incubation at room temperature, the cross-linked samples were subjected to reduction and alkylation before digestion by trypsin.

The samples were analysed by positive mode LC-MS/MS and then by LC-IM-MS/MS. The LC-MS/MS data was exported as a mgf file and processed by MeroX to identify cross-linked peptides. These ions were then manually cross referenced to the LC-IM-MS/MS data. Unfortunately the algorithms and software currently available for identifying cross-links are limited to single bio-conjugations and higher ordered cross-links are not identified. Therefore peptides with more than 1 Lys modification require manual calculation and identification. As this would take a prohibitive amount of time and the determination of singular cross-links is sufficient to provide insight into the 3D structure of proteins [274], [457], identification of higher order cross-links is not included in this structural study of the *h*PFD subunits.

All CXL experiments were conducted a minimum of 3 times with each reaction analysed by LC-MS/MS and LC-IM-MS/MS a minimum of 2 repeats. The CXL products seen in this section were seen at least twice from independent repeats.

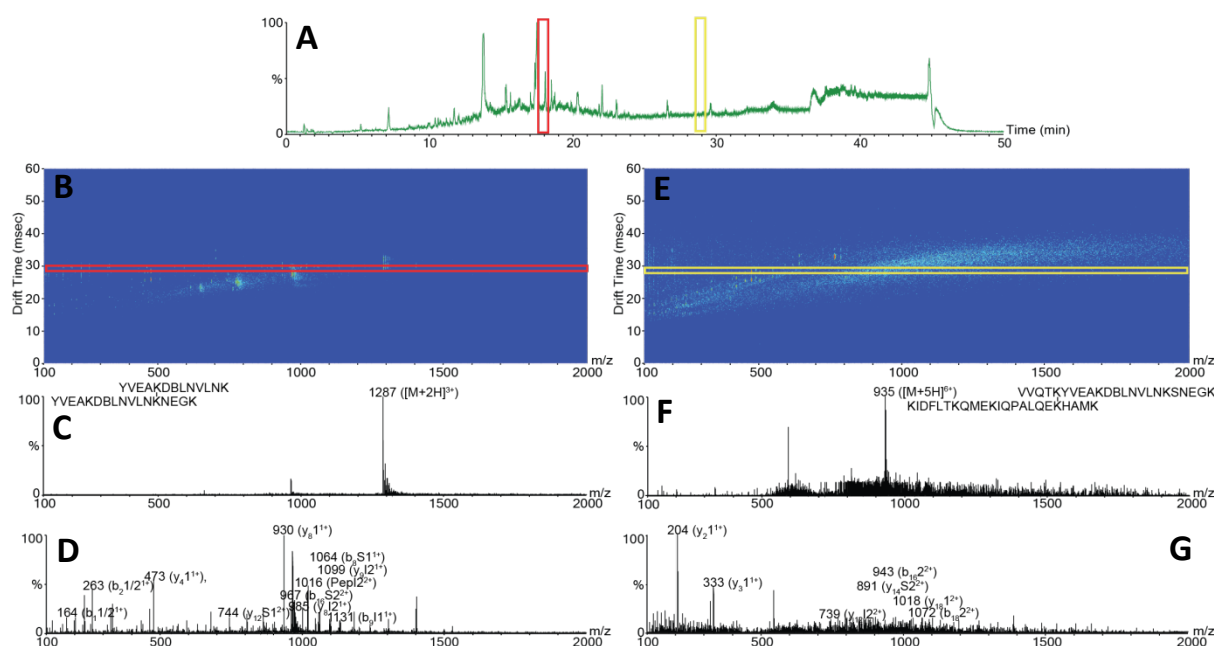
#### 7.2.5.1: *h*PFD Subunit 5 Cross-linked with S55'THS

The LC separation of *h*PFD subunit 5 treated with S55'THS reveals a plethora of products as seen in the chromatogram shown in Figure 7.10A. The selected LC bins of 17.9-18.1 minutes and the selected drift time of 32.3-32.5 milliseconds (Figure 7.10B red highlight) reveals the intermolecular cross-link of YVEAK<sub>47</sub>DBLNVLNK<sub>55</sub>-YVEAK<sub>47</sub>DBLNVLNK<sub>55</sub>NEGK<sub>59</sub> observed at 1286.6  $m/z$  ( $[M+2H]^{3+}$ ) with a calculated monoisotopic mass of 3857.9 (Figure 7.10C). The cross-link is verified through the observation of the time aligned CID fragmentation products seen in Figure 7.10D. Ions corresponding to the simultaneous cleavage of spacer arm and backbone of peptide 1, YVEAK<sub>47</sub>DBLNVLNK<sub>55</sub> are seen at 1131.0, 1064.0 and 744.1  $m/z$ , corresponding to  $b_9$  with the 6-membered iminotetrahydropyran ( $b_9I1^{1+}$ ),  $b_8$  with the *s*-methylthiopentanoyl modification ( $b_8S1^{1+}$ ) and  $y_{12}$  also with the *s*-methylthiopentanoyl ( $y_{12}S1^{2+}$ ), respectively. In conjunction, simultaneous

fragmentation of spacer arm and the peptide YVEAK<sub>47</sub>DBLNVLNK<sub>55</sub>NEGK<sub>59</sub> are seen at 1098.6, 984.5 and 966.7  $m/z$ ; these relate to both  $y_9$  and  $y_8$  with an iminotetrahydropyran modification each ( $y_9I2^{1+}$ ) ( $y_8I2^{1+}$ ) and  $b_{16}$  plus a s-methylthiopentanoyl attachment ( $b_{16}S2^{2+}$ ). In addition, a peak correlating to this intact peptide with the iminotetrahydropyran modification is also seen. Determination of these ions enables the identification of the specific cross-linked Lys residues and all other related fragments observed in the spectrum without the attachment of a spacer arm such as 473.3 ( $y_4I1^{1+}$ ), 930.0 ( $y_8I1^{1+}$ ), 164.1 ( $b_{11}/2^{1+}$ ) and 263.1  $m/z$  ( $b_{21}/2^{1+}$ ) serve as reinforcements to the located bio-conjugation.

To further demonstrate the effectiveness of our approach, Figure 7.10A shows the selected retention time of 29.5-29.7 minutes and the selected drift time of 28.5-28.7 milliseconds, to reveal the intermolecular cross-link of VVQTK<sub>42</sub>YVEAK<sub>47</sub>DBLNVLNK<sub>55</sub>SNEGK<sub>60</sub>-K<sub>106</sub>IDFLTK<sub>112</sub>QMEK<sub>116</sub>IQPALQEK<sub>124</sub>HAMK<sub>128</sub> observed at 934.8  $m/z$  ( $[M+5H]^{6+}$ ) with a calculated monoisotopic mass of 5603.95. The MS/MS fragmentation frame verifies the cross-link through the observation of the time aligned CID fragmentation products of 738.9 and 891.0  $m/z$ , corresponding to the backbone cleavage of K<sub>106</sub>IDFLTK<sub>112</sub>QMEK<sub>116</sub>IQPALQEK<sub>124</sub>HAMK<sub>128</sub> at  $y_{12}$  with an iminotetrahydropyran attached and  $y_{14}$  plus a s-methylthiopentanoyl modification.

Backbone fragmentations without bio-conjugations are also present in the spectrum for VVQTK<sub>42</sub>YVEAK<sub>47</sub>DBLNVLNK<sub>55</sub>SNEGK<sub>60</sub> with  $y$  ions observed at 204.1 ( $y_2I1^{1+}$ ), 333.2 ( $y_3I1^{1+}$ ), and 1018.3 ( $y_{18}I1^{1+}$ )  $m/z$ . The second peptide K<sub>106</sub>IDFLTK<sub>112</sub>QMEK<sub>116</sub>IQPALQEK<sub>124</sub>HAMK<sub>128</sub> is also seen to cleave at  $b_{16}$  and  $b_{18}$  resulting in ions at 943.0 ( $b_{16}I2^{2+}$ ), and 1071.6 ( $b_{18}I2^{2+}$ )  $m/z$ . Although these time aligned fragmentations do not show spacer arm attachments, they are in agreeance with the other related ions and therefore assist with the determination of the site of bio-conjugation.



**Figure 7.10:** LC-IM-MS/MS analysis of S55'THS cross-linked *h*PFD subunit 5 using an Agilent Infinity LC-system coupled to an Agilent 6560 IM-MS system. Highlighted are selected regions of the (A) LC chromatogram which give rise to (B&E) drift time plots from which extracted mass spectra (C, D, F & G) can be generated by selection of drift time bins to identify (C&D) intermolecular cross-link YVEAK<sub>47</sub>DBLNVLNK<sub>55</sub>-YVEAK<sub>47</sub>DBLNVLNK<sub>55</sub>NEGK<sub>59</sub>, & (F&G) VVQTK<sub>42</sub>YVEAK<sub>47</sub>DBLNVLNK<sub>55</sub>SNEGK<sub>60</sub>-K<sub>106</sub>IDFLTK<sub>112</sub>QMEK<sub>116</sub>IQPALQEK<sub>124</sub>HAMK<sub>128</sub>. The cross-linker modified ions are labelled I & S for the iminotetrahydropyran & s-methylthiopentanoyl fragments respectively, while the intermolecular cross-linked peptides are labelled 1 & 2, & precursor ion labelled M.

CID labile CX-MS reagents of the type discussed in this chapter rely on the cross-linker fragmentation to occur at lower energies than any of the backbone cleavages, to allow the peptides involved in any intermolecular cross-link to be separated and sequenced independently. However, it was observed that fragmentation of the precursor ions resulted in many simultaneous cleavages of the spacer arm of S55'THS and backbone. This was also observed in Section 7.2.3 and 7.2.4. It is likely that the calculated collision energy ramp is set at a voltage that enables the fragmentation of both the spacer arm and the peptide. Although this parameter can be adjusted, ions that are higher in charges have been found to fragment more easily than ions with a lower charge.

In total for S55'THS cross-linked *h*PFD subunit 5 analysed by Merox there were 4 intermolecular cross-links found and these are summarised in Table 7.6.

**Table 7.6: Summary of observed cross-linking products of S55'-THS treated hPFD subunit 5 analysed by MeroX & manually identified in the LC-IM-MS/MS analysis.**

Peptide	Monoisotopic Mass (Da)	Observed Ions	Retention Time (min)	m/z	Drift Time (msec)
<b>Intermolecular Cross-Links</b>					
DFFK <sub>104</sub> R-QMEK <sub>116</sub> IQPALQEK <sub>124</sub> <b>Observed fragments:</b> 217.1 (y <sub>2</sub> S1 <sup>2+</sup> ), 364.2 (y <sub>4</sub> S1 <sup>2+</sup> ), 685.4 (y <sub>6</sub> 2 <sup>1+</sup> )	2382.2	2 <sup>+</sup> , 3 <sup>+</sup>	18.9	1191.6, 794.7	33.4, 24.1
VVQTK <sub>42</sub> YVEAK <sub>47</sub> DBLNVLNK <sub>55</sub> SNEGK <sub>60</sub> - K <sub>106</sub> IDFLTK <sub>112</sub> QMEK <sub>116</sub> IQPALQEK <sub>124</sub> HAMK <sub>128</sub> <b>Observed fragments:</b> 333.2 (y <sub>3</sub> 1 <sup>1+</sup> ), 204.1 (y <sub>2</sub> 1 <sup>1+</sup> ), 943.0 (b <sub>16</sub> 2 <sup>2+</sup> ), 891.0 (y <sub>14</sub> S2 <sup>2+</sup> ), 1071.6 (b <sub>18</sub> 2 <sup>2+</sup> ), 738.9 (y <sub>12</sub> l2 <sup>2+</sup> ), 1018.3 (y <sub>18</sub> 1 <sup>2+</sup> )	5603.95	6 <sup>+</sup>	29.6	934.8	28.6
TAEDAK <sub>100</sub> DFFK <sub>104</sub> RK <sub>106</sub> - HAMK <sub>128</sub> QAVMEMMSQK <sub>138</sub> IQQLTALGAAQATAK <sub>153</sub> <b>Observed fragments:</b> 1167.6 (y <sub>21</sub> l2 <sup>2+</sup> ), 1264.6 (y <sub>22</sub> S2 <sup>2+</sup> ), 1283.6 (b <sub>10</sub> S1 <sup>1+</sup> ), 1355.7 (y <sub>22</sub> 2 <sup>3+</sup> ), 1610.5 ([M-OH]+2H) <sup>3+</sup> , 966.4 ([M-H <sub>2</sub> O]+4H) <sup>5+</sup>	4846.4	5 <sup>+</sup>	26.1	970.1	26.1
YVEAK <sub>47</sub> DBLNVLNK <sub>55</sub> -YVEAK <sub>47</sub> DBLNVLNK <sub>55</sub> NEGK <sub>59</sub> <b>Observed fragments:</b> 1131.0 (b <sub>9</sub> l1 <sup>1+</sup> ), 473.3 (y <sub>4</sub> 1 <sup>1+</sup> ), 930.0 (y <sub>8</sub> 1 <sup>1+</sup> ), 1064.0 (b <sub>8</sub> S1 <sup>1+</sup> ), 744.1 (y <sub>12</sub> S1 <sup>2+</sup> ), 1098.6 (y <sub>9</sub> l2 <sup>1+</sup> ), 984.5 (y <sub>8</sub> l2 <sup>1+</sup> ), 1016.3 (PepI2 <sup>2+</sup> ), 966.7 (b <sub>16</sub> S2 <sup>2+</sup> ), 164.1 (b <sub>1</sub> 1/2 <sup>1+</sup> ), 263.1 (b <sub>2</sub> 1/2 <sup>1+</sup> )	3857.9	4 <sup>+</sup> , 3 <sup>+</sup>	18.0	1286.6, 965.2	32.4, 27.6
<b>Dead End Modifications</b>					
TAEDAK <sub>100</sub> DFFK <sub>104</sub> RK <sub>106</sub> -S55'-THS-OH <b>Observed fragments:</b> 173.1 (b <sub>2</sub> l <sup>1+</sup> ), 350.2 (b <sub>6</sub> l <sup>2+</sup> ), 431.3 (y <sub>3</sub> S <sup>1+</sup> ), 618.8 (b <sub>10</sub> l <sup>2+</sup> )	1684.8	2 <sup>+</sup> , 3 <sup>+</sup>	23.7	843.4, 562.6	25.1, 22.6

\*Fragment ion annotations: corresponding b or y ion fragment, the attachment of the l or S modification, the corresponding 1 or 2 peptide and the overall charge of the ion.

In these experiments it is difficult to distinguish between cross-links forming within a protein or between multiple protein molecules, due to the propensity for oligomerisation previously exhibited in other chapters. For simplification, the intermolecular cross-links obtained for hPFD subunit 5 were plotted onto a schematic diagram as intramolecular protein links (Figure 7.11) to demonstrate the potential 3D constructs of the complex and for comparisons to cross-links observed between the chaperone and Aβ(1-40). The CXL data indicates that K<sub>104</sub> is in close proximity to a K<sub>116</sub> and a K<sub>128</sub> and that K<sub>42</sub> is close to a K<sub>124</sub>. As residue K<sub>104</sub> has more than one bio-conjugation identified, this suggests that this residue is highly solvent accessible.

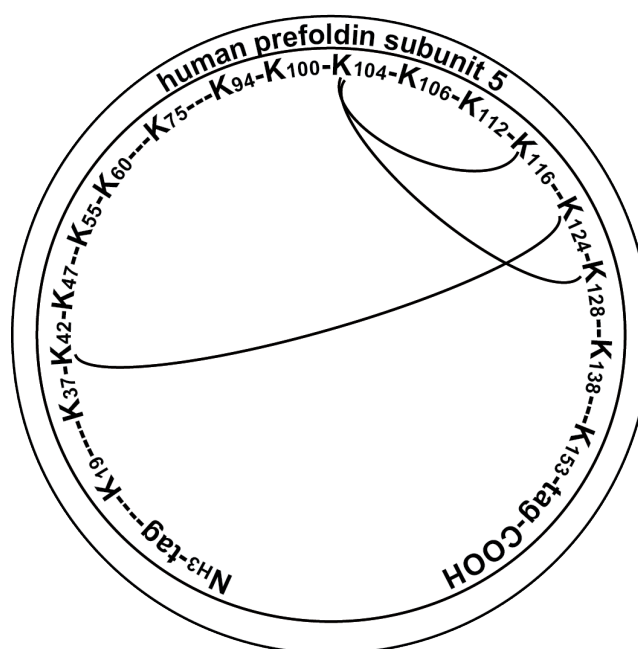


Figure 7.11: Schematic diagram of the identified cross-links for *h*PFD subunit 5. Only Lys residues are shown. Dashes represent amino acids. Solid lines represent the intra peptide cross-links observed.

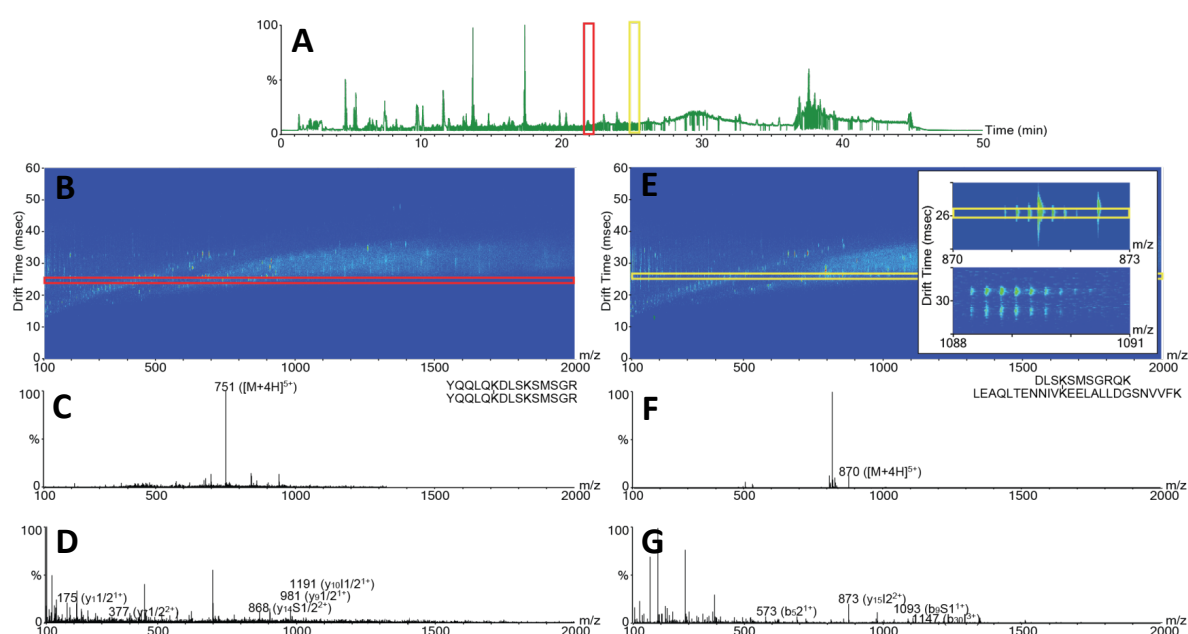
The previously proposed structure of eukaryotic PFD [89], [91], [93], [95], suggests that *h*PFD subunit 5 within the complex has an elongated tertiary structure with a  $\beta$ -sheet central domain ranging from K<sub>55</sub> to T<sub>95</sub> in the sequence to form a tweezer like construction. It is suggested that the residues within this region interact with the other subunits of PFD to hold the complex in its claw like structure, while the N-terminus and C-terminus branch out, both forming hydrophobic  $\alpha$ -helices that are likely to interact together. The data compiled here reinforces the literature derived *h*PFD subunit 5 tweezer like assembly with the turn of the sequence occurring between the residues K<sub>47</sub> and K<sub>94</sub>. It is interesting to note that no bio-conjugations were observed for residues near the termini. This can be explained through the hydrophobic nature of the distal end that causes it to coil and bury into itself, therefore causing the Lys residues to become inaccessible.

#### 7.2.5.2: *h*PFD Subunit 6 Cross-linked with S55'THS

In the same manner as *h*PFD subunit 5, *h*PFD subunit 6 was cross-linked using the positive ion mode cross-linking reagent S55'THS, digested by trypsin and analysed by LC-MS/MS and LC-IM-MS/MS. As an example, summation of the chromatogram between the range 22.3-22.4 minutes (Figure 7.12A) reveals the corresponding drift plot seen in Figure 7.12B and the selected drift time of 24.6-27.9 milliseconds reveals the intermolecular cross-link of YQQLQK<sub>49</sub>DLSK<sub>53</sub>SMSGY-QQQLQK<sub>49</sub>DLSK<sub>53</sub>SMSGY observed at 750.6  $m/z$  ( $[M+4H]^{5+}$ ) with a calculated monoisotopic mass of 3748.9.



The MS/MS fragmentation frame verifies the cross-link through the observation of the time aligned CID fragmentation products of 868.4 and 1190.6  $m/z$  (Figure 7.12D), corresponding to the simultaneous spacer arm and backbone fragmentation:  $y_{14}$  with the s-methylthiopentanoyl modification ( $y_{14}S1/2^{2+}$ ) and  $y_{10}$  with the 6-membered iminotetrahydropyran ( $y_{10}I1^{1+}$ ), respectively. Backbone fragmentations without bio-conjugations are also present in the spectrum at 175.1 ( $y_11^{1+}/y_12^{1+}$ ), 376.7 ( $y_71^{2+}/y_72^{2+}$ ) and 980.5 ( $y_91^{1+}/y_92^{1+}$ )  $m/z$ . As there are no time aligned fragmentations observed correlating to the modification occurring at K<sub>53</sub>, it is proposed that both peptides are cross-linked at K<sub>49</sub>.



**Figure 7.12:** LC-IM-MS/MS analysis of S55'THS cross-linked hPF D subunit 6 using an Agilent Infinity LC-system coupled to an Agilent 6560 IM-MS system. Highlighted are selected regions of the (A) LC chromatogram which give rise to (B&E) drift time plots from which extracted mass spectra (C, D, F & G) can be generated by selection of drift time bins to identify (C & D) intermolecular cross-link YQQLQK<sub>49</sub>DLSK<sub>53</sub>SMSGR-YQQLQK<sub>49</sub>DLSK<sub>53</sub>SMSGR, & (F&G) DLSK<sub>53</sub>SMSGR-QK<sub>60</sub>LEAQLTENNIVK<sub>72</sub>EELALLDGSNVVFK<sub>86</sub> or DLSK<sub>53</sub>SMSGR-QK<sub>60</sub>LEAQLTENNIVK<sub>72</sub>EELALLDGSNVVFK<sub>86</sub> or dead-end DLSK<sub>53</sub>SMSGRQK<sub>60</sub>LEAQLTENNIVK<sub>72</sub>EELALLDGSNVVFK<sub>86</sub>-S55'-THS-OH. The cross-linker modified ions are labelled I & S for the iminotetrahydropyran & s-methylthiopentanoyl fragments respectively, while the intermolecular cross-linked peptides are labelled 1 & 2, & precursor ion labelled M.

Extraction of features over the retention time 25.1-25.3 minutes of the LC chromatogram (Figure 7.12A yellow highlight) reveals a drift plot with ions 870.9  $m/z$  ( $[M+4H]^{5+}$ ) and 1088.3  $m/z$  ( $[M+3H]^{4+}$ ) (Figure 7.12E). The inserts of Figure 7.12E display the ATD of these ions and it can be observed that 870.9  $m/z$  has one ATD whereas 1088.3  $m/z$  is seen to have 2 ATD. Unfortunately, the simplified mass spectra from the drift times correlating to ion 1088.3  $m/z$  revealed that the ions did not fragment at the set CID voltage. However, extraction of the drift time 26.0-26.3 milliseconds shows

the spectra seen in Figure 7.12F-G. It can be seen that the precursor ion 870.9  $m/z$  in Figure 7.17F is low in relative abundance and consequently a high abundance of unwanted overlapping time aligned fragmentations are observed in Figure 7.17G.

The cross-links identified by MeroX and manually confirmed in the LC-IM-MS/MS data are listed in Table 7.7. Figure 7.17G illustrates the fragment ions correlating to the precursor 870.9  $m/z$  are seen at 872.9, 1092.5, 573.3 and 1146.6  $m/z$  and the intermolecular cross-link that corresponds to this is  $\text{DLSK}_{53}\text{SMSGRQK}_{60}\text{-LEAQLTENNIVK}_{72}\text{EELALLDGSNVVFK}_{86}$ . The ions at 872.9  $m/z$  corresponds to fragment  $y_{15}$  of  $\text{LEAQLTENNIVK}_{72}\text{EELALLDGSNVVFK}_{86}$  plus an iminotetrahydropyran modification ( $y_{15}|2^{2+}$ ), while 572.3  $m/z$  relates to  $b_5$  of the same peptide ( $b_5 2^{1+}$ ), and 1092.5  $m/z$  is suggested to correlate to the backbone cleavage  $b_9$  of  $\text{DLSK}_{53}\text{SMSGRQK}_{60}$  plus a s-methylthiopentanoyl ( $b_7 1^{1+}$ ).

Fragment ions 872.9 and 1092.5  $m/z$  are capable of supporting 2 different intermolecular cross-linking scenarios of  $\text{DLSK}_{53}\text{SMSGRQK}_{60}\text{-LEAQLTENNIVK}_{72}\text{EELALLDGSNVVFK}_{86}$  and  $\text{DLSK}_{53}\text{SMSGRQK}_{60}\text{LEAQLTENNIVK}_{72}\text{EELALLDGSNVVFK}_{86}$ . Fortunately, the observation of 573.3 ( $b_5 2^{1+}$ )  $m/z$  allows the identification of the bio-conjugated peptides  $\text{DLSK}_{53}\text{SMSGRQK}_{60}\text{-LEAQLTENNIVK}_{72}\text{EELALLDGSNVVFK}_{86}$ .

It is seen in Figure 7.12E insert that there are 2 ATD observed for 1088.3  $m/z$  which ultimately suggests the presence of 2 cross-linked structural isomers. Table 7.7 shows that the fragment ion 1146.6  $m/z$  seen in Figure 7.17G associates with the backbone cleavage and spacer arm iminotetrahydropyran of dead-end cross-linked peptide  $\text{DLSK}_{53}\text{SMSGRQK}_{60}\text{LEAQLTENNIVK}_{72}\text{EELALLDGSNVVFK}_{86}\text{-S55'-THS-OH}$  ( $b_{30}|3^{3+}$ ). Unfortunately, the precise location of the dead-end cross-link cannot be determined from the acquired data with the possible Lys modification likely to occur at  $K_{53}$ ,  $K_{60}$ , or  $K_{72}$ .

If necessary, further sequencing of the identified cross-links from the LC-IM-MS/MS analysis can be accomplished through the selection of targeted ions and CID parameters can be set specifically for fragmentation of the selected ion. However, as the identification and site specific determination of the intermolecular cross-link was achieved, and the location of the dead-end cross-link of  $\text{DLSK}_{53}\text{SMSGRQK}_{60}\text{LEAQLTENNIVK}_{72}\text{EELALLDGSNVVFK}_{86}\text{-S55'-THS-OH}$  will only give solvation information, further sequencing was not conducted.

**Table 7.7: Summary of observed cross-linking products of S55'-THS treated hPFD subunit 6 analysed by MeroX & manually identified in the LC-IM-MS/MS analysis.**

Peptide	Monoisotopic Mass (Da)	Observed Ions	Retention Time (min)	m/z	Drift Time (msec)
<b>Intermolecular Cross-Links</b>					
DLSK <sub>25</sub> SMSG <sub>RQK</sub> <sub>32</sub> - LEAQLTENNIVK <sub>44</sub> EELALLDGSNVVFK <sub>58</sub>	4350.3	4 <sup>+</sup> , 5 <sup>+</sup>	25.2	1088.3, 870.9	31.1, 28.2, 26.1
<b>Observed fragments:</b> 872.9 (y <sub>15</sub> l <sup>2+</sup> ), 1092.5 (b <sub>9</sub> S1 <sup>1+</sup> ), 573.3 (b <sub>5</sub> 2 <sup>1+</sup> )					
AQAAK <sub>122</sub> AGAPGK <sub>128</sub> -DLSK <sub>25</sub> SMSG <sub>RQK</sub> <sub>32</sub>	2433.3	4 <sup>+</sup>	16.6	609.1	22.7
<b>Observed fragments:</b> 116.0 (b <sub>1</sub> 2 <sup>1+</sup> ), 514.8 (y <sub>10</sub> S1 <sup>2+</sup> ), 229.1 (b <sub>2</sub> 2 <sup>1+</sup> ), 261.5 (y <sub>8</sub> l <sup>3+</sup> ), 397.2 (y <sub>7</sub> 2 <sup>2+</sup> )					
RLDYITAEIK <sub>88</sub> R-AQAAK <sub>150</sub> AGAPGK <sub>128</sub> AAAAFESR	3361.8	3 <sup>+</sup>	36.8	1121.3	31.3
<b>Observed fragments:</b> 217.1 (y <sub>2</sub> S1 <sup>2+</sup> ), 221.1 (b <sub>5</sub> 1 <sup>3+</sup> ), 628.3 (y <sub>4</sub> l <sup>1+</sup> ), 816.9 (y <sub>16</sub> S2 <sup>2+</sup> ), 823.4 (b <sub>10</sub> 2 <sup>1+</sup> ), 962.5 (b <sub>8</sub> 1 <sup>1+</sup> ), 1090.6 (y <sub>10</sub> l <sup>2+</sup> ), 1115.6 precursor(-17) <sup>3+</sup>					
K <sub>8</sub> LQGEVEK <sub>15</sub> - AGAPGK <sub>128</sub> AAAAFESRACSLACGK <sub>146</sub>	3522.7	5 <sup>+</sup>	14.6	705.4	24.2
<b>Observed fragments:</b> 606.8 (b <sub>13</sub> l <sup>2+</sup> ), 924.0 (b <sub>18</sub> S2 <sup>2+</sup> )					
YQQLQK <sub>21</sub> DLSK <sub>53</sub> SMSG <sub>R</sub> -YQQLQK <sub>21</sub> DLSK <sub>53</sub> SMSG <sub>R</sub>	3748.9	5 <sup>+</sup>	22.4	750.6	24.8
<b>Observed fragments:</b> 175.1 (y <sub>1</sub> 1/2 <sup>1+</sup> ), 376.7 (y <sub>7</sub> 1/2 <sup>2+</sup> ), 868.4 (y <sub>14</sub> S1/2 <sup>2+</sup> ), 980.5 (y <sub>9</sub> 1/2 <sup>1+</sup> ), 1190.6 (y <sub>10</sub> l <sup>1+</sup> )					
<b>Dead-end Cross-Links</b>					
DLSK <sub>25</sub> SMSG <sub>RQK</sub> <sub>32</sub> LEAQLTENNIVK <sub>44</sub> EELALLDGSNV VFK <sub>58</sub> -S55'-THS-OH	4350.3	4 <sup>+</sup> , 5 <sup>+</sup>	25.2	1088.3, 870.9	31.1, 28.2, 26.1
<b>Observed fragments:</b> 1146.6 (b <sub>30</sub> l <sup>3+</sup> )					
QELGEARATVGK <sub>78</sub> R-S55'-THS-OH	1644.9	3 <sup>+</sup>	20.4	549.0	21.7
<b>Observed fragments:</b> 181.4 (y <sub>4</sub> l <sup>3+</sup> ), 459.2 (y <sub>7</sub> S <sup>2+</sup> )					

\*Fragment ion annotations: corresponding b or y ion fragment, the attachment of the l or S modification, the corresponding 1 or 2 peptide and the overall charge of the ion.

The identified intra protein cross-links are plotted on a schematic diagram to demonstrate the potential 3D construct of the hPFD subunit 6 in Figure 7.13. K<sub>25</sub> and K<sub>128</sub> are involved in at least 2 bio-conjugations, which indicate that these Lys are easily accessible for solvation and modification. The observed bio-conjugations expose the proximity with which a K<sub>25</sub> and a K<sub>128</sub> are to the opposite end of the primary sequence.

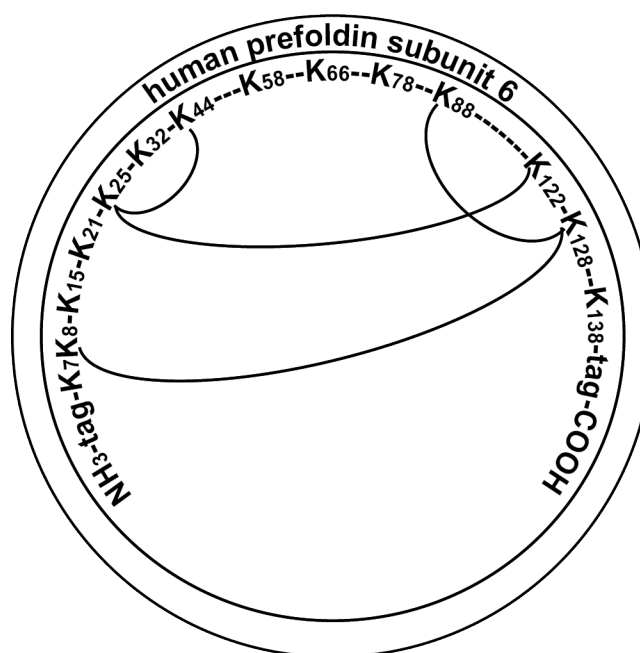


Figure 7.13: Schematic diagram of identified cross-links for *hPFD* subunit 6. Only Lys residues are shown. Dashes represent amino acids. Solid lines represent the intra peptide cross-links observed.

Similar to *hPFD* subunit 5, *hPFD* subunit 6 is proposed to form a tweezer like conformation in the quaternary structure of eukaryotic PFD with a central  $\beta$ -sheet domain located between residues L<sub>49</sub> and E<sub>68</sub> [89], [91], [93], [95]. From the data seen here, it is possible that *hPFD* subunit 6 has formed this elongated tertiary structure, as the intermolecular cross-links formed are seen to occur from residues near the N-terminus to the residues near the C-terminus. Alternatively, this could show interactions from other *hPFD* subunit 6 units. The central  $\beta$ -sheet turn in the tweezer like structure is likely to occur between the residues K<sub>44</sub> and K<sub>88</sub>, and allows the proximity of K<sub>8</sub> to K<sub>128</sub> and K<sub>25</sub> to K<sub>122</sub>. The data seen for CXL *hPFD* subunit 6 could potentially reinforce the literature proposed structure of the eukaryotic PFD subunit.

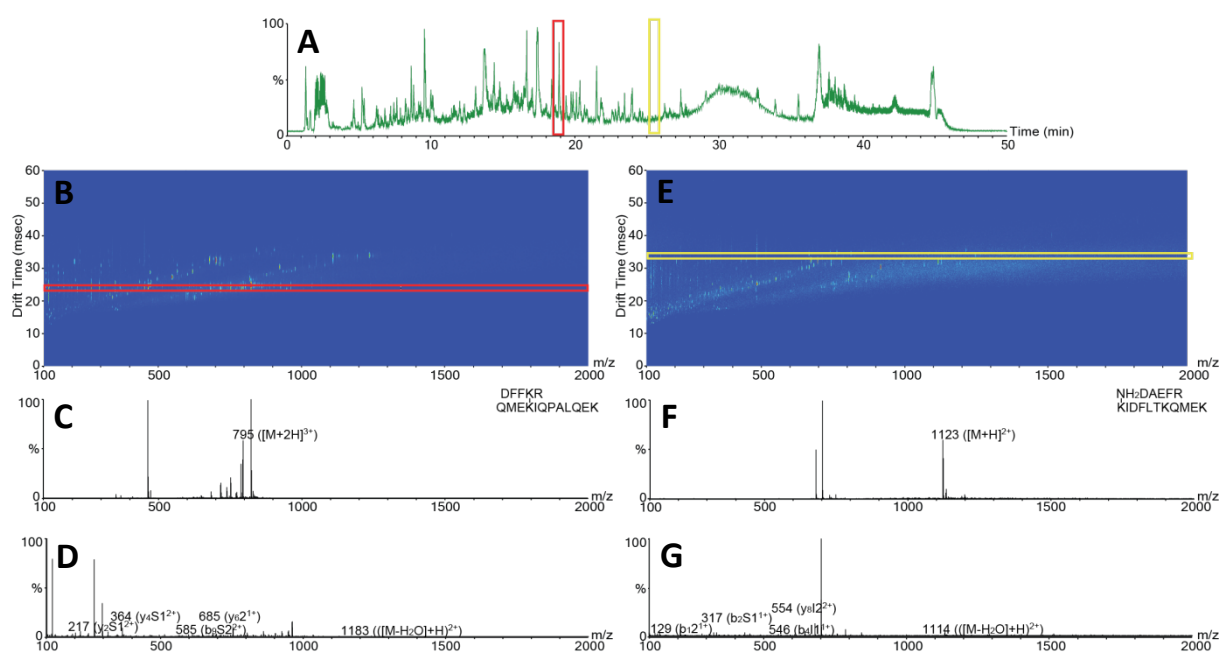
#### 7.2.6: Cross-linking of *hPFD* Subunits 5 and 6 to A $\beta$ (1-40)

To ascertain the binding interactions and interfaces of the expressed *hPFD* subunits to A $\beta$  (1-40), *hPFD* subunit 5 and 6 were each added to a solution of A $\beta$ (1-40) at a 1:1 ratio and incubated at 37 °C overnight to react. Cross-linking reagent was then added to the mixture at a 50:1 ratio of reagent: protein at room temperature and allowed to cross-link overnight. Following the previous workflow, samples were then proteolytically digested and subjected to analysis by LC-MS/MS and LC-IM-MS/MS. Data from LC-MS/MS was exported as an mgf file as input into MeroX. MeroX calculated

and identified cross-links were manually confirmed in the LC-IM-MS/MS spectra. Merox identified cross-links not observed in the LC-IM-MS/MS spectra were removed and omitted from the data.

#### 7.2.6.1: *h*PFD Subunit 5 Cross-linked to A $\beta$ (1-40) with S55'THS

The sample of *h*PFD subunit 5 and A $\beta$ (1-40) treated with S55'THS contains a mixture of products as seen in the chromatogram of Figure 7.14A. As an example of the data recorded, the LC chromatogram region from 18.8-19.0 minutes was selected to extract the drift plot seen in 7.14B. From this plot, the drift time of 24.0-24.3 milliseconds was selected to reveal the simplified mass spectrum (Figure 7.14C) of the intermolecular cross-link of DFFK<sub>104</sub>R-QMEK<sub>116</sub>IQPALQEK<sub>124</sub> observed at 794.7 ([M+2H]<sup>3+</sup>) *m/z* with a calculated monoisotopic mass of 2382.2 Da. This cross-link is verified through the observation of the time aligned CID fragmentation products (Figure 7.14D). Ions related to the  $\gamma_2$  and  $\gamma_4$  fragmentation of DFFK<sub>104</sub>R and the attached modified cross-linker *s*-methylthiopentanoyl are seen at 217.1 ( $\gamma_2$ S1<sup>2+</sup>) and 364.2 ( $\gamma_4$ S1<sup>2+</sup>) *m/z*. A backbone fragmentation without the modified cross-linker is also observed in the spectrum corresponding to the backbone fragmentation of the peptide QMEK<sub>116</sub>IQPALQEK<sub>124</sub> at 685.4 ( $\gamma_6$ 2<sup>1+</sup>) *m/z*. Lastly, a peak related to the loss of water from the precursor is also observed at 1182.6 ([M-H<sub>2</sub>O]+H)<sup>2+</sup>) *m/z*.



**Figure 7.14:** LC-IM-MS/MS analysis of S55'THS cross-linked *h*PFD subunit 5 & A $\beta$ (1-40) using an Agilent Infinity LC system coupled to an Agilent 6560 IM-MS system. Highlighted are selected regions of the (A) LC chromatogram which give rise to (B&E) drift time plots from which extracted (C&F) mass spectra at low collisional energy & (D&G) high collisional energy can be generated by selection of drift time bins to identify intermolecular cross-link (C & D) DFFK<sub>104</sub>R-QMEK<sub>116</sub>IQPALQEK<sub>124</sub> & (F & G) (A $\beta$ -(1-40)) NH<sub>2</sub>DAEFR-K<sub>106</sub>IDFLTK<sub>116</sub>QMEK<sub>124</sub>. The cross-linker modified ions are labelled I & S for the iminotetrahydropyran & s-methylthiopentanoyl fragments respectively, while the intermolecular cross-linked peptides are labelled 1 & 2, & precursor ion labelled M.

In the same manner, the intermolecular cross-link of NH<sub>2</sub>DAEFR-K<sub>106</sub>IDFLTK<sub>112</sub>QMEK<sub>116</sub> between A $\beta$ (1-40) and *h*PFD subunit 5 (1123.1  $m/z$  ([M]+H)<sup>2+</sup>) is shown in Figure 7.14E-G. The selected LC retention time of 25.5-25.8 minutes gives rise to an extracted drift plot (Figure 7.14E) from which the ATD of 1123.1  $m/z$  at 32.0-32.4 milliseconds can be selected. The MS/MS frame reveals the time aligned fragment ions. Fragments observed with the attached cross-linker modification include 554.3  $m/z$  ( $y_8$ I<sup>2+</sup>); FLTK<sub>140</sub>QMEK<sub>144</sub> and the iminotetrahydropyran modification, 546.2 ( $b_4$ I<sup>1+</sup>); DAEF and the iminotetrahydropyran modification, 317.1  $m/z$  ( $b_2$ S<sup>1+</sup>); DA and the s-methylthiopentanoyl modification. A backbone fragment ion is observed at 129.1  $m/z$  ( $b_1$ I<sup>2+</sup>) and the loss of water from the precursor at 1114.1  $m/z$  ([M-H<sub>2</sub>O]+H)<sup>2+</sup> is also seen in the spectrum. These time aligned fragmentations ultimately allow confident identification of the cross-linked Lys amino acids. All cross-links in the acquired spectrum that were recognised by MeroX and manually confirmed in the LC-IM-MS/MS analysis are listed in Table 7.8.

**Table 7.8: Summary of observed cross-linking products of S55'-THS treated *h*PFD subunit 5 & A $\beta$  (1-40) analysed by MeroX & manually identified in the LC-IM-MS/MS analysis.**

Peptide	Monoisotopic Mass (Da)	Observed Ions	Retention Time (min)	m/z	Drift Time (msec)
<b>Intermolecular Cross-Links</b>					
DFFK <sub>104</sub> R-QMEK <sub>116</sub> IQPALQEK <sub>124</sub> <b>Observed fragments:</b> 217.1 (y <sub>2</sub> S1 <sup>2+</sup> ), 364.2 (y <sub>4</sub> S1 <sup>2+</sup> ), 685.4 (y <sub>6</sub> 2 <sup>1+</sup> ), 585.3 (b <sub>9</sub> S2 <sup>2+</sup> ), 1182.6 ([M-H <sub>2</sub> O]+H) <sup>2+</sup> )	2382.2	2 <sup>+</sup> , 3 <sup>+</sup>	18.9	1191.6, 794.7	33.4, 24.1
IDFLTK <sub>112</sub> QMEK <sub>116</sub> - IQQLTALGAAQATAK <sub>153</sub> AAAAFESRACSLEACG TK <sub>171</sub> <b>Observed fragments:</b> 423.8 (y <sub>9</sub> S1 <sup>3+</sup> ), 798.9 (b <sub>15</sub> S2 <sup>2+</sup> ), 989.5 (y <sub>30</sub> l2 <sup>3+</sup> ), 1241.6 (y <sub>23</sub> l2 <sup>2+</sup> ), 805.4 ([M-H <sub>2</sub> O]+5H) <sup>6+</sup> )	4845.4	5 <sup>+</sup> , 6 <sup>+</sup>	26.2	969.9, 808.4	26.7, 25.2
K <sub>106</sub> IDFLTK <sub>112</sub> -DCLNVLNK <sub>55</sub> SNEGK <sub>60</sub> <b>Observed fragments:</b> 699.4 (b <sub>5</sub> l1 <sup>1+</sup> ), 848.4 (b <sub>6</sub> S1 <sup>1+</sup> )	2566.3	3 <sup>+</sup>	19.1	856.1	24.8
(A $\beta$ (1-40)) NH <sub>2</sub> DAEFR- K <sub>106</sub> IDFLTK <sub>112</sub> QMEK <sub>116</sub> <b>Observed fragments:</b> 129.1 (b <sub>1</sub> 2 <sup>1+</sup> ), 546.2 (b <sub>4</sub> l1 <sup>1+</sup> ), 554.3 (y <sub>8</sub> l2 <sup>2+</sup> ), 317.1 (b <sub>2</sub> S1 <sup>1+</sup> ), 1114.1 ([M-H <sub>2</sub> O]+H) <sup>2+</sup> )	2245.1	2 <sup>+</sup>	25.7	1123.1	32.2
(A $\beta$ (1-40)) NH <sub>2</sub> DAEFRHDSGYEVHHQK <sub>16</sub> - QAVMEMMSQK <sub>138</sub> IQQLTALGAAQATAK <sub>153</sub> <b>Observed fragments:</b> 223.6 (b <sub>3</sub> S1 <sup>1+</sup> ), 521.2 (b <sub>8</sub> l1 <sup>2+</sup> )	4846.3	5 <sup>+</sup>	16.4	970.0	27.5
<b>Dead End Modifications</b>					
IQQLTALGAAQATAK <sub>153</sub> AAAAFESRABSLEACG TK <sub>171</sub> -S55'-THS-OH <b>Observed fragments:</b> 147.1 (y <sub>1</sub> <sup>1+</sup> ), 837.3 (PepI <sup>4+</sup> )	3595.8	4 <sup>+</sup>	25.6	899.7	27.6

\*Fragment ion annotations: corresponding b or y ion fragment, the attachment of the l or S modification, the corresponding 1 or 2 peptide and the overall charge of the ion.

To summarise, 3 intra-protein, 2 inter-protein and 1 dead-end cross-links were confidently assigned from the acquired data. The intra-protein bio-conjugations of K<sub>55</sub> to K<sub>106</sub>, K<sub>104</sub> to K<sub>116</sub>, and K<sub>112</sub> to K<sub>153</sub> are drawn on a schematic diagram including the 2 N-terminal inter-protein connections to reveal the low resolution conformation of *h*PFD subunit 5 and the interactions it has to A $\beta$ (1-40) (Figure 7.15). The schematic diagram of Figure 7.15 illustrates the inter-protein cross-links of A $\beta$ (1-40) to the Lys residues near the C-terminus of *h*PFD subunit 5. This suggests that there are surface binding interfaces near the N-terminus of A $\beta$ (1-40) and the *h*PFD subunit 5 residues of K<sub>112</sub> and K<sub>138</sub> during chaperone holdase activity.

It is interesting that the binding interactions are seen to occur at the flexible N-terminus of the A $\beta$  peptide as the cluster of hydrophobic residues are located in the middle and at the C-terminus of the sequence [42]. It was determined in Section 4.2.4 that monomeric A $\beta$ (1-40) adopts a partial  $\beta$ -sheet structure in solution and previously published structural studies indicate an antiparallel  $\beta$ -sheet structure forming between residues L<sub>17</sub> to D<sub>23</sub> and A<sub>30</sub> to V<sub>36</sub>, while the C-terminus is buried to shield its hydrophobic amino acids and the N-terminus has been determined to be flexible and unstructured [373]–[376]. Previous chaperoning studies of the A $\beta$  monomer found stabilisation at

the N-terminus was through electrostatic interactions, whilst stabilisation at the C-terminus was through hydrophobic interactions [479], [480]. This ultimately indicates that the *h*PF<sub>2</sub> subunit 5 stabilises the misfolding substrate at the N-terminus through electrostatic interactions.

In addition, the lack of inter-protein cross-links between the peptide to the Lys residues at the N-terminus of the *h*PF<sub>2</sub> subunit 5, suggests that the misfolding peptide is not encapsulated by chaperone. It is possible that A $\beta$ (1-40) has a transient and possibly surface superficial interaction with *h*PF<sub>2</sub> subunit 5 where holdase activity occurs through the peripheral contact of the C-terminal distal end of the chaperone to the N-terminus of the misfolding peptide via electrostatic interactions.

Although, the chaperoning interactions observed here indicate superficial binding through electrostatic interactions, it must be noted that *h*PF<sub>2</sub> subunit 5 may behave differently within the full complex due to the presence of the other multivalent subunits. It should also be considered that without the other restrictive subunits, flexibility of the *h*PF<sub>2</sub> subunit 5 may be present and the interactions observed here ultimately demonstrate the mechanisms for this particular  $\beta$ -type subunit when incubated individually, that may not occur when the subunit is within the full complex. The other eukaryotic subunits may have a more prominent role in the chaperoning of misfolding substrates. Alternatively, *h*PF<sub>2</sub> subunit 5 may behave as indicated above whilst within the complex regardless.

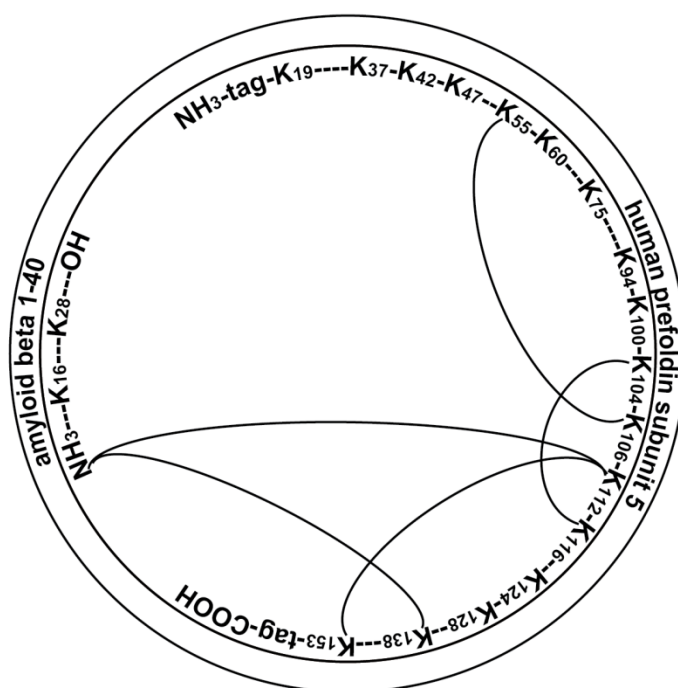


Figure 7.15: Schematic diagram of identified cross-links of *h*PF<sub>2</sub> subunit 5 to A $\beta$ (1-40). Only Lys residues are shown. Dashes represent amino acids. Solid lines represent the intra peptide cross-links observed.



A comparison of the data obtained for the CXL of *h*PF<sub>D</sub> subunit 5 and the CXL of co-incubated *h*PF<sub>D</sub> subunit 5 and A $\beta$ (1-40) reveals the change in complex conformation with the addition of the misfolding peptide. As previously detailed in Section 7.2.5.2, the intra-protein CXL of K<sub>42</sub> to K<sub>124</sub> in *h*PF<sub>D</sub> subunit 5 suggests the proximity of the N-terminal to the C-terminal. In contrast, Figure 7.15 reveals that the addition of A $\beta$ (1-40) to the CXL experiment changes the intra-protein bio-conjugations identified. The data shows that there are no bio-conjugations linking the distal ends of *h*PF<sub>D</sub> subunit 5, subsequently revealing that the coiled tentacles of the chaperone are no longer in proximity with each other. It is proposed that A $\beta$ (1-40) may have disrupted this cross-link and redirected the bio-conjugation elsewhere, or more likely, the *h*PF<sub>D</sub> subunit may have undergone conformational changes to chaperone the misfolding substrate. The modified Lys residues of K<sub>55</sub> to K<sub>106</sub> and K<sub>112</sub> to K<sub>153</sub> were not previously seen in the independently cross-linked *h*PF<sub>D</sub> subunit 5 data. This suggests that a structure change of the subunit occurred to accommodate the misfolding substrate as different Lys residues were exposed for cross-linking modification, and the residues K<sub>55</sub>, K<sub>106</sub>, K<sub>112</sub> and K<sub>153</sub> were previously buried.

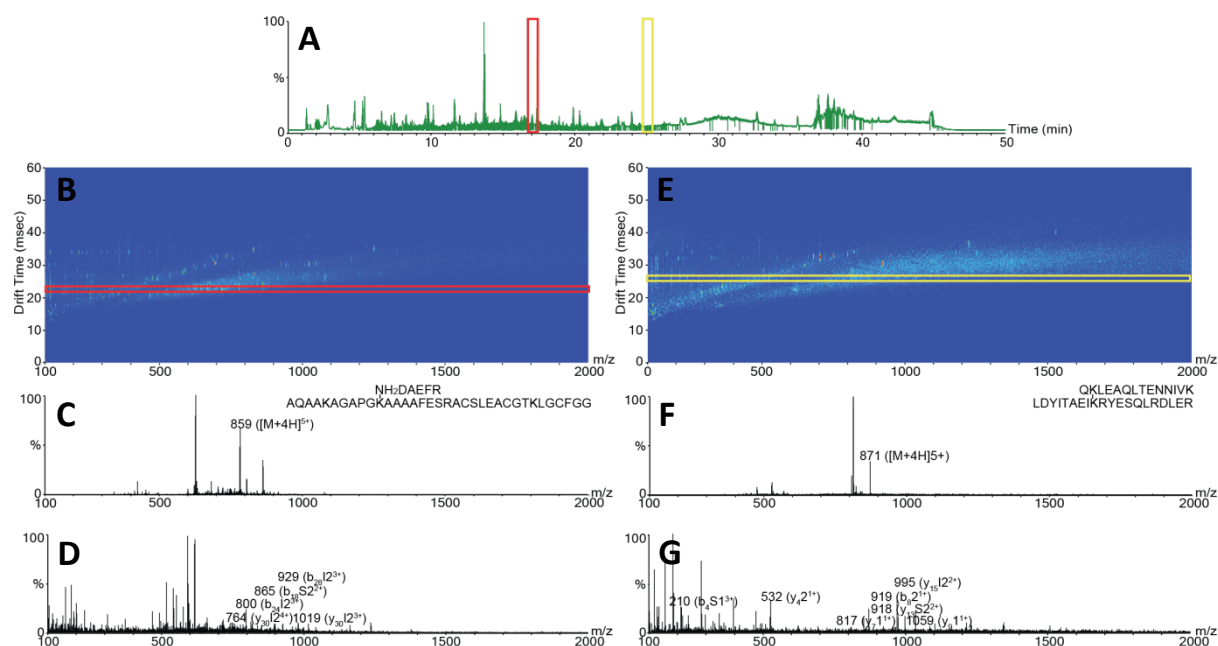
In addition, comparison of the number of modified Lys residues on the *h*PF<sub>D</sub> subunit 5 can reveal the possible restructuring of the protein. The CXL data of the co-incubated A $\beta$ (1-40) and *h*PF<sub>D</sub> subunit 5 reveal 7 different Lys residue modifications on the chaperone compared with 5 for the independent *h*PF<sub>D</sub> subunit. Although the majority of the cross-linked Lys residues are observed to vary between the independent and co-incubation CXL data sets, there is some consistency seen in the data. The bio-conjugation of K<sub>104</sub>-K<sub>116</sub> is observed in both experiments, ultimately suggesting that these residues remain in a relatively solvent attainable position in both experiments. It is proposed that the chaperoning mechanism of the *h*PF<sub>D</sub> subunit 5 involves the disengagement of the C-terminal distal end of the chaperone from the N-terminal distal coil to widen the principal cavity. This conformation exposes residues on the C-terminal branch that assists in the identification and stabilisation of the misfolding A $\beta$ (1-40), leading to holdase activity.

#### 7.2.6.2: *h*PF<sub>D</sub> Subunit 6 Cross-linked to A $\beta$ (1-40) with S55'THS

The complex formed by *h*PF<sub>D</sub> subunit 6 and A $\beta$ (1-40) was cross-linked with S55'THS to reveal several intra-protein cross-links and of inter-protein cross-links between the *h*PF<sub>D</sub> substrate and the disordered peptide. To illustrate this, the full LC chromatogram showing separation of the products of the CXL reaction is seen in Figure 7.16A. Summation of the ions detected between the retention time range of 16.8-17.0 minutes is shown in the drift plot of Figure 7.16B and from this data,

selection of the drift time bins 24.2-24.4 milliseconds gives rise to the mass spectra at low and high collisional energy (Figure 7.16C & D). The spectrum shown without CID exposes a peak at 858.8  $m/z$  with a monoisotopic mass of 4290.0 Da corresponding to the intermolecular cross-link between hPFD subunit 6 and A $\beta$ (1-40) of NH<sub>2</sub>DAEFR-AQAAK<sub>122</sub>AGAPGK<sub>128</sub>AAAAFESRACSLACGK<sub>146</sub>LGCFGG. The subsequent MS/MS frame exposes the time aligned fragments and ions pertaining to the usual backbone and spacer arm fragmentation seen at 764.4 ( $y_{30}I2^{4+}$ ), 799.7 ( $b_{24}I2^{3+}$ ), 864.9 ( $b_{18}S2^{2+}$ ), 929.5 ( $b_{28}I2^{3+}$ ), and 1018.8 ( $y_{30}I2^{3+}$ )  $m/z$ . It is interesting to note that these fragments all correspond to peptide 2; AQAAK<sub>122</sub>AGAPGK<sub>128</sub>AAAAFESRACSLACGK<sub>146</sub>LGCFGG. Although no time aligned fragments are detected for peptide 1, the bio-conjugation is confidently assigned to the N-terminus of DAEFR as there are no other possible intermolecular cross-linking scenarios from the corresponding time aligned fragments of peptide 2.

Figure 7.16 also shows the intermolecular cross-link of QK<sub>32</sub>LEAQLTENNIVK<sub>72</sub>-LDYITAEIK<sub>88</sub>RYESQLRDLER (Figure 7.16A, E-G). The selected bins of the chromatogram from 25.1-25.3 minutes, as highlighted in Figure 7.16A, were extracted to generate the drift plot in Figure 7.16E. From the drift plot, the range 26.1-26.3 milliseconds was selected to extract the mass spectra shown in Figure 7.15F & G. Figure 7.16F reveals the ions at 870.9  $m/z$  as the second most abundant peak of the extracted mass spectrum at low collisional energy. These ions are consistent with the mass of the intermolecular cross-link K<sub>32</sub>LEAQLTENNIVK<sub>44</sub>-LDYITAEIK<sub>88</sub>RYESQLRDLER (calculated monoisotopic mass of 4350.3 Da). This cross-link is confidently assigned with the observation of time aligned fragmentation of peptide backbone and cross-linker reagent spacer arm. These are seen in Figure 7.16E at 210.4 ( $b_4S1^{3+}$ ), 918.5 ( $y_{13}S2^{2+}$ ) and 994.5  $m/z$  ( $y_{15}I2^{2+}$ ). Backbone fragmentations of peptide 1; K<sub>32</sub>LEAQLTENNIVK<sub>44</sub>, are seen at 817.4  $m/z$  ( $y_7I1^{1+}$ ) and 1058.6  $m/z$  ( $y_9I1^{1+}$ ). Backbone cleavages of peptide 2; LDYITAEIK<sub>88</sub>RYESQLRDLER, are displayed in Figure 7.16G at 532.3  $m/z$  ( $y_4I2^{1+}$ ) and 919.5  $m/z$  ( $b_8I2^{1+}$ ).



**Figure 7.16:** LC-IM-MS/MS analysis of S55'THS cross-linked *h*PFD subunit 6 to A $\beta$ (1-40) using an Agilent Infinity LC system coupled to an Agilent 6560 IM-MS system. Highlighted are selected regions of the (A) LC chromatogram which give rise to (B&E) drift time plots from which extracted mass spectra at (C&F) low collisional energy & (D&G) high collisional energy can be generated by selection of drift time bins to identify intermolecular cross-link (C & D) (A $\beta$ -(1-40)) NH<sub>2</sub>DAEFR-AQAAK<sub>122</sub>AGAPGK<sub>128</sub>AAAAFESRACSLACGK<sub>146</sub>LGCFFG & (F & G) QKLEAQLTENNIVK<sub>44</sub>-LDYITAEIK<sub>88</sub>RYESQLRDLER. The cross-linker modified ions are labelled I & S for the iminotetrahydropyran & s-methylthiopentanoyl fragments respectively, while the intermolecular cross-linked peptides are labelled 1 & 2, & precursor ion labelled M.

All cross-links identified by MeroX and manually confirmed in the LC-IM-MS/MS analysis are listed in Table 7.9. It is observed from Table 7.9 that MeroX was able to confidently identify 6 different CXL modifications; 4 intermolecular cross-links and 2 dead-end cross-links. Of the 4 intermolecular cross-links, 2 are inter-protein bio-conjugations between A $\beta$ (1-40) and *h*PFD subunit 6 and 2 are intra-protein within the *h*PFD subunit 6.

**Table 7.9: Summary of observed cross-linking products of S55'-THS treated *h*PFD subunit 6 & Aβ (1-40) analysed by MeroX & manually identified in the LC-IM-MS/MS analysis.**

Peptide	Monoisotopic Mass (Da)	Observed Ions	Retention Time (min)	m/z	Drift Time (msec)
<b>Intermolecular Cross-Links</b>					
DLSK <sub>25</sub> SMSGGR-QK <sub>32</sub> LEAQLTENNIVK <sub>44</sub> EELALLDGSNNVFK <sub>58</sub>	4350.3	5 <sup>+</sup>	25.3	870.9	26.7
<b>Observed fragments:</b> 841.8 (b <sub>21</sub> S2 <sup>3+</sup> ), 872.9 (y <sub>15</sub> I2 <sup>2+</sup> ), 1092.5 (b <sub>9</sub> S1 <sup>1+</sup> ), 573.3 (b <sub>5</sub> 2 <sup>1+</sup> )					
QK <sub>32</sub> LEAQLTENNIVK <sub>44</sub> -LDYITAEIK <sub>88</sub> RYESQLRDLER	4350.3	5 <sup>+</sup>	25.2	870.9	26.2
<b>Observed fragments:</b> 210.4 (b <sub>4</sub> S1 <sup>3+</sup> ), 532.3 (y <sub>4</sub> 2 <sup>1+</sup> ), 817.4 (y <sub>7</sub> 1 <sup>1+</sup> ), 918.5 (y <sub>13</sub> S2 <sup>2+</sup> ), 919.5 (b <sub>8</sub> 2 <sup>1+</sup> ), 994.5 (y <sub>15</sub> I2 <sup>2+</sup> ), 1058.6 (y <sub>9</sub> 1 <sup>1+</sup> )					
(Aβ(1-40)) NH <sub>2</sub> DAEFR-AQAAK <sub>122</sub> AGAPGK <sub>128</sub> AAAFAESRACSLACGK <sub>146</sub> LGCFG	4290.0	5 <sup>+</sup>	16.9	858.8	24.3
<b>Observed fragments:</b> 764.4 (y <sub>30</sub> I2 <sup>4+</sup> ), 799.7 (b <sub>24</sub> I2 <sup>3+</sup> ), 864.9 (b <sub>18</sub> S2 <sup>2+</sup> ), 929.5 (b <sub>28</sub> I2 <sup>3+</sup> ), 1018.8 (y <sub>30</sub> I2 <sup>3+</sup> )					
Aβ(1-40) HDSGYEVHHQK <sub>17</sub> LVFFAEDVGSNNK <sub>28</sub> GAIIGLMVGGVV-K <sub>8</sub> LQGEVEK <sub>15</sub>	4867.5	5 <sup>+</sup> , 6 <sup>+</sup>	26.3	974.3, 812.1	26.1, 25.0
<b>Observed fragments:</b> 454.7 (y <sub>18</sub> S2 <sup>4+</sup> ), 683.6 (y <sub>26</sub> I2 <sup>4+</sup> ), 755.1 (y <sub>22</sub> I2 <sup>3+</sup> )					
<b>Dead End Modifications</b>					
DLSK <sub>25</sub> SMSGRQK <sub>32</sub> LEAQLTENNIVK <sub>44</sub> EELALLDGSNNVFK <sub>5</sub>	4350.3	5 <sup>+</sup>	25.3	870.9	26.7
<b>Observed fragments:</b> 914.7 (y <sub>32</sub> I <sup>4+</sup> ), 1146.6 (b <sub>30</sub> I <sup>3+</sup> )					
QK <sub>32</sub> LEAQLTENNIVK <sub>44</sub> -S55'-THS-OH	1858.0	3 <sup>+</sup>	16.0	620.0	22.7
<b>Observed fragments:</b> 817.4 (y <sub>7</sub> 2 <sup>1+</sup> )					

\*Fragment ion annotations: corresponding b or y ion fragment, the attachment of the I or S modification, the corresponding 1 or 2 peptide and the overall charge of the ion.

The identified cross-links are plotted onto a schematic diagram (Figure 7.17) to illustrate the low resolution conformation of *h*PFD subunit 6 and the interactions it has to Aβ(1-40) (Figure 7.17). The schematic diagram reveals that the identified inter-protein cross-links occur to both the N- and C-terminal distal ends of the *h*PFD subunit 6, bio-conjugated to the misfolding substrate at 2 separate locations. One inter-protein cross-link is observed to occur at the N-terminus of Aβ(1-40) to K<sub>128</sub> of *h*PFD subunit 6, while the other is seen at K<sub>28</sub> of Aβ(1-40) to K<sub>8</sub> of *h*PFD subunit 6. As previously discussed in Section 4.2.4 and 7.2.6.1, the Aβ(1-40) monomer is able to adopt a partial β-sheet fold in an effort to shield its hydrophobic residues. K<sub>28</sub> located on the Aβ peptide has been demonstrated by previous studies to be part of the β-hairpin turn of the partial structure adopted [373]–[376]. This suggests that the amyloidogenic peptide is encapsulated by the *h*PFD subunit 6.

The inter-protein bio-conjugations observed between the chaperone to Aβ(1-40) can also give insight into the structural change of *h*PFD subunit 6 with chaperoning activity. Comparison of the CXL intermolecular links found in Section 7.2.5.2 reveal that the protein residues of K<sub>8</sub> and K<sub>128</sub> were cross-linked. This intra-protein bio-conjugation suggests proximity of the termini without the addition of Aβ(1-40). As this cross-link is no longer observed upon co-incubation, it can be assumed

that the termini of *h*PF<sub>2</sub> subunit 6 are no longer within proximity following the addition of the misfolding peptide. In a similar manner seen for *h*PF<sub>2</sub> subunit 5, the central cavity appears to have widened to accommodate the substrate. However, little change is observed for *h*PF<sub>2</sub> subunit 6 in terms of residue modification. Figure 7.13 reveals that the independently S55' THS treated *h*PF<sub>2</sub> subunit 6 has a total of 6 different Lys residues modified and this is also observed in Figure 7.17. Almost all the same residues have been modified, with the exception of residues K<sub>122</sub> and K<sub>32</sub>. As the residues K<sub>8</sub>, K<sub>25</sub>, K<sub>44</sub>, K<sub>88</sub> and K<sub>128</sub> are solvent accessible in Section 7.2.5.2, the addition of Aβ(1-40) reveals that the solvation of these residues do not change. This ultimately shows that the amphipathic coils of *h*PF<sub>2</sub> subunit 6 undergoes little unfolding and reorientation whilst undertaking holdase chaperone activity of Aβ(1-40).

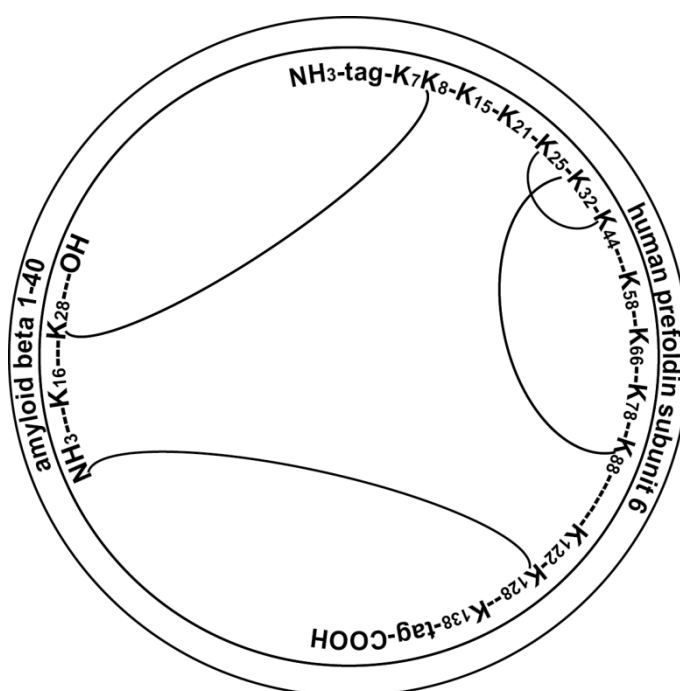


Figure 7.17: Schematic diagram of identified cross-links of *h*PF<sub>2</sub> subunit 6 to Aβ(1-40). Only Lys residues are shown. Dashes represent amino acids. Solid lines represent the intra peptide cross-links observed.

Section 7.2.5.2 identified that the central turn and β-sheet in the tweezer like structure of *h*PF<sub>2</sub> subunit 6 is likely to occur between the residues K<sub>44</sub> and K<sub>88</sub>. The intra-protein cross-links seen here appear consistent with this structural arrangement. Although little restructuring of the *h*PF<sub>2</sub> subunit 6 coils seem to occur during holdase chaperone activity, it cannot be ruled out that the reorientation and exposure of the amphipathic coils occurs before the binding of the substrate as these temporal interactions and flexibility of the protein were not integrated into these experiments.

Comparison of the CXL data for co-incubated *h*PFD subunits 5 and 6 with A $\beta$ (1-40) reveals that both subunits may have increased the central cavity size between their respective amphipathic coils as there are no bio-conjugations linking the C-terminal to the N-terminal during chaperone activity. The distal coils of subunit 6 seem to have less of a structure change than subunit 5. More Lys residues on *h*PFD subunit 5 are modified with the addition of A $\beta$ (1-40) than for *h*PFD subunit 6, which suggests that subunit 5 undergoes a larger change during chaperone activity than subunit 6.

In a previous structural study for prokaryotic PFD  $\alpha$  and  $\beta$  subunits, it was identified that hydrophobic grooves in the distal coils formed the central cavity in both cases [84], [85]. The published molecular dynamic simulations of the archaeal PFD subunits reveal the importance of this cavity for binding to misfolding substrates and chaperoning activity, similar to the results seen here. Previous literature chaperone studies suggest that the prokaryotic PFD  $\alpha$  subunit extends its central cavity through the loosening of the distal coils and multiple conformational changes are observed contingent on the size of the substrate. The results in this Chapter indicate that eukaryotic PFD  $\alpha$  subunit 5 also has the ability to loosen its coils. It was published that the prokaryotic  $\alpha$  subunit provides the multiple substrate recognition mechanism of *ph*PFD, while the  $\beta$  subunit is essential for the binding of non-native proteins. MD simulations showed that the distal regions of the prokaryotic  $\beta$  subunit formed van der Waal contacts with the misfolding substrate [84], [85]. In particular, I<sub>107</sub> of the PFD  $\beta$  subunit near the C-terminal (*ph*PFD has a total of 117 amino acids) has been indicated to be a crucial residue for binding and redirecting of the misfolding substrate. The  $\beta$  subunit first interacts with the surface of the unfolded proteins by the hydrophobic interaction to protect them from aggregation. These published studies seem to align with the CXL results of *h*PFD subunit 6 seen here, as both termini have cross-links to the misfolding substrate. However, the absence of cross-links between the N-terminus to the C-terminus suggests *h*PFD subunit 6 extends its distal ends and is therefore flexible, while the prokaryotic PFD  $\beta$  subunit has been reported to have a rigid structure [84], [85]. Further, CXL assays for the eukaryotic PFD subunits 1, 2, 3 and 4, may reveal further information on the chaperoning mechanism of intact PFD.

It has been suggested in previous studies of the hexameric PFD that the substrate binding mechanism depends on the flexibility of the molecular clamp, similar to the results seen for *h*PFD subunit 5, as a means to grip substrates of varying shapes and sizes [85]. It was thought that eukaryotic PFD may be more limited in substrate binding than prokaryotic PFD due the 6 divergent subunits and potentially specialised subunits, utilising a combination of its coiled-coil distal ends and its spatial cavity to hold misfolded proteins rather than interactions only through the coiled-coil distal ends seen for prokaryotic PFD [85], [95]. It is interesting to see that the archaeal PFD complex

utilises only its tips for holdase interactions, whilst the archaeal PFD  $\alpha$  subunit is seen to have conformational flexibility. It is possible that the confines of the intact prokaryotic PFD constrict movement of the subunits, while eukaryotic PFD has evolved to have flexibility in all the subunits, as seen in the results here, enabling the use of its spatial cavity. These previously published studies also suggested that the prokaryotic PFD subunits interact in a concerted or multivalent fashion to chaperone misfolding substrates. A concerted interaction of the eukaryotic PFD subunit is likely to occur based on the CXL results seen here as *h*PFD subunit 5 is seen to stabilise one end of A $\beta$ (1-40) while *h*PFD subunit 6 is observed to engulf the substrate. It is highly likely that the subunits interact in a multivalent fashion to increase the overall chaperone efficiency, as was observed in Section 3.2.4.1.

To summarise, the novel approach of CXL with CID labile CX-MS reagents analysed by LC-IM-MS/MS was developed to address some of the analytical challenges associated with higher order protein structures investigated by CXL-MS. It is shown that the orthogonal separation power of IM allows for simplification of spectra arising from complex tryptic digest mixtures following cross-linking, in turn simplifying the detection of cross-linking products. In addition, the use of MS cleavable cross-linkers and post IM CID allows for the detection of time aligned precursor and fragment ions, providing high confidence identification of cross-linking modifications with residue specific location, and effectively distinguishing between dead-end, intra- and inter-molecular cross-linked peptides. The data presented herein indicate that LC-IM-MS/MS offers a clear advantage over traditional CXL-MS methodologies for the identification of cross-linked peptides from complex mixtures. Utilising this approach, independent CXL assays of *h*PFD subunit 5 and *h*PFD subunit 6 with S55'THS were conducted and confident identification of the cross-linked peptides contribute to the construction of low-resolution models for these eukaryotic chaperones. Further CXL assays of these subunits with the addition of monomeric A $\beta$ (1-40) revealed inherent flexibility of both *h*PFD subunits, with *h*PFD subunit 5 demonstrating a higher degree of conformational changes.

## 7.3: Experimental

All proteins and chemicals were purchased from Sigma Aldrich (Missouri, USA) unless otherwise specified, and were used as received. DSP was purchased from Thermo Scientific (Rockford, USA) and sequencing grade modified trypsin was purchased from Promega (Madison, USA). Rapigest was purchased from Waters (Manchester, UK). Ac-AAKA was synthesised in-house using standard Fmoc solid phase methods on 2-chlorotriyl chloride resin (GL Biochem, Shanghai, China) and purified by HPLC to greater than 95 % purity.

### 7.3.1: Synthesis of S-methyl 5,5'-thiodipentanoylhydroxysuccinimide

The following procedure was adapted from Lu *et. al.* [466]. Compounds were characterised through NMR with an Agilent 500 spectrometer ( $^1\text{H}$  at 500 MHz,  $^{13}\text{C}$  at 125 MHz) in  $\text{CDCl}_3$  as the solvent unless specified otherwise.  $^1\text{H}$  chemical shifts are reported in ppm relative to TMS ( $\delta$  0.0) and  $^{13}\text{C}$  NMR are reported in ppm relative to TMS ( $\delta$  0.0). All *J* values were quoted to the nearest 0.1 Hz. Multiplicities are reported as (br) broad, (s) singlet, (d) doublet, (t) triplet, (q) quartet, (qnt) quintet, (sext) sextet and (m) multiplet. Mass spectra were obtained on a nanospray Waters Micro mass spectrometer and samples were dissolved in a mixture of 50 % methanol and water unless otherwise stated and parameters were set to capillary voltage, 2.1. kV; cone voltage, 60 eV; trap collision energy, 10 V; transfer collision energy, 10 V; source temperature, 80 °C; backing pressure, 5 mBar.

**5-Mercaptopentanoic acid.** 5-bromovaleric acid (2.2 g, 12.2 mmol) and thiourea (1.4 g, 18.4 mmol) were dissolved in EtOH (30 mL) and refluxed for 24 hours. The solvent was removed by rotatory evaporation and 7.5 M NaOH (30 mL, 225 mmol) was added. The mixture was stirred for an additional 24 hours at 90 °C. This solution was then cooled to 0 °C on ice and whilst cooled 2 M  $\text{H}_2\text{SO}_4$  (5 mL) was added slowly with stirring until the pH was 1 and the product was extracted with DCM (2 x 100 mL). The combined extracts were dried over anhydrous  $\text{Mg}_2\text{SO}_4$  and concentrated by rotatory evaporation to give the 5-mercaptopentanoic as colourless oil in quantitative yield. The product was used without further purification (1.5 g, 91%).  $^1\text{H}$  NMR (500 mHz,  $\text{CDCl}_3$ ):  $\delta$  1.36 (t, 1H, *J*=7.8), 1.65-1.78 (m, 4H), 2.38 (t, 2H, *J*=7.5), 2.55 (q, 2H, *J*=6.9), 9.80 (s, broad, 1H).  $^{13}\text{C}$  NMR (126 MHz,  $\text{CDCl}_3$ ):  $\delta$  25.97, 26.80, 31.13, 36.08, 182.30. MS ( $\text{C}_5\text{H}_{10}\text{O}_2\text{S}$ , nanoESI): calculated  $[\text{M}+\text{H}]^+$  135.20, found 134.92.

**5,5'-Thiodipentanoic acid.** 5-bromovaleric acid (2.4 g, 13.38 mmol) in 16.7 M NaOH (8 mL) was added dropwise to a cooled solution (0 °C) of freshly prepared 5-mercaptopentanoic (1.5 g, 11.2 mmol) dissolved in 16.7 M NaOH (8 mL). The resulting reaction was stirred for 40-50 °C for 24 hours.



The solution was then cooled to room temperature and acidified with conc. HCl to pH 1 and repeatedly extracted with DCM (5 x 50 mL). The extracts were combined and dried over Mg<sub>2</sub>SO<sub>4</sub> and concentrated by rotatory evaporation to give 5,5'-thiodipentanoic acid as a white solid. The product was used without further purification (2.4 g, 92%). <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>): δ 1.63-1.79 (m, 8H), 2.40 (t, 4H, J=7.2), 2.54 (t, 4H, J=7.2). <sup>13</sup>C NMR (75 MHz, CDCl<sub>3</sub>): δ 26.45, 31.53, 34.26, 36.18 and 181.95. MS (C<sub>10</sub>H<sub>18</sub>O<sub>4</sub>S, nanoESI): calculated [M+H]<sup>+</sup> 235.31, found 235.09.

**5,5'-Thiodipentanoylhydroxysuccinimide.** 5,5'-Thiodipentanoic acid (150 mg, 0.64 mmol) and NHS (184.3 mg, 16.02 mmol) were dissolved in dry THF (20 mL) and stirred for 5 mins at room temperature under nitrogen, in darkness. EDC (307.4 mg, 16.02 mmol) was added to the reaction which was stirred overnight. The solvent was removed by rotatory evaporation and the crude product was dissolved with ETAC (20 mL). The resulting solution was washed with water (20 mL) and then brine (20 mL). Precipitated by-products were removed by filtration through a thin layer of silica. The solution was dried over Mg<sub>2</sub>SO<sub>4</sub> and concentrated by rotatory evaporation to yield a colourless oil (210 mg, 77 %). <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>): δ 1.68-1.72 (m, 4H), 1.80-1.86 (m, 4H), 2.53 (t, 4H, J=7.5), 2.62 (t, 4H, J=7.5), 2.81 (s, broad, 8H). <sup>13</sup>C NMR (75 MHz, CDCl<sub>3</sub>): δ 26.34, 28.24, 31.10, 33.15, 33.79, 171.04 and 171.89. MS (C<sub>18</sub>H<sub>24</sub>O<sub>8</sub>N<sub>2</sub>S, nanoESI): calculated [M+H]<sup>+</sup> 429.46, found 429.37.

**S-Methyl 5,5'-Thiodipentanoylhydroxysuccinimide Iodide.** 5,5'-Thiodipentanoylhydroxysuccinimide (50 mg, 0.12 mmol) and iodomethane (68 mg, 0.48 mmol) in chloroform (5 mL) were reacted for four days. The solvent was removed by rotary evaporation to yield a crude pale yellow solid. <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>): δ 1.82-1.86 (m, 8H), 2.55 (t, 4H, J=6.5), 2.84 (s, 8H), 3.25 (s, 3H), 4.35 (t, 4H, J=7.5). <sup>13</sup>C NMR (126 MHz, CDCl<sub>3</sub>): δ 24.83, 25.41, 25.38, 28.32, 33.18, 43.63, 170.72, and 171.95. MS (C<sub>19</sub>H<sub>27</sub>O<sub>8</sub>N<sub>2</sub>S, nanoESI): calculated [M+H]<sup>+</sup> 443.49, found 443.16.

### 7.3.2: Cross-linking of Ac-AAKA and Caerin 1.10

The peptides Ac-AAKA and caerin 1.10 (GLLSVLGSAKHVLPVVPVIAEKL) were dissolved in anhydrous dimethyl sulfoxide (DMSO) to a concentration of 1 mM. The cross-linkers were dissolved in dry DMSO and added to the respective peptides at a 1:1 mole ratio (Ac-AAKA) and a 3:1 mole ratio (caerin 1.10). The mixtures were incubated at room temperature for 1 hour before being concentrated to dryness using a SpeedVac concentrator. The remaining precipitates were reconstituted with water and either desalted by C<sub>18</sub> ZipTip (Millipore, USA) before analysis or used without further modifications.

### 7.3.3: Cross-linking of Ubiquitin

Ubiquitin was dissolved in phosphate buffered saline (PBS, pH 8) to a concentration of 100  $\mu$ M (99  $\mu$ L), and the appropriate cross-linker dissolved in DMSO (1  $\mu$ L, 100 mM) was added to a final molar ratio of 10:1. Following 1 hour incubation at room temperature to allow cross-linking to occur, the solution was diluted with  $\text{NH}_4\text{CO}_3\text{H}$  (90  $\mu$ L, 25 mM) and 1 % (w/v) of Rapigest (10  $\mu$ L, Waters, UK) in  $\text{NH}_4\text{CO}_3\text{H}$  (25 mM) was added. This was briefly vortexed and heated at 80 °C for 10 min. Dithiothreitol (5.5  $\mu$ L, 100 mM) was added to the solution and briefly vortexed. The solution was heated at 60 °C for 10 min and then cooled to room temperature. Iodoacetamide (5.5  $\mu$ L, 300 mM) was added and the solution briefly vortexed and stored in darkness at room temperature for 30 mins. Trypsin (1  $\mu$ L, 1 ng/ $\mu$ L) was then added and the sample was incubated at 37 °C overnight. The resultant solution was stored at -20 °C until analysed by either IM-MS or LC-IM-MS.

### 7.3.4: Cross-linking of *h*PFD Subunits 5 and 6 to A $\beta$ (1-40)

*h*PFD subunits 5 and 6 were buffer exchanged into PBS (pH 8) to a concentration of 50  $\mu$ M. Independent solutions or mixtures of the respective subunit and A $\beta$ (1-40) at a 1:1 ratio were incubated at room temperature for 1 hour. S55'THS dissolved in DMSO (1  $\mu$ L, 250 mM) was added to each solution at a final molar ratio of 50:1. Following an overnight incubation at room temperature to allow cross-linking to occur, the solution was briefly vortexed and heated at 80 °C for 10 min. DTT (5.5  $\mu$ L, 100 mM) was added to the solution and briefly vortexed. The solution was heated at 60 °C for 10 min and then cooled to room temperature. IAA (5.5  $\mu$ L, 300 mM) was added and the solution briefly vortexed and stored in darkness at room temperature for 30 mins. Trypsin (50  $\mu$ L, 20 ng/ $\mu$ L) was then added and the sample was incubated at 37 °C overnight. The resultant solution was stored at -20 °C until analysed by either LC-MS/MS or LC-IM-MS/MS.

### 7.3.5: Waters Synapt IM-MS

IM-MS analysis was performed using a Synapt system (Waters, Manchester, UK). The sample was ionised using nano-electrospray ionisation, and IM-MS instrument parameters for positive ion mode were typically as follows: capillary voltage, 1.6 kV; cone voltage, 60 V; trap collision energy, 6 V; source temperature, 100 °C; backing pressure, 2.5 mBar; IMS cell pressure ( $\text{N}_2$ ), 0.5 mBar; travelling wave velocity, 300  $\text{m.s}^{-1}$ ; travelling wave height, 8 V. IM-MS instrument parameter changes for negative ion mode were: travelling wave velocity, 200  $\text{m.s}^{-1}$ ; travelling wave height, 9 V. After separating according to mobility, ions were fragmented in the transfer region of the instrument for CID experiments using a transfer collision energy of 30 eV for S55'THS linked peptides and 70 eV for DSP linked peptides.

MS data were processed using the program MassLynx (Waters, Manchester, UK, version 4.1). Cross-linked peptides were predicted using MeroX software [273] and confirmed by manual identification of expected cross-linker fragments aligned in drift time with the precursor ion.

### 7.3.6: Waters Synapt LC-IM-MS

LC-IM-MS analysis was performed using a Waters nanoACQUITY UPLC system fitted with an Atlantis C18 300  $\mu\text{m}$  x 150 mm column (3  $\mu\text{m}$  particle size) held at 35 °C coupled to a Synapt HDMS system (Waters, Manchester, UK). Chromatographic separation at a flow rate of 10  $\mu\text{L}\cdot\text{min}^{-1}$  was performed using a solvent composition profile of 1 % mobile phase B held for 1 min, followed by a linear gradient to 40 % mobile phase B over 15 min, the column was then washed with 99 % mobile phase B from 15.1 – 17 min followed by re-equilibration for 3 min at 1 % mobile phase B. Mobile phase A and B consisted of 999:1 (v:v)  $\text{H}_2\text{O}$ :FA and 999:1 (v:v) ACN:FA respectively.

The LC eluent was ionised using electrospray ionisation, and IM-MS instrument parameters were as described in Section 6.3.5 with the following exceptions: fragmentation in the transfer region of the instrument for CID experiments using a transfer collision energy of 42 eV for S55'THS linked peptides and 45 eV for DSP linked peptides.

MS data were processed using the program MassLynx (Waters, Manchester, UK, version 4.1). The LC chromatogram was binned into 20 equal bins, and drift-time plots and were generated for each 1 min bin. Similarly, mass spectra were generated from 0.5 ms bins in drift time, and were then searched for cross-link products. Cross-linked peptides were predicted using MeroX software and confirmed by manual identification of expected cross-linker fragments aligned in drift time with the precursor ion.

### 7.3.6: Agilent 6560 LC-MS/MS

LC-MS/MS analysis of CXL peptides were performed using an Infinity BioHPLC connected to a 6560 IM-MS system (Agilent, California, USA). Agilent 1290 Infinity II BioHPLC system fitted with an Advance Bio Peptide Map column 2.1 x 250 mm column (2.7  $\mu\text{m}$  particle size) held at 60 °C coupled to a 6560 IM-MS system (Agilent, California, USA). The sample was ionised by electrospray ionisation, and LC-MS/MS instrument parameters for positive mode are as follows:

Chromatographic separation at a flow rate of 0.4 mL/min was performed using a solvent composition profile of 3 % mobile phase B held for 1 min, followed by a linear gradient to 45 % mobile phase B over 30 min, followed by a linear gradient to 60 % mobile phase B over 5 min and the column was then washed with 97 % mobile phase B from 36-50 min followed by re-equilibration

for 6 min at 3 % mobile phase B. Mobile phase A and B consisted of 999:1 (v:v) water:formic acid and 999:1 (v:v) acetonitrile:formic acid respectively.

The LC eluent was ionised using electrospray ionisation, and MS instrument parameters were typically as follows: capillary voltage, 3.5 kV; nozzle voltage, 1000 V; source temperature, 100 °C; IMS cell pressure (N<sub>2</sub>), 0.5 mBar; auto MS/MS, top 10 abundant ions. Ions were fragmented in the collision cell of the instrument using a calculated ramped collision energy. MS data were processed using the program MassHunter Qualitative Analysis (Agilent, California, USA, version B.07.00). Cross-linked peptides analysed by LC-MS/MS were matched using MeroX software [273].

### 7.3.7: Agilent 6560 LC-IM-MS/MS

LC-IM-MS/MS analysis of CXL peptides were analysed following parameters previously described in Section 7.3.6. IM spectra were acquired, and following separation according to mobility, ions were fragmented in the collision cell of the instrument using a calculated ramped collision energy. IM-MS data were processed using the program IM-MS Browser (Agilent, California, USA, version B.07.01) and cross-links identified in the LC-MS/MS analysis by MeroX were manually found in the LC-IM-MS/MS acquired spectra.

# Chapter 8: Summary, Conclusions and Future Directions

---

Previous investigations into the biological functions and structure of PFD have primarily utilised species of prokaryotic origins. This hexameric protein is ubiquitously expressed and has been found to chaperone misfolding substrates into their correct native conformation. PFD selectively inhibits the aggregation of the substrate through a holdase mechanism, forming transient interactions with the target protein through its amphipathic coils. These distal tips are an important structural aspect of the chaperone and are part of an overall jelly fish like shape. The distal tentacles of eukaryotic PFD are formed from 6 homologous subunits, and the assembly of these subunits give rise to a large central amphipathic groove within the interior of the chaperone. It is within this highly conserved groove that the majority of refolding interactions are known to occur, in an independent manner.

The ubiquitous nature and chaperoning mechanism of eukaryotic PFD has implicated the protein for potential therapy in various diseases, including cancer and neurodegenerative diseases such as Huntington's, Parkinson's and Alzheimer's disease. Understanding the precise mechanism of PFD has prompted much research, but as yet, has provided little clarity in terms of the potential role of PFD and its subunits to these debilitating diseases.

Spectroscopic and biophysical analyses of the eukaryotic PFD and its subunits presented here provide further structural and mechanistic insights to this elusive protein. The effect of PFD on the misfolding of substrate A $\beta$ (1-40) was investigated to further our understanding of the potential therapy of PFD in the aforementioned disease states.

## 8.1: Summary of Results

The first focus of work presented here was to determine whether the co-incubation of eukaryotic PFD with the monomeric A $\beta$ (1-40) would affect the conversion of the misfolding substrate to amyloid fibrils. The purpose of these experiments was to provide some mechanistic insight to the possible chaperoning effects of PFD. The interaction of eukaryotic PFD with A $\beta$ (1-40) was assessed *in vitro* using ThT fluorescence assays, and it was determined that eukaryotic PFD and its subunits were able to prevent fibril formation by the Alzheimer's disease related protein. *b*PFD was seen to reduce fibril formation by 76 %, while a reduction of 67 % and 18 % was observed for *h*PFD subunit 6 and 5, respectively. The chaperones were found to redirect the aggregation of A $\beta$ (1-40), leading to the formation of monomeric intermediates. Cytotoxicity studies of the chaperoned product using

cultured PC12 cells revealed that *h*PFD subunit 6 formed an intermediate that had a higher toxicity than the substrate formed from *h*PFD subunit 5.

A combination of high and low resolution techniques were then applied to investigate the structure of PFD at various levels, with the aim of gaining further insight into its biological activity from a molecular view point. Far-UV CD analysis of the *h*PFD subunits 5 and 6 confirmed the proteins were primarily of  $\alpha$ -helical structure. The studies conducted in here demonstrated the adverse effect on secondary structure with increases in pH for both subunits. However, a slight decrease in pH was seen to increase the MRE and chromophore absorption. Melting profiles of the subunits were also obtained to give insight into the folding and unfolding mechanism of the chaperones. Far-UV analysis of *h*PFD subunit 5 following thermal aggregation revealed consistent data with previous protein melting profiles while *h*PFD subunit 6 was found to have a high resistance to the thermal stimulus and did not follow typical general melting curves.

In addition, the secondary structures of the incubated product following co-incubation of *h*PFD subunit 5 or 6 with the monomeric A $\beta$ (1-40) gave new insight to the overall structure of chaperoned A $\beta$ (1-40) and the high associated cytotoxicity.

Using a combination of 2D and 3D NMR spectroscopy, the determination of a high resolution structure of *h*PFD subunits 5 and 6 was attempted. Unfortunately, the results in this Chapter concluded that these protein complexes are not suitable for this technique under the conditions employed. It was found that 3D NMR experiments required long periods of acquisition that ultimately resulted in protein precipitation. Although there were some well resolved peaks obtained in 2D NMR spectroscopy, the intensity of the peaks were low, peak broadening was observed and many spin systems were missing. By analysis of spectra and secondary shifts  $^1\text{H}$ - $^1\text{H}$  NMR TOSCY of *h*PFD subunits 5 and 6, the majority of spin systems seen in the spectra were found to indicate  $\alpha$ -helical content. It was determined that the low number of peaks seen in the spectra was most likely due to molecular dynamics as typical oligomerisation of the protein can cause line-broadening. Precipitation observed following acquisition also revealed aggregation and low solubility of the *h*PFD subunits in conditions required for NMR analysis.

Native IM-MS analysis of *b*PFD allowed the CCS of the intact hexamer to be determined. Utilising the developed CTC nanospray source enabled IM-MS analysis at a range of temperatures, allowing the delineation of the equilibria for the protein complex before analysis. Increases in temperature were seen to shift the equilibria away from the intact hexameric form into sub-complexes that increased in size and polydispersity at elevated temperatures.

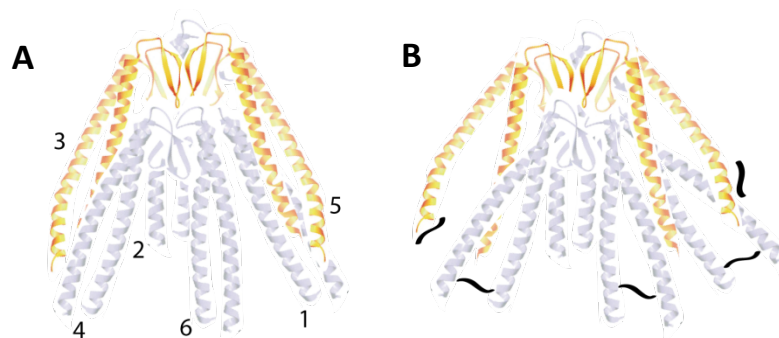
In addition, the IM-MS analysis of *h*PFD subunits 5 and 6 revealed a combination of monomer, dimer and trimeric species. It was therefore determined that *h*PFD subunits 5 and 6 have the ability to form oligomers as a result of self-chaperoning activity, whereby intact monomers bind to slightly unfolded monomers to prevent their aggregation. The incorporation of these additional subunits increases the oligomeric size and dispersity so that there may be a potential increase in substrate reactivity and improved chaperone efficiency.

Thermal dissociation assays conducted on the subunits revealed the resilience of the proteins to thermal stimuli. In particular, *h*PFD subunit 6 was seen to uphold its oligomeric states at high temperatures. The calculated CCS values and CCSDs were observed to increase in value with the progression of the experiment, demonstrating the structural unfolding effects as a result of thermal aggregation.

Finally, the focus of this study was to determine the binding interfaces of *h*PFD subunits 5 and 6, including possible structure reorganisation to give an insight into the potential chaperoning mechanism of eukaryotic PFD. Both subunits were first independently subjected to CXL using facile CID cleavable cross-linker reagents and analysed by LC-IM-MS/MS. The results reinforced previous tweezer-like models of the subunits. The addition of the misfolding substrate A $\beta$ (1-40) to the assay demonstrated flexibility of the apolar amphipathic distal branches of the tweezer like chaperones. It was determined that A $\beta$ (1-40) was encapsulated by *h*PFD subunit 6, while the substrate formed a more superficial binding to *h*PFD subunit 5. Both subunits revealed extension of their central cavity upon chaperoning. A higher number of cross-linker modifications was seen for *h*PFD subunit 5 and it was determined that a larger extension occurred for the PFD  $\alpha$  subunit. In totality however, the evidence presented indicated that the chaperoning of A $\beta$ (1-40) resulted in a significant structural reorganisation of the *h*PFD subunits 5 and 6. Furthermore, this work demonstrated the utility of IM in analysis of protein CXL studies.

## 8.2: Overall Conclusion

The work presented in this thesis demonstrates the vast array of well established and emerging biophysical techniques in structural biology. Low resolution studies, which combine structural data from a variety of techniques in an integrative approach, are becoming commonplace to characterise many biological processes. The application and development of existing methods enabled the potential structural model of eukaryotic PFD as illustrated in Figure 8.1.



**Figure 8.1: Schematic representation of eukaryotic PFD. (A) Possible eukaryotic PFD structure at equilibria. (B) Possible eukaryotic PFD structure during chaperone activity of Aβ(1-40). Aβ(1-40) is represented by the black curved lines.**

Figure 8.1A reveals the possible substoichiometric subunit arrangement of eukaryotic PFD, where a possible internal arrangement occurs between subunits 1, 4, 6 and 2, 3, 5 as evidenced in Section 6.2.2. Unfortunately, the construction of the models in Figure 8.1 is based on several assumptions obtained from the data of the individual PFD subunits. The distal branches of the hexamer in Figure 8.1A are of  $\alpha$ -helical content and the regions of subunit contact are seen to be of  $\beta$ -sheet structure. Although the secondary structures of the intact complex were not elucidated by far-UV CD, it is highly likely that the results in Section 4.2.1 are able to represent the collective  $\alpha$  and  $\beta$  subunits of the chaperone. The primary sequence typically dictates the secondary structure of proteins and this sequence remains the same regardless of the tertiary and quaternary protein level.

Figure 8.1B illustrates the possible flexibility of the distal branches during chaperone activity of Aβ(1-40). Flexibility was seen in the CXL data obtained in Section 7.2.6 for the *h*PFD subunits 5 and 6 during chaperone activity. Unfortunately, the mechanism of *b*PFD was not explored by CXL it must be noted that Figure 8.1B is based on the assumption that the subunits behave in a similar flexible manner when they are part of the intact PFD complex. Therefore, it is cautioned that the schematic diagram in Figure 8.1B is highly speculative and further work is required to determine the binding mechanisms of the eukaryotic chaperone. The multivalent subunits of eukaryotic PFD are suggested to identify and interact with the misfolding substrate through the extending of its branches. Although, the chaperoning abilities of *h*PFD subunits 5 and 6 may not necessarily reflect the interactions of eukaryotic PFD, this flexible scenario seen in Figure 8.1B, aligns with the data obtained in Section 3.2.4, where chaperone efficiency was seen to significantly increase between the eukaryotic PFD subunits and the intact chaperone complex. In conclusion, this thesis has increased our structural knowledge of the elusive eukaryotic PFD and has given us an insight into the possible mechanisms involved in its chaperoning activity.



### 8.3: Future Directions

Defining the structure-function relationship of eukaryotic PFD is crucial to understanding the mechanisms involved in protein chaperoning. More importantly, delineating this relationship has the potential to inform therapeutic designs, the exploitation of the ubiquitous PFD could combat neurodegenerative diseases that are associated with protein aggregation. For example, drugs could potentially be designed to trigger the neurological upregulation of PFD, which could inhibit and prevent protein aggregation.

Further work is required to obtain structure and mechanistic insights for eukaryotic PFD rather than making assumptions based on the studies of individual eukaryotic PFD subunits. Unfortunately work on the *b*PFD was restricted due to minimal amounts of sample. It is suggested here that continued structure investigation of eukaryotic PFD should consider high resolution crystallography. Further studies in CXL and LC-IM-MS/MS are also suggested, utilising *b*PFD and other misfolding substrates to investigate the previously proposed structural extension of the central cavity. Mutation studies could also aid in determining the specific residues of PFD involved in substrate protein binding and interaction.

Further biological studies such as ThT assays, to be conducted on the dissociating sub-complexes of *b*PFD, will allow the elucidation of the significance of these trimers. The possible dynamics of *b*PFD should also be investigated in the presence of crowding agents that can mimic the intracellular environment such as polyethylene glycol. To investigate the possibility of molten-globule binding characteristics of PFD, analyses conducted using near-UV CD and ANS-binding experiments would be able to determine whether the constructs had molten-globule character.

Overall, there are many more studies and researches to be conducted on eukaryotic PFD before its potential therapeutic benefits might be realised.

## References:

- [1] F. H. C. Crick, "On protein synthesis," *Symp. Soc. Exp. Biol.*, vol. XII, pp. 139–163, 1956.
- [2] W. Wedemeyer and H. Scheraga, "Protein folding: overview of pathways," *Encycl. Life Sci.*, pp. 1–9, 2001.
- [3] A. J. Cozzzone, "Proteins: fundamental chemical properties," *Life Sci.*, pp. 1–10, 2002.
- [4] J. C. Kendrew, R. E. Dickerson, B. E. Strandberg, R. G. Hart, D. R. Davies, D. C. Phillips, and V. C. Shore, "Structure of myoglobin: a three-dimensional fourier synthesis at 2 Å resolution," *Nature*, vol. 185, no. 4711, pp. 422–427, 1960.
- [5] J. C. Kendrew, G. Bodo, H. M. Dintzis, R. G. Parrish, and H. Wyckoff, "A three-dimensional model of the lyoglobin molecule obtained by x-ray analysis," *Nature*, vol. 181, pp. 662–666, 1958.
- [6] C. J. Oldfield and K. A. Dunker, "Intrinsically disordered proteins and intrinsically disordered protein regions," *Annu. Rev. Biochem.*, vol. 83, pp. 553–84, 2014.
- [7] E. Jurneczko, F. Cruickshank, M. Porrini, P. Nikolova, I. D. G. Campuzano, M. Morris, and P. E. Barran, "Intrinsic disorder in proteins: a challenge for (un)structural biology met by ion mobility-mass spectrometry," *Biochem. Soc. Trans.*, vol. 40, no. 5, pp. 1021–1026, 2012.
- [8] H. J. Dyson and P. E. Wright, "Intrinsically unstructured proteins and their functions," *Nat. Rev. Mol. Cell Biol.*, vol. 6, no. 3, pp. 197–208, 2005.
- [9] C. V Robinson, A. Sali, and W. Baumeister, "The molecular sociology of the cell," *Nature*, vol. 450, no. 7172, pp. 973–982, 2007.
- [10] K. Teilum, J. G. Olsen, and B. B. Kragelund, "Protein stability, flexibility and function," *Biochim. Biophys. Acta*, vol. 1814, no. 8, pp. 969–976, 2011.
- [11] E. J. Hendy, P. J. Tomiak, M. J. Collins, J. Hellstrom, A. W. Tudhope, J. M. Lough, and K. E. H. Penkman, "Assessing amino acid racemization variability in coral intra-crystalline protein for geochronological applications," *Geochim. Cosmochim. Acta*, vol. 86, pp. 338–353, 2012.
- [12] F. U. Hartl and M. Hayer-Hartl, "Molecular chaperones in the cytosol: from nascent chain to folded protein," *Science (80-. )*, vol. 295, no. 5561, pp. 1852–8, 2002.
- [13] A. M. C. Marcelino and L. M. Gierasch, "Roles of  $\beta$ -turns in protein folding: from peptide models to protein engineering," *Biopolymers*, vol. 89, no. 5, pp. 380–391, 2008.
- [14] E. H. C. Bromley and K. J. Channon, "Alpha-helical peptide assemblies: giving new function to designed structures," *Prog. Mol. Biol. Transl. Sci.*, vol. 103, pp. 231–275, 2011.
- [15] A. J. Doig, "Stability and design of  $\alpha$ -helical peptides," *Prog. Mol. Biol. Transl. Sci.*, vol. 83, pp. 1–52, 2008.

- [16] S. Vijay-Kumar, C. E. Bugg, and W. J. Cook, "Structure of ubiquitin refined at 1.8 Å resolution," *J. Mol. Biol.*, vol. 194, no. 3, pp. 531–544, Apr. 1987.
- [17] H. Noguchi, K. L. Campbell, C. Ho, S. Unzai, S. Y. Park, and J. R. H. Tame, "Structures of haemoglobin from woolly mammoth in liganded and unliganded states," *Acta Crystallogr. Sect. D*, vol. 68, no. 11, pp. 1441–1449, 2012.
- [18] J. H. Tomlinson, S. Ullah, P. E. Hansen, and M. P. Williamson, "Characterization of salt bridges to lysines in the protein G B1 domain," *J. Am. Chem. Soc.*, vol. 131, no. 13, pp. 4674–4684, 2009.
- [19] G. I. Makhatadze, V. V. Loladze, D. N. Ermolenko, X. Chen, and S. T. Thomas, "Contribution of surface salt bridges to protein stability: guidelines for protein engineering," *J. Mol. Biol.*, vol. 327, no. 5, pp. 1135–1148, 2003.
- [20] J. M. Mas, P. Aloy, M. A. Martí-Renom, B. Oliva, R. De Llorens, F. X. Avilés, and E. Querol, "Classification of protein disulphide-bridge topologies," *J. Comput. Aided. Mol. Des.*, vol. 15, no. 5, pp. 477–487, 2001.
- [21] A. J. Doig and D. H. Williams, "Is the hydrophobic effect stabilizing or destabilizing in proteins? The contribution of disulphide bonds to protein stability," *J. Mol. Biol.*, vol. 217, no. 2, pp. 389–398, 1991.
- [22] R. L. Baldwin, "Dynamic hydration shell restores Kauzmann's 1959 explanation of how the hydrophobic factor drives protein folding," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 111, no. 36, pp. 13052–6, 2014.
- [23] A. Kaushik and D. Gupta, "Protein folding grand challenge: hydrophobic vs. hydrophilic forces," *J. Biomol. Struct. Dyn.*, vol. 31, no. 9, pp. 1008–10, 2013.
- [24] C. B. Anfinsen, "Principles that govern the folding of protein chains," *Nature*, vol. 20, no. 516, pp. 493–494, Sep. 1979.
- [25] R. L. Baldwin, "Energetics of protein folding kinetics," *J. Mol. Biol.*, no. 371, pp. 283–301, 2007.
- [26] A. Verma, S. M. Gopal, A. Schug, T. Herges, K. Klenin, and W. Wenzel, "All-atom protein folding with free-energy forcefields," *Prog. Mol. Biol. Transl. Sci.*, vol. 83, pp. 181–253, 2008.
- [27] A. R. Dinner, A. Šali, L. J. Smith, C. M. Dobson, and M. Karplus, "Understanding protein folding via free-energy surfaces from theory and experiment," *Trends Biochem. Sci.*, vol. 25, no. 7, pp. 331–339, 2000.
- [28] R. L. Karpel, "The illusive search for the lowest free energy state of globular proteins and RNAs," *DNA Repair (Amst.)*, vol. 21, pp. 158–162, 2014.
- [29] C. M. Dobson and M. Karplus, "The fundamentals of protein folding: bringing together theory

- and experiment," *Curr. Opin. Struct. Biol.*, vol. 9, no. 1, pp. 92–101, 1999.
- [30] C. M. Dobson, "Principles of protein folding, misfolding and aggregation," *Semin. Cell Dev. Biol.*, vol. 15, no. 1, pp. 3–16, 2004.
  - [31] F. U. Hartl and M. Hayer-Hartl, "Converging concepts of protein folding in vitro and in vivo," *Nat. Struct. Mol. Biol.*, vol. 16, no. 6, pp. 574–581, 2009.
  - [32] T. R. Jahn and S. E. Radford, "Folding versus aggregation: polypeptide conformations on competing pathways," *Arch. Biochem. Biophys.*, vol. 469, no. 1, pp. 100–117, 2008.
  - [33] I. Tavernelli, S. Cotesta, and E. E. Di Iorio, "Protein dynamics, thermal stability, and free-energy landscapes: a molecular dynamics investigation," *Biophys. J.*, vol. 85, no. 4, pp. 2641–2649, 2003.
  - [34] S. E. Jackson, "How do small single domain proteins fold?," *Fold. Des.*, vol. 3, no. 4, pp. 81–91, 1998.
  - [35] D. J. Brockwell and S. E. Radford, "Intermediates: ubiquitous species on folding energy landscapes?," *Curr. Opin. Struct. Biol.*, vol. 17, no. 1, pp. 30–37, 2007.
  - [36] D. J. Brockwell, D. A. Smith, and S. E. Radford, "Protein folding mechanisms: new methods and emerging ideas," *Curr. Opin. Struct. Biol.*, vol. 10, no. 1, pp. 16–25, 2000.
  - [37] Y. Yoshimura, Y. Lin, H. Yagi, Y.-H. Lee, H. Kitayama, K. Sakurai, M. So, H. Ogi, H. Naiki, and Y. Goto, "Distinguishing crystal-like amyloid fibrils and glass-like amorphous aggregates from their kinetics of formation," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 109, no. 36, pp. 14446–14451, 2012.
  - [38] H. Ecroyd and J. A. Carver, "Unraveling the mysteries of protein folding and misfolding," *Int. Union Biochem. Mol. Biol. Life*, vol. 60, no. 12, pp. 769–774, 2008.
  - [39] H. Ecroyd and J. A. Carver, "Crystallin proteins and amyloid fibrils," *Cell. Mol. Life Sci.*, vol. 66, no. 1, pp. 62–81, 2009.
  - [40] K. F. Winklhofer, J. Tatzelt, and C. Haass, "The two faces of protein misfolding: gain- and loss-of-function in neurodegenerative diseases," *Eur. Mol. Biol. Organ. J.*, vol. 27, no. 2, pp. 336–349, 2008.
  - [41] P. T. Lansbury, "Evolution of amyloid: what normal protein folding may tell us about fibrillogenesis and disease," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 96, no. March, pp. 3342–3344, 1999.
  - [42] L. C. Serpell, "Alzheimer's amyloid fibrils: structure and assembly," *Biochim. Biophys. Acta*, vol. 1502, no. 1, pp. 16–30, 2000.
  - [43] J. L. Jimenez, J. I. Guijarro, E. Orlova, and H. R. Saibil, "Cryo-electron microscopy structure of an SH3 amyloid fibril and model of the molecular packing," *Eur. Mol. Biol. Organ. J.*, vol. 18, no.

- 4, pp. 815–821, 1999.
- [44] M. Sunde and C. C. F. Blake, “From the globular to the fibrous state: protein structure and structural conversion in amyloid formation,” *Q. Rev. Biophys.*, vol. 31, no. 1, pp. 1–39, 1998.
  - [45] T. E. Golde, D. R. Borchelt, B. I. Giasson, and J. Lewis, “Thinking laterally about neurodegenerative proteinopathies,” *J. Clin. Invest.*, vol. 123, no. 5, pp. 1847–1855, 2013.
  - [46] D. C. Thorn, “Amyloid Fibril Formation by Bovine Milk  $\alpha$ s2-and  $\kappa$ -casein, and Its Inhibition by The Molecular Chaperones  $\alpha$ s- and  $\beta$ -casein,” The University of Adelaide, South Australia, 2010.
  - [47] M. Stefani and C. M. Dobson, “Protein aggregation and aggregate toxicity: new insights into protein folding, misfolding diseases and biological evolution,” *J. Mol. Med.*, vol. 81, no. 11, pp. 678–699, 2003.
  - [48] J. L. Tomic, A. Pensalfini, E. Head, and C. G. Glabe, “Soluble fibrillar oligomer levels are elevated in Alzheimer’s disease brain and correlate with cognitive dysfunction,” *Neurobiol. Dis.*, vol. 35, no. 3, pp. 352–358, 2009.
  - [49] C. M. Sondag, G. Dhawan, and C. K. Combs, “Beta amyloid oligomers and fibrils stimulate differential activation of primary microglia,” *J. Neuroinflammation*, vol. 6, no. 1, pp. 1–13, 2009.
  - [50] V. Novitskaya, O. V Bocharova, I. Bronstein, and I. V Baskakov, “Amyloid fibrils of mammalian prion protein are highly toxic to cultured cells and primary neurons,” *J. Biol. Chem.*, vol. 281, no. 19, pp. 13828–13836, 2006.
  - [51] M. Sakono and T. Zako, “Amyloid oligomers: formation and toxicity of A $\beta$  oligomers,” *Fed. Eur. Biochem. Soc. J.*, vol. 277, no. 6, pp. 1348–1358, 2010.
  - [52] M. Stefani, “Structural features and cytotoxicity of amyloid oligomers: implications in Alzheimer’s disease and other diseases with amyloid deposits,” *Prog. Neurobiol.*, vol. 99, no. 3, pp. 226–245, 2012.
  - [53] M. Fändrich, “Oligomeric intermediates in amyloid formation: structure determination and mechanisms of toxicity,” *J. Mol. Biol.*, vol. 421, no. 4–5, pp. 427–440, 2012.
  - [54] M. Verma, A. Vats, and V. Taneja, “Toxic species in amyloid disorders: oligomers or mature fibrils,” *Ann. Indian Acad. Neurol.*, vol. 18, no. 2, pp. 138–145, 2015.
  - [55] A. L. Gharibyan, V. Zamotin, K. Yanamandra, O. S. Moskaleva, B. A. Margulis, I. A. Kostanyan, and L. A. Morozova-Roche, “Lysozyme amyloid oligomers and fibrils induce cellular death via different apoptotic/necrotic pathways,” *J. Mol. Biol.*, vol. 365, no. 5, pp. 1337–1349, 2007.
  - [56] C. Haass and D. J. Selkoe, “Soluble protein oligomers in neurodegeneration: lessons from the Alzheimer’s amyloid  $\beta$ -peptide,” *Nat. Rev. Mol. Cell Biol.*, vol. 8, no. 2, pp. 101–112, 2007.

- [57] P. Salahuddin, M. T. Fatima, A. S. Abdelhameed, S. Nusrat, and R. H. Khan, "Structure of amyloid oligomers and their mechanisms of toxicities: Targeting amyloid oligomers using novel therapeutic approaches," *Eur. J. Med. Chem.*, vol. 114, pp. 41–58, 2016.
- [58] I. Benilova, E. Karran, and B. D. Strooper, "The toxic A $\beta$  oligomer and Alzheimer's disease: an emperor in need of clothes," *Nat. Neurosci.*, vol. 15, no. 3, pp. 349–357, 2012.
- [59] A. Demuro, I. Parker, and G. E. Stutzmann, "Calcium signaling and amyloid toxicity in Alzheimer disease," *J. Biol. Chem.*, vol. 285, no. 17, pp. 12463–12468, 2010.
- [60] S. M. Butterfield and H. A. Lashuel, "Amyloidogenic protein-membrane interactions: mechanistic insight from model systems," *Angew. Chemie Int. Ed.*, vol. 49, no. 33, pp. 5628–5654, 2010.
- [61] T. Nakagawa, H. Zhu, N. Morishima, E. Li, J. Xu, B. A. Yankner, and J. Yuan, "Caspase-12 mediates endoplasmic-reticulum-specific apoptosis and cytotoxicity by amyloid- $\beta$ ," *Nature*, vol. 403, no. 6765, pp. 98–103, 2000.
- [62] R. J. Ellis, "Do molecular chaperones have to be proteins?," *Biochem. Biophys. Res. Commun.*, vol. 238, no. 3, pp. 687–692, 1997.
- [63] F. U. Hartl, A. Bracher, and M. Hayer-Hartl, "Molecular chaperones in protein folding and proteostasis," *Nature*, vol. 475, pp. 324–332, 2011.
- [64] J. E. Gestwicki, "Special series: molecular chaperones in protein folding and disease," *Biopolymers*, vol. 93, no. 3, pp. 209–210, 2010.
- [65] T. M. Treweek, A. M. Morris, and J. A. Carver, "Intracellular protein unfolding and aggregation: the role of small heat-shock chaperone proteins," *Aust. J. Chem.*, vol. 56, no. 5, pp. 357–367, 2003.
- [66] H. Saibil, "Chaperone machines for protein folding, unfolding and disaggregation," *Nat. Rev. Mol. Cell Biol.*, vol. 14, no. 10, pp. 630–642, 2013.
- [67] C. Jolly and R. I. Morimoto, "Role of the heat shock response and molecular chaperones in oncogenesis and cell death," *J. Natl. Cancer Inst.*, vol. 92, no. 19, pp. 1564–1572, 2000.
- [68] Y. E. Kim, M. S. Hipp, A. Bracher, M. Hayer-Hartl, and F. U. Hartl, "Molecular chaperone functions in protein folding and proteostasis," *Annu. Rev. Biochem.*, no. 82, pp. 323–355, 2013.
- [69] D. M. Hatters, R. A. Lindner, J. A. Carver, and G. J. Howlett, "The molecular chaperone,  $\alpha$ -crystallin, inhibits amyloid formation by apolipoprotein C-II," *J. Biol. Chem.*, vol. 276, no. 36, pp. 33755–33761, 2001.
- [70] J. A. Carver, A. Rekas, D. C. Thorn, and M. R. Wilson, "Small heat-shock proteins and clusterin: intra- and extracellular molecular chaperones with a common mechanism of action and

- function?," *Int. Union Biochem. Mol. Biol. Life*, vol. 55, no. 12, pp. 661–668, 2003.
- [71] B. Bukau and A. L. Horwich, "The Hsp70 and Hsp60 chaperone machines," *Cell*, vol. 92, no. 3, pp. 351–366, 1998.
  - [72] B. A. Kaufman, J. E. Kolesar, P. S. Perlman, and R. A. Butow, "A function for the mitochondrial chaperonin Hsp60 in the structure and transmission of mitochondrial DNA nucleoids in *Saccharomyces cerevisiae*," *J. Cell Biol.*, vol. 163, no. 3, pp. 457–461, 2003.
  - [73] M. Okochi, T. Yoshida, T. Maruyama, Y. Kawarabayasi, H. Kikuchi, and M. Yohda, "Pyrococcus prefoldin stabilizes protein-folding intermediates and transfers them to chaperonins for correct folding," *Biochem. Biophys. Res. Commun.*, vol. 291, no. 4, pp. 769–774, 2002.
  - [74] S. Kim, K. R. Willison, and A. L. Horwich, "Cytosolic chaperonin subunits have a conserved ATPase domain but diverged polypeptide-binding domains," *Trends Biochem. Sci.*, vol. 19, no. 12, pp. 543–548, 1994.
  - [75] Y. Cong, M. L. Baker, J. Jakana, D. Woolford, E. J. Miller, S. Reissmann, R. N. Kumar, A. M. Redding-Johanson, T. S. Batth, A. Mukhopadhyay, S. J. Ludtke, J. Frydman, and W. Chiu, "4.0-Å resolution cryo-EM structure of the mammalian chaperonin TRiC/CCT reveals its unique subunit arrangement," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 107, no. 11, pp. 4967–4972, 2010.
  - [76] A. Y. Yam, Y. Xia, H.-T. J. Lin, A. Burlingame, M. Gerstein, and J. Frydman, "Defining the TRiC/CCT interactome links chaperonin function to stabilization of newly made proteins with complex topologies," *Nat. Struct. Mol. Biol.*, vol. 15, no. 12, pp. 1255–1262, 2008.
  - [77] S. Geissler, K. Siegers, and E. Schiebel, "A novel protein complex promoting formation of functional  $\alpha$ - and  $\gamma$ -tubulin," *Eur. Mol. Biol. Organ. J.*, vol. 17, no. 4, pp. 952–966, 1998.
  - [78] W. J. Hansen, N. J. Cowan, and W. J. Welch, "Prefoldin-nascent chain complexes in the folding of cytoskeletal proteins," *J. Cell Biol.*, vol. 145, no. 2, pp. 265–277, 2000.
  - [79] I. E. Vainberg, S. A. Lewis, H. Rommelaere, C. Ampe, J. Vandekerckhove, H. L. Klein, and N. J. Cowan, "Prefoldin, a chaperone that delivers unfolded proteins to cytosolic chaperonin," *Cell*, vol. 93, no. 5, pp. 863–873, 1998.
  - [80] M. Takano, E. Tashiro, A. Kitamura, H. Maita, S. M. M. Iguchi-Ariga, M. Kinjo, and H. Ariga, "Prefoldin prevents aggregation of  $\alpha$ -synuclein," *Brain Res.*, no. 1542, pp. 186–194, 2014.
  - [81] M. Sakono, T. Zako, H. Ueda, M. Yohda, and M. Maeda, "Formation of highly toxic soluble amyloid beta oligomers by the molecular chaperone prefoldin," *Fed. Eur. Biochem. Soc. J.*, vol. 275, no. 23, pp. 5982–5993, 2008.
  - [82] K. M. Sörgjerd, T. Zako, M. Sakono, P. C. Stirling, M. R. Leroux, T. Saito, P. Nilsson, M. Sekimoto, T. C. Saido, and M. Maeda, "Human prefoldin inhibits amyloid- $\beta$  (A $\beta$ ) fibrillation and contributes to formation of nontoxic A $\beta$  aggregates," *Biochemistry*, vol. 52, no. 20, pp.

- 3532–3542, 2013.
- [83] T. Zako, Y. Murase, R. Iizuka, T. Yoshida, T. Kanzaki, N. Ide, M. Maeda, T. Funatsu, and M. Yohda, "Localization of prefoldin interaction sites in the hyperthermophilic group II chaperonin and correlations between binding rate and protein transfer rate," *J. Mol. Biol.*, vol. 364, no. 1, pp. 110–120, 2006.
  - [84] A. Ohtaki, H. Kida, Y. Miyata, N. Ide, A. Yonezawa, T. Arakawa, R. Iizuka, K. Noguchi, A. Kita, M. Odaka, K. Miki, and M. Yohda, "Structure and molecular dynamics simulation of archaeal prefoldin: the molecular mechanism for binding and recognition of nonnative substrate proteins," *J. Mol. Biol.*, vol. 376, no. 4, pp. 1130–1141, 2008.
  - [85] V. F. Lundin, P. C. Stirling, J. Gomez-Reino, J. C. Mwenifumbo, J. M. Obst, J. M. Valpuesta, and M. R. Leroux, "Molecular clamp mechanism of substrate binding by hydrophobic coiled-coil residues of the archaeal chaperone prefoldin," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 101, no. 13, pp. 4367–4372, 2004.
  - [86] T. A. Whitehead, B. B. Boonyaratanakornkit, V. Höllrigl, and D. S. Clark, "A filamentous molecular chaperone of the prefoldin family from the deep-sea hyperthermophile *Methanocaldococcus jannaschii*," *Protein Sci.*, vol. 16, pp. 626–634, 2007.
  - [87] D. J. Glover and D. S. Clark, "Oligomeric assembly is required for chaperone activity of the filamentous  $\gamma$ -prefoldin," *Fed. Eur. Biochem. Soc. J.*, vol. 282, no. 15, pp. 2985–2997, 2015.
  - [88] H. Rommelaere, M. De Neve, K. Neirynck, D. Peelaers, D. Waterschoot, M. Goethals, N. Fraeyman, J. Vandekerckhove, and C. Ampe, "Prefoldin recognition motifs in the nonhomologous proteins of the actin and tubulin families," *J. Biol. Chem.*, vol. 276, no. 44, pp. 41023–41028, 2001.
  - [89] C. T. Simons, A. Staes, H. Rommelaere, C. Ampe, S. A. Lewis, and N. J. Cowan, "Selective contribution of eukaryotic prefoldin subunits to actin and tubulin binding," *J. Biol. Chem.*, vol. 279, no. 6, pp. 4196–4203, 2004.
  - [90] G. Millán-Zambrano and S. Chávez, "Nuclear functions of prefoldin," *Open Biol.*, vol. 4, no. 140085, pp. 1–10, 2014.
  - [91] J. Martín-Benito, J. Gómez-Reino, P. C. Stirling, V. F. Lundin, P. Gómez-Puertas, J. Boskovic, P. Chacón, J. J. Fernández, J. Berenguer, M. R. Leroux, and J. M. Valpuesta, "Divergent substrate-binding mechanisms reveal an evolutionary specialization of eukaryotic prefoldin compared to its archaeal counterpart," *Structure*, vol. 15, no. 1, pp. 101–110, 2007.
  - [92] D. Bogumil, D. Alvarez-Ponce, G. Landan, J. O. McInerney, and T. Dagan, "Integration of two ancestral chaperone systems into one: the evolution of eukaryotic molecular chaperones in light of eukaryogenesis," *Mol. Biol. Evol.*, vol. 31, no. 2, pp. 410–418, 2014.



- [93] R. Siegert, M. R. Leroux, C. Scheufler, F. U. Hartl, and I. Moarefi, "Structure of the molecular chaperone prefoldin: unique interaction of multiple coiled coil tentacles with unfolded proteins," *Cell*, vol. 103, no. 4, pp. 621–632, 2000.
- [94] M. Fändrich, M. A. Tito, M. R. Leroux, A. A. Rostom, F. U. Hartl, C. M. Dobson, and C. V. Robinson, "Observation of the noncovalent assembly and disassembly pathways of the chaperone complex MtGimC by mass spectrometry," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 97, no. 26, pp. 14151–14155, 2000.
- [95] J. Martín-Benito, J. Boskovic, P. Gomez-Puertas, J. L. Carrascosa, C. T. Simons, S. A. Lewis, F. Bartolini, N. J. Cowan, and J. M. Valpuesta, "Structure of eukaryotic prefoldin and of its complexes with unfolded actin and the cytosolic chaperonin CCT," *Eur. Mol. Biol. Organ. J.*, vol. 21, no. 23, pp. 6377–6386, 2002.
- [96] M. Okochi, T. Nomura, T. Zako, T. Arakawa, R. Iizuka, H. Ueda, T. Funatsu, M. Leroux, and M. Yohda, "Kinetics and binding sites for interaction of the prefoldin with a group II chaperonin. Contiguous non-native substrate and chaperonin binding sites in the archaeal prefoldin," *J. Biol. Chem.*, vol. 279, no. 30, pp. 31788–31795, 2004.
- [97] H. P. Erickson, "Evolution of the cytoskeleton," *Bioessays*, vol. 29, no. 7, pp. 668–677, 2007.
- [98] M. R. Leroux, M. Fändrich, D. Klunker, K. Siegers, A. N. Lupas, J. R. Brown, E. Schiebel, C. M. Dobson, and F. U. Hartl, "MtGimC, a novel archaeal chaperone related to the eukaryotic chaperonin cofactor GimC/prefoldin," *Eur. Mol. Biol. Organ. J.*, vol. 18, no. 23, pp. 6730–6743, 1999.
- [99] V. F. Lundin, M. R. Leroux, and P. C. Stirling, "Quality control of cytoskeletal proteins and human disease," *Trends Biochem. Sci.*, vol. 35, no. 5, pp. 288–297, 2010.
- [100] K. Siegers, T. Waldmann, M. R. Leroux, K. Grein, A. Shevchenko, E. Schiebel, and F. U. Hartl, "Compartmentation of protein folding in vivo: sequestration of non-native polypeptide by the chaperonin-GimC system," *Eur. Mol. Biol. Organ. J.*, vol. 18, no. 1, pp. 75–84, 1999.
- [101] T. Zako, R. Iizuka, M. Okochi, T. Nomura, T. Ueno, H. Tadakuma, M. Yohda, and T. Funatsu, "Facilitated release of substrate protein from prefoldin by chaperonin," *Fed. Eur. Biochem. Soc. Lett.*, vol. 579, no. 17, pp. 3718–3724, 2005.
- [102] E. Tashiro, T. Zako, H. Muto, Y. Ito, K. Sörgjerd, N. Terada, A. Abe, M. Miyazawa, A. Kitamura, H. Kitaura, H. Kubota, M. Maeda, T. Momoi, S. M. M. Iguchi-Arigo, M. Kinjo, and H. Ariga, "Prefoldin protects neuronal cells from polyglutamine toxicity by preventing aggregation formation," *J. Biol. Chem.*, vol. 288, no. 27, pp. 19958–19972, 2013.
- [103] S. Cao, G. Carlesso, A. B. Osipovich, J. Llanes, Q. Lin, K. L. Hoek, W. N. Khan, and H. E. Ruley, "Subunit 1 of the prefoldin chaperone complex is required for lymphocyte development and

- function," *J. Immunol.*, no. 181, pp. 476–484, 2008.
- [104] J. Zhang, Q. Han, Y. Song, Q. Chen, and X. Xia, "Analysis of subcellular prefoldin 1 redistribution during rabies virus infection," *Jundishapur J. Microbiol.*, vol. 8, no. 7, pp. 1–7, 2015.
- [105] H. Tsuchiya, T. Iseda, and O. Hino, "Identification of a novel protein (VBP-1) binding to the von Hippel-Lindau (VHL) tumor suppressor gene product," *Cancer Res.*, vol. 56, no. 13, pp. 2881–2885, 1996.
- [106] S. Hennecke, J. Beck, K. Bornemann-Kolatzki, S. Neumann, H. M. Escobar, I. Nolte, S. C. Hammer, M. Hewicker-Trautwein, J. Junginger, F. J. Kaup, B. Brenig, and E. Schütz, "Prevalence of the prefoldin subunit 5 gene deletion in canine mammary tumors," *Public Libr. Sci. One*, vol. 10, no. 7, pp. 1–10, 2015.
- [107] X. Li, P. Mooney, S. Zheng, C. Booth, M. B. Braunfeld, D. A. Agard, and Y. Cheng, "Electron counting and beam-induced motion correction enable near atomic resolution single particle cryoEM," *Nat. Methods*, vol. 10, no. 6, pp. 584–590, 2013.
- [108] C. Huang and C. G. Kalodimos, "Structures of Large Protein Complexes Determined by Nuclear Magnetic Resonance Spectroscopy," *Annu. Rev. Biophys.*, vol. 46, no. 1, pp. 317–336, 2017.
- [109] E. Callaway, "The revolution will not be crystallized," *Nature*, vol. 525, pp. 172–174, 2015.
- [110] S. Jonic and C. Vénien-Bryan, "Protein structure determination by electron cryo-microscopy," *Curr. Opin. Pharmacol.*, vol. 9, no. 5, pp. 636–642, 2009.
- [111] H. D. T. Mertens and D. I. Svergun, "Structural characterization of proteins and complexes using small-angle X-ray solution scattering," *J. Struct. Biol.*, vol. 172, no. 1, pp. 128–141, 2010.
- [112] N. Berova and K. Nakanishi, *Circular Dichroism: Principles and Applications*, 2nd ed. New York: John Wiley & Sons, 2000.
- [113] E. D. Merkley, E. S. Baker, K. L. Crowell, D. J. Orton, T. Taverner, C. Ansong, Y. M. Ibrahim, M. C. Burnet, J. R. Cort, G. A. Anderson, R. D. Smith, and J. N. Adkins, "Mixed-isotope labeling with LC-IMS-MS for characterization of protein-protein interactions by chemical cross-linking," *J. Am. Soc. Mass Spectrom.*, vol. 24, no. 3, pp. 444–449, 2014.
- [114] S. D. Pringle, K. Giles, J. L. Wildgoose, J. P. Williams, S. E. Slade, K. Thalassinou, R. H. Bateman, M. T. Bowers, and J. H. Scrivens, "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.*, vol. 261, no. 1, pp. 1–12, 2007.
- [115] M. R. H. Krebs, E. H. C. Bromley, and A. M. Donald, "The binding of thioflavin-T to amyloid fibrils: localisation and implications," *J. Struct. Biol.*, vol. 149, no. 1, pp. 30–37, 2005.
- [116] M. Biancalana and S. Koide, "Molecular mechanism of Thioflavin-T binding to amyloid fibrils,"

- Proteins and Proteomics*, vol. 1804, no. 7, pp. 1405–1412, 2010.
- [117] R. Khurana, C. Coleman, C. Ionescu-Zanetti, S. A. Carter, V. Krishna, R. K. Grover, R. Roy, and S. Singh, "Mechanism of thioflavin T binding to amyloid fibrils," *J. Struct. Biol.*, vol. 151, no. 3, pp. 229–238, 2005.
  - [118] M. R. Wilson, J. J. Yerbury, and S. Poon, "Potential roles of abundant extracellular chaperones in the control of amyloid formation and toxicity," *Mol. Biosyst.*, vol. 4, no. 1, pp. 42–52, 2008.
  - [119] M. V Berridge and A. S. Tan, "Characterization of the cellular reduction of 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT): subcellular localization, substrate dependence, and involvement of mitochondrial electron transport in MTT reduction," *Arch. Biochem. Biophys.*, vol. 303, no. 2, pp. 474–482, 1993.
  - [120] M. V Berridge, P. M. Herst, and A. S. Tan, "Tetrazolium dyes as tools in cell biology: new insights into their cellular reduction," *Biotechnol. Annu. Rev.*, vol. 11, pp. 127–152, 2005.
  - [121] D. H. Williams and I. Fleming, *Spectroscopic Methods on Organic Chemistry*, 5th ed. Mc Graw-Hill, 1995.
  - [122] E. D. Becker, "A brief history of nuclear magnetic resonance," *Anal. Chem.*, vol. 65, no. 6, pp. 295–302, 1993.
  - [123] K. Wüthrich, "Protein structure determination in solution by NMR spectroscopy," *J. Biol. Chem.*, vol. 265, no. 36, pp. 22059–22062, 1990.
  - [124] R. M. Silverstein, F. X. Webster, D. J. Kiemle, and D. L. Bryce, *Spectrometric identification of organic compounds*, 8th ed. New York: Wiley, 2014.
  - [125] J. B. Lambert and E. P. Mazzola, *Nuclear Magnetic Resonance: An Introduction to Principles, Applications and Experimental Methods*, 2nd ed. Pearson Education, 2004.
  - [126] J. N. S. Evans, *Biomolecular NMR Spectroscopy*, 1st ed. Oxford, 1995.
  - [127] D. S. Wishart, B. D. Sykes, and F. M. Richards, "Relationship between nuclear magnetic resonance chemical shift and protein secondary structure," *J. Mol. Biol.*, vol. 222, no. 2, pp. 311–333, 1991.
  - [128] A. N. Calabrese, "Characterisation of Protein Structure and Interactions: Novel Applications to the Study of Bioactive Peptides," The University of Adelaide, South Australia, 2013.
  - [129] A. M. Gronenborn and G. M. Clore, "Protein structure determination in solution by two-dimensional and three-dimensional nuclear magnetic resonance spectroscopy," *Anal. Chem.*, vol. 62, no. 1, pp. 2–15, 1990.
  - [130] G. M. Clore and A. M. Gronenborn, "Determination of three-dimensional structures of proteins and nucleic acids in solution by nuclear magnetic resonance spectroscopy," *Crit. Rev. Biochem. Mol. Biol.*, vol. 24, no. 5, pp. 479–564, 1989.

- [131] P. Legault, J. Li, J. Mogridge, L. E. Kay, and J. Greenblatt, "NMR structure of the bacteriophage  $\lambda$  N peptide/boxB RNA complex: Recognition of a GNRA fold by an arginine-rich motif," *Cell*, vol. 93, no. 2, pp. 289–299, 1998.
- [132] R. W. Kriwacki and P. T. Pitner, "Current aspects of practical two-dimensional (2D) nuclear magnetic resonance (NMR) spectroscopy: applications to structure elucidation," *Pharm. Res.*, vol. 6, no. 7, pp. 531–555, 1989.
- [133] W. P. Aue, E. Bartholdi, and R. R. Ernst, "Two-dimensional spectroscopy. Application to nuclear magnetic resonance," *J. Chem. Phys.*, vol. 64, no. 5, pp. 2229–2248, 1976.
- [134] D. G. Davis and A. Bax, "Assignment of complex proton NMR spectra via two-dimensional homonuclear Hartmann-Hahn spectroscopy," *J. Am. Chem. Soc.*, vol. 107, no. 9, pp. 2820–2821, 1985.
- [135] L. Braunschweiler and R. R. Ernst, "Coherence transfer by isotropic mixing: Application to proton correlation spectroscopy," *J. Magn. Reson.*, vol. 53, no. 3, pp. 521–528, 1983.
- [136] S. R. Hartmann and E. L. Hahn, "Nuclear double resonance in the rotating frame," *Phys. Rev.*, vol. 128, no. 5, pp. 2042–2053, 1962.
- [137] M. H. Levitt, D. Suter, and R. R. Ernst, "Spin dynamics and thermodynamics in solid state NMR cross polarization," *J. Chem. Phys.*, vol. 84, no. 8, pp. 4243–4255, 1986.
- [138] T. Herrmann, P. Güntert, and K. Wüthrich, "Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA," *J. Mol. Biol.*, vol. 319, no. 1, pp. 209–227, 2002.
- [139] K. Wüthrich, *NMR of Proteins and Nucleic Acids*. New York: Wiley, 1986.
- [140] D. Neuhaus and M. P. Williamson, *The Nuclear Overhauser Effect in Structural and Conformational Analysis*. New York: VCH Publishers, 1989.
- [141] S. W. Englander and A. J. Wand, "Main-chain-directed strategy for the assignment of  $^1\text{H}$  NMR spectra of proteins," *Biochemistry*, vol. 26, no. 19, pp. 5953–5958, 1987.
- [142] D. S. Wishart, C. G. Bigam, A. Holm, R. S. Hodges, and B. D. Sykes, " $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects," *J. Biomol. NMR*, vol. 5, no. 1, pp. 67–81, 1995.
- [143] D. L. D. Stefan and J. A. Wand, "Two-dimensional  $^1\text{H}$  NMR study of human ubiquitin: a main chain directed assignment and structure analysis," *Biochemistry*, no. 26, pp. 7272–7281, 1987.
- [144] D. C. Dalgarno, B. A. Levine, and R. J. P. Williams, "Structural information from NMR secondary chemical shifts of peptide  $\alpha$  C-H protons in proteins," *Biosci. Rep.*, vol. 3, no. 5, pp. 443–452, 1983.

- [145] I. D. Kuntz, P. A. Kosen, and E. C. Craig, "Amide chemical shifts in many helices in peptides and proteins are periodic," *J. Am. Chem. Soc.*, vol. 113, no. 8, pp. 1406–1408, 1991.
- [146] A. Pastore and V. Saudek, "The relationship between chemical shift and secondary structure in proteins," *J. Magn. Reson.*, vol. 90, no. 1, pp. 165–176, 1990.
- [147] A. Bax and L. Lerner, "Two-dimensional nuclear magnetic resonance spectroscopy," *Am. Assoc. Adv. Sci.*, vol. 232, no. 4753, pp. 960–967, 1986.
- [148] A. Bax, "Two-dimensional NMR and protein structure," *Annu. Rev. Biochem.*, vol. 58, pp. 223–256, 1989.
- [149] D. Marion and K. Wüthrich, "Application of phase sensitive two-dimensional correlated spectroscopy (COSY) for measurements of  $^1\text{H}$ - $^1\text{H}$  spin-spin coupling constants in proteins," *Biochem. Biophys. Res. Commun.*, vol. 113, no. 3, pp. 967–974, 1983.
- [150] D. H. Corrêa and C. H. I. Ramos, "The use of circular dichroism spectroscopy to study protein folding, form and function," *African J. Biochem. Res.*, vol. 3, no. 5, pp. 164–173, 2009.
- [151] I. H. V Stokkum, H. J. Spoelder, M. Bloemendal, R. V Grondelle, and F. C. Groen, "Estimation of protein secondary structure and error analysis from circular dichroism spectra," *Anal. Biochem.*, vol. 191, no. 1, pp. 110–118, 1990.
- [152] N. J. Greenfield, "Using circular dichroism spectra to estimate protein secondary structure," *Nat. Protoc.*, vol. 1, no. 6, pp. 2876–2890, 2007.
- [153] N. J. Greenfield, "Determination of the folding of proteins as a function of denaturants, osmolytes or ligands using circular dichroism," *Nat. Protoc.*, vol. 1, no. 6, pp. 2733–2741, 2006.
- [154] A. J. Adler, N. J. Greenfield, and G. D. Fasman, "Circular dichroism and optical rotatory dispersion of proteins and polypeptides," *Methods Enzymol.*, vol. 27, no. 1965, pp. 675–735, 1973.
- [155] G. Holzwarth and P. Doty, "The ultraviolet circular dichroism of polypeptides," *J. Am. Chem. Soc.*, vol. 87, no. 4, pp. 218–228, 1965.
- [156] S. M. Kelly and N. C. Price, "The use of circular dichroism in the investigation of protein structure and function," *Curr. Proteins Pept. Sci.*, vol. 1, no. 4, pp. 349–84, 2000.
- [157] N. J. Greenfield, "Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions," *Nat. Protoc.*, vol. 1, no. 6, pp. 2527–2535, 2009.
- [158] E. A. Roman, J. Santos, and F. L. G. Flecha, "The use of circular dichroism methods to monitor unfolding transitions in peptides, globular and membrane proteins," in *Circular Dichroism: Theory and Spectroscopy*, D. S. Rodgers and S. C. Wihile, Eds. New York: Nova Science

- Publishers, 2017, pp. 217–254.
- [159] N. Sreerama and R. W. Woody, “Computational and analysis of protein circular dichroism spectra,” *Methods Enzymol.*, vol. 13, no. 1, pp. 318–351, 2004.
  - [160] L. Whitmore and B. A. Wallace, “DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data,” *Nucleic Acids Res.*, vol. 32, pp. 668–673, 2004.
  - [161] A. Lobley, L. Whitmore, and B. A. Wallace, “DICHROWEB: an interactive website for the analysis of protein secondary structure from circular dichroism spectra,” *Bioinformatics*, vol. 18, no. 1, pp. 211–212, 2002.
  - [162] L. Whitmore and B. A. Wallace, “Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases,” *Biopolymers*, vol. 89, no. 5, pp. 392–400, 2007.
  - [163] P. Manavalan and W. C. Johnson, “Variable selection method improves the prediction of protein secondary structure from circular dichroism spectra,” *Anal. Biochem.*, vol. 167, no. 1, pp. 76–85, 1987.
  - [164] A. J. Miles, L. Whitmore, and B. A. Wallace, “Spectral magnitude effects on the analyses of secondary structure from circular dichroism spectroscopic data,” *Protein Sci.*, vol. 14, no. 2, pp. 368–374, 2005.
  - [165] J. R. I. Yates, “A century of mass spectrometry: from atoms to proteomes,” *Nat. Methods*, vol. 8, no. 8, pp. 633–637, 2011.
  - [166] A. Chaudhry and H. Kleinpoppen, *Analysis of Excitation and Ionisation of Atoms and Molecules by Electron Impact*. Berlin: Springer, 2011.
  - [167] F. H. Field, “Chemical ionization mass spectrometry,” *Anal. Chem.*, vol. 1, pp. 42–49, 1968.
  - [168] K. Downard, *Mass Spectrometry*. London: The Royal Society of Chemistry, 2004.
  - [169] J. L. Norris and R. M. Caprioli, “Analysis of tissue specimens by matrix-assisted laser desorption/ionization imaging mass spectrometry in biological and clinical research,” *Chem. Rev.*, vol. 113, pp. 2309–2342, 2013.
  - [170] J. K. Lewis, J. Wei, and G. Siuzdak, “Matrix-assisted laser desorption/ionization mass spectrometry in peptide and protein analysis,” *Encycl. Anal. Chem.*, pp. 5880–5894, 2000.
  - [171] L. F. Marvin, M. A. Roberts, and L. B. Fay, “Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry in clinical chemistry,” *Clin. Chim. Acta*, no. 337, pp. 11–21, 2003.
  - [172] R. Aebersold and M. Mann, “Mass spectrometry-based proteomics,” *Nature*, vol. 422, no. 6928, pp. 198–207, 2003.
  - [173] J. L. P. Benesch and C. V Robinson, “Dehydrated but unharmed,” *Nature*, vol. 462, no. 3, pp.

- 576–577, 2009.
- [174] J. B. Fenn, M. Mann, C. K. Meng, S. F. Wong, and C. M. Whitehouse, “Electrospray ionization for mass spectrometry of large biomolecules,” *Am. Assoc. Adv. Sci.*, vol. 246, no. 4926, pp. 64–71, 1989.
- [175] R. J. Beuhler and L. Friedman, “Rapid heating in fragile molecule mass spectrometry,” *Int. J. Mass Spectrom. Ion Process.*, vol. 78, pp. 1–15, 1987.
- [176] M. Yamashita and J. B. Fenn, “Electrospray ion source. Another variation on the free-jet theme,” *J. Phys. Chem.*, vol. 88, no. 20, pp. 4451–4459, 1984.
- [177] C. M. Whitehouse, R. N. Dreyer, M. Yamashita, and J. B. Fenn, “Electrospray interface for liquid chromatographs and mass spectrometers,” *Anal. Chem.*, vol. 57, no. 3, pp. 675–679, 1985.
- [178] P. J. Dyson, I. Khalaila, S. Luetzgen, J. S. McIndoe, and D. Zhao, “Direct probe electrospray (and nanospray) ionization mass spectrometry of neat ionic liquids,” *Chem. Commun.*, no. 19, pp. 2204–2205, 2004.
- [179] A. Gapeev, A. Berton, and D. Fabris, “Current-controlled nanospray ionization mass spectrometry,” *J. Am. Soc. Mass Spectrom.*, vol. 20, no. 7, pp. 1334–1341, 2009.
- [180] R. Juraschek, T. Dülcks, and M. Karas, “Nanoelectrospray-more than just a minimized-flow electrospray ionization source,” *J. Am. Soc. Mass Spectrom.*, vol. 10, no. 4, pp. 300–8, 1999.
- [181] C. V Robinson, “Keeping it together,” *Royal Society of Chemistry*, 2003. [Online]. Available: <http://www.rsc.org/chemistryworld/Issues/2003/February/together.asp>. [Accessed: 21-Apr-2016].
- [182] A. F. M. Gavriilidou, B. Gülbakan, and R. Zenobi, “Influence of ammonium acetate concentration on receptor-ligand binding affinities measured by native nano ESI-MS: a systematic study,” *Anal. Chem.*, vol. 87, no. 20, pp. 10378–10384, 2015.
- [183] S. Banerjee and S. Mazumdar, “Electrospray ionization mass spectrometry: a technique to access the information beyond the molecular weight of the analyte,” *Int. J. Anal. Chem.*, pp. 1–40, 2012.
- [184] J. V. D. Greef, W. M. A. Niessen, and U. R. Tjaden, “Liquid chromatography-mass spectrometry approach the need for a multidimensional,” *J. Chromatogr.*, vol. 474, pp. 5–19, 1989.
- [185] J. J. Pitt, “Principles and applications of liquid chromatography-mass spectrometry in clinical biochemistry,” *Clin. Biochem. Rev.*, vol. 30, no. 1, pp. 19–34, 2009.
- [186] R. E. Ardrey, *Liquid Chromatography-Mass Spectrometry: An Introduction*, vol. 1. New York: Wiley, 2003.

- [187] I. V Chernushevich, A. V Loboda, and B. A. Thomson, "An introduction to quadrupole-time-of-flight mass spectrometry," *J. Mass Spectrom.*, vol. 36, no. 8, pp. 849–865, 2001.
- [188] W. Paul, "Electromagnetic traps for charged and neutral particles," *Rev. Mod. Phys.*, vol. 62, no. 3, pp. 531–540, 1990.
- [189] A. G. Marshall, C. L. Hendrickson, and G. S. Jackson, "Fourier transform ion cyclotron resonance mass spectrometry: a primer," *Mass Spectrom. Rev.*, vol. 17, pp. 1–35, 1998.
- [190] Q. Hu, R. J. Noll, H. Li, A. Makarov, M. Hardman, and G. R. Cooks, "The orbitrap: a new mass spectrometer," *J. Mass Spectrom.*, vol. 40, no. 4, pp. 430–443, 2005.
- [191] W. Paul, H. P. Reinhard, and U. von Zahn, "Das elektrische massenfilter als massenspektrometer und isotopentrenner," *Zeitschrift für Phys.*, vol. 152, no. 2, pp. 143–182, 1958.
- [192] J. Griffiths, "A brief history of mass spectrometry," *Anal. Chem.*, vol. 80, pp. 5678–5683, 2008.
- [193] D. J. Douglas, "Linear quadrupoles in mass spectrometry," *Mass Spectrom. Rev.*, vol. 28, pp. 937–960, 2009.
- [194] P. H. Dawson, *Quadrupole Mass Spectrometry and Its Applications*. Amsterdam: Elsevier, 1976.
- [195] D. A. Skoog, J. F. Holler, and S. R. Crouch, *Principles of Instrumental Analysis*, 5th ed. Belmont: Thomson, 2007.
- [196] J. Smolanoff, A. Lapicki, and S. L. Anderson, "Use of a quadrupole production mass filter for high energy resolution ion beam," *Adv. Mass Spectrom.*, vol. 66, no. 6, pp. 3706–3708, 1995.
- [197] V. H. Wysocki, K. A. Resing, Q. Zhang, and G. Cheng, "Mass spectrometry of peptides and proteins," *Methods*, no. 35, pp. 211–222, 2005.
- [198] K. F. Medzihradszky and R. J. Chalkley, "Lessons in de novo peptide sequencing by tandem mass spectrometry," *Mass Spectrom. Rev.*, vol. 34, pp. 43–63, 2013.
- [199] A. M. Palumbo, S. A. Smith, C. L. Kalcic, M. Dantus, P. M. Stemmer, and G. E. Reid, "Tandem mass spectrometry strategies for phosphoproteome analysis," *Mass Spectrom. Rev.*, vol. 30, pp. 600–625, 2011.
- [200] M. M. Wolff and W. E. Stephens, "A pulsed mass spectrometer with time dispersion," *Rev. Sci. Instrum.*, vol. 24, no. 8, pp. 616–617, 1953.
- [201] M. L. Alexandrov, N. V Krasnov, L. R. Lokshin, and A. V Chuprikov, "Discrimination effects in inorganic ion-cluster detection by secondary-electron multiplier in mass spectrometry experiments," *Rapid Commun. Mass Spectrom.*, vol. 4, no. 1, pp. 9–12, 1990.
- [202] K. L. Busch, "Detecting ions in mass spectrometers with the faraday cup," *Spectroscopy*, vol. 26, no. 11, 2011.



- [203] X. Liang, Y. Zhang, W. Chen, P. Cai, S. Zhang, X. Chen, and S. Shi, "High-speed counter-current chromatography coupled online to high performance liquid chromatography-diode array detector-mass spectrometry for purification, analysis and identification of target compounds from natural products," *J. Chromatogr. A*, vol. 1385, pp. 69–76, 2015.
- [204] J. L. Wiza, "Microchannel plate detectors," *Nucl. Instrum. Methods*, vol. 162, pp. 587–601, 1979.
- [205] D. M. Murphy and K. Mauersberger, "Operation of a microchannel plate counting system in a mass spectrometer," *Rev. Sci. Instrum.*, vol. 56, no. 2, pp. 220–226, 1985.
- [206] A. A. Shvartsburg and R. D. Smith, "Fundamentals of traveling wave ion mobility spectrometry," *Anal. Chem.*, vol. 80, no. 24, pp. 9689–9699, 2008.
- [207] S. Wickramasekara, F. Zandkarimi, J. Morré, J. Kirkwood, L. Legette, Y. Jiang, A. Gombart, J. Stevens, and C. Maier, "Electrospray quadrupole travelling wave ion mobility time-of-flight mass spectrometry for the detection of plasma metabolome changes caused by xanthohumol in obese zucker (fa/fa) rats," *Metabolites*, no. 3, pp. 701–717, 2013.
- [208] M. Kłoniecki, A. Jabłonowska, J. Poznański, J. Langridge, C. Hughes, I. D. G. Campuzano, K. Giles, and M. Dadlez, "Ion mobility separation coupled with MS detects two structural states of Alzheimer's disease A $\beta$ 1-40 peptide oligomers," *J. Mol. Biol.*, vol. 407, no. 1, pp. 110–124, 2011.
- [209] J. C. May, C. R. Goodwin, N. M. Lareau, K. L. Leaptrot, C. B. Morris, R. T. Kurulugama, A. Mordehai, C. Klein, W. Barry, E. Darland, G. Overney, K. Imatani, G. C. Stafford, J. C. Fjeldsted, and J. A. McLean, "Conformational ordering of biomolecules in the gas phase: nitrogen collision cross sections measured on a prototype high resolution drift tube ion mobility-mass spectrometer," *Anal. Chem.*, vol. 86, no. 4, pp. 2107–2116, 2014.
- [210] J. C. May and J. A. McLean, "Ion mobility-mass spectrometry: time-dispersive instrumentation," *Anal. Chem.*, vol. 87, no. 3, pp. 1422–1436, 2015.
- [211] C. W. N. Damen, W. Chen, A. B. Chakraborty, M. V Oosterhout, J. R. Mazzeo, J. C. Gebler, J. H. M. Schellens, H. Rosing, and J. H. Beijnen, "Electrospray ionization quadrupole ion-mobility time-of-flight mass spectrometry as a tool to distinguish the lot-to-lot heterogeneity in N-glycosylation profile of the therapeutic monoclonal antibody trastuzumab," *J. Am. Soc. Mass Spectrom.*, vol. 20, no. 11, pp. 2021–2033, 2009.
- [212] B. C. Bohrer, S. I. Merenbloom, S. L. Koeniger, A. E. Hilderbrand, and D. E. Clemmer, "Biomolecule analysis by ion mobility spectrometry," *Annu. Rev. Anal. Chem.*, vol. 1, no. 1, pp. 293–327, 2008.
- [213] G. F. Verbeck, B. T. Ruotolo, H. A. Sawyer, K. J. Gillig, and D. H. Russell, "A fundamental

- introduction to ion mobility mass spectrometry applied to the analysis of biomolecules,” *J. Biomol. Tech.*, vol. 13, no. 2, pp. 56–61, 2002.
- [214] A. Heptner, T. Reinecke, J. Langejuergen, and S. Zimmermann, “A gated atmospheric pressure drift tube ion mobility spectrometer-time-of-flight mass spectrometer,” *J. Chromatogr. A*, vol. 1356, pp. 241–248, 2014.
- [215] M. Fasciotti, P. M. Lalli, G. Heerdt, R. A. Steffen, Y. E. Corilo, G. F. de Sá, R. J. Daroda, F. d A. M. Reis, N. H. Morgon, R. C. L. Pereira, M. N. Eberlin, and C. F. Klitzke, “Structure-drift time relationships in ion mobility mass spectrometry,” *Int. J. Ion Mobil. Spectrom.*, vol. 16, no. 2, pp. 117–132, 2013.
- [216] R. Fernández-Maestre, “Ion mobility spectrometry: history, characteristics and applications,” *Rev. U.D.C.A Actual. Divulg. Científica*, vol. 15, no. 2, pp. 467–479, 2012.
- [217] E. Jurneczko, J. Kalapothakis, I. D. G. Campuzano, M. Morris, and P. E. Barran, “Effects of drift gas on collision cross sections of a protein standard in linear drift tube and traveling wave ion mobility mass spectrometry,” *Anal. Chem.*, vol. 84, no. 20, pp. 8524–8531, 2012.
- [218] F. Lanucara, S. W. Holman, C. J. Gray, and C. E. Eyers, “The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics,” *Nat. Chem.*, vol. 6, no. 4, pp. 281–294, 2014.
- [219] E. A. Mason and E. W. McDaniel, *Transport properties of ions in gases*. New York: Wiley, 1988.
- [220] A. R. Shah, K. Agarwal, E. S. Baker, M. Singhal, A. M. Mayampurath, Y. M. Ibrahim, L. J. Kangas, M. E. Monroe, R. Zhao, M. E. Belov, G. A. Anderson, and R. D. Smith, “Machine learning based prediction for peptide drift times in ion mobility spectrometry,” *Bioinformatics*, vol. 26, no. 13, pp. 1601–1607, 2010.
- [221] B. Wang, J. Zhang, P. Chen, Z. Ji, S. Deng, and C. Li, “Prediction of peptide drift time in ion mobility mass spectrometry from sequence-based features,” *Biomed. Cent.*, vol. 14, pp. 1–9, 2013.
- [222] R. Guevremont and R. W. Purves, “High field asymmetric waveform ion mobility spectrometry-mass spectrometry: an investigation of leucine enkephalin ions produced by electrospray ionization,” *J. Am. Soc. Mass Spectrom.*, vol. 10, no. 6, pp. 492–501, 1999.
- [223] D. M. Williams and T. L. Pukala, “Novel insight into protein misfolding diseases revealed by ion mobility-mass spectrometry,” *Mass Spectrom. Rev.*, vol. 32, pp. 169–187, 2013.
- [224] M. F. Mesleh, J. M. Hunter, A. A. Shvartsburg, G. C. Schatz, and M. F. Jarrold, “Structural information from ion mobility measurements: effects of the long-range potential,” *J. Phys. Chem.*, vol. 100, no. 40, pp. 16082–16086, 1996.
- [225] B. T. Ruotolo, S.-J. Hyung, P. M. Robinson, K. Giles, R. H. Bateman, and C. V Robinson, “Ion

- mobility–mass spectrometry reveals long-lived, unfolded intermediates in the dissociation of protein complexes,” *Angew. Chemie Int. Ed.*, vol. 46, no. 42, pp. 8001–8004, 2007.
- [226] B. T. Ruotolo, K. Giles, I. D. G. Campuzano, A. M. Sandercock, R. H. Bateman, and C. V. Robinson, “Evidence for macromolecular protein rings in the absence of bulk water,” *Sci. Rep.*, vol. 310, no. 12, pp. 1658–1661, 2005.
- [227] G. N. Sivalingam, J. Yan, H. Sahota, and K. Thalassinos, “Amphitrite: a program for processing travelling wave ion mobility mass spectrometry data,” *Int. J. Mass Spectrom.*, vol. 345–347, pp. 54–62, 2013.
- [228] E. R. Schenk, F. Nau, and F. Fernandez-Lima, “Theoretical predictor for candidate structure assignment from IMS data of biomolecule-related conformational space,” *Int. J. Ion Mobil. Spectrom.*, vol. 18, no. 1, pp. 1–13, 2015.
- [229] T. W. Knapman, J. T. Berryman, I. D. G. Campuzano, S. A. Harris, and A. E. Ashcroft, “Considerations in experimental and theoretical collision cross-section measurements of small molecules using travelling wave ion mobility spectrometry-mass spectrometry,” *Int. J. Mass Spectrom.*, vol. 298, no. 1–3, pp. 17–23, 2010.
- [230] D. P. Smith, T. W. Knapman, I. D. G. Campuzano, R. W. Malham, J. T. Berryman, S. E. Radford, and A. E. Ashcroft, “Deciphering drift time measurements from travelling wave ion mobility spectrometry-mass spectrometry studies,” *Eur. J. Mass Spectrom.*, vol. 15, no. 2, pp. 113–130, 2009.
- [231] C. A. Scarff, K. Thalassinos, G. R. Hilton, and J. H. Scrivens, “Travelling wave ion mobility mass spectrometry studies of protein structure: biological significance and comparison with X-ray crystallography and nuclear magnetic resonance spectroscopy measurements,” *Rapid Commun. Mass Spectrom.*, vol. 22, pp. 3297–3304, 2008.
- [232] C. Uetrecht, R. J. Rose, E. van Duijn, K. Lorenzen, and A. J. R. Heck, “Ion mobility mass spectrometry of proteins and protein assemblies,” *Chem. Soc. Rev.*, vol. 39, no. 5, pp. 1633–1655, 2010.
- [233] A. A. Shvartsburg and M. F. Jarrold, “An exact hard-spheres scattering model for the mobilities of polyatomic ions,” *Chem. Phys. Lett.*, vol. 261, no. 1–2, pp. 86–91, 1996.
- [234] S. E. Anderson, C. Bleiholder, E. R. Bocker, P. J. Stang, and M. T. Bowers, “A novel projection approximation algorithm for the fast and accurate computation of molecular collision cross sections (III): Application to supramolecular coordination-driven assemblies with complex shapes,” *Int. J. Mass Spectrom.*, vol. 330–332, pp. 78–84, 2012.
- [235] C. Bleiholder, T. Wyttenbach, and M. T. Bowers, “A novel projection approximation algorithm for the fast and accurate computation of molecular collision cross sections (II). Method,” *Int.*

- J. Mass Spectrom.*, vol. 308, no. 1, pp. 1–10, 2011.
- [236] C. Bleiholder, S. Contreras, and M. T. Bowers, “A novel projection approximation algorithm for the fast and accurate computation of molecular collision cross sections (IV). Application to polypeptides,” *Int. J. Mass Spectrom.*, vol. 354–355, pp. 275–280, 2013.
  - [237] C. Bleiholder, T. Wyttenbach, and M. T. Bowers, “A novel projection approximation algorithm for the fast and accurate computation of molecular collision cross sections (I). Method,” *Int. J. Mass Spectrom.*, vol. 308, no. 1, pp. 1–10, 2011.
  - [238] T. L. Pukala, B. T. Ruotolo, M. Zhou, A. Politis, R. Stefanescu, J. A. Leary, and C. V Robinson, “Subunit architecture of multiprotein assemblies determined using restraints from gas-phase measurements,” *Structure*, vol. 17, no. 9, pp. 1235–1243, 2009.
  - [239] S. Mehmood, T. M. Allison, and C. V Robinson, “Mass spectrometry of protein complexes: from origins to applications,” *Annu. Rev. Phys. Chem.*, no. 1, pp. 453–474, 2015.
  - [240] J. A. McLean, B. T. Ruotolo, K. J. Gillig, and D. H. Russell, “Ion mobility-mass spectrometry: a new paradigm for proteomics,” *Int. J. Mass Spectrom.*, vol. 240, pp. 301–315, 2005.
  - [241] Waters Corporation, *Waters Synapt High Definition Mass Spectrometry System; Quick Start Guide*. Massachusetts, 2007.
  - [242] Waters Corporation, *Waters Synapt High Definition Mass Spectrometry System; Operator’s Guide*. Massachusetts, 2007.
  - [243] R. Kurulugama, K. Imatani, and L. Taylor, *The Agilent Ion Mobility Q-TOF Mass Spectrometer System: Technical Overview*. California, 2013.
  - [244] *Agilent 6200 Series TOF and 6500 Series Q-TOF LC/MS System: Concepts Guide*. California: Agilent, 2014.
  - [245] N. F. Zinnel and D. H. Russell, “Size-to-charge dispersion of collision-induced dissociation product ions for enhancement of structural information and product ion identification,” *Anal. Chem.*, vol. 86, no. 10, pp. 4791–4798, 2014.
  - [246] J. H. Bowie, C. S. Brinkworth, and S. Dua, “Collision-induced fragmentations of the (M-H)<sup>−</sup> parent anions of underivatized peptides: an aid to structure determination and some unusual negative ion cleavages,” *Mass Spectrom. Rev.*, vol. 21, no. 2, pp. 87–107, 2002.
  - [247] Y. Hua, S. B. Wainhaus, Y. Yang, L. Shen, Y. Xiong, X. Xu, F. Zhang, J. L. Bolton, and R. B. van Breemen, “Comparison of negative and positive ion electrospray tandem mass spectrometry for the liquid chromatography tandem mass spectrometry analysis of oxidized deoxynucleosides,” *J. Am. Soc. Mass Spectrom.*, vol. 12, no. 1, pp. 80–87, 2000.
  - [248] K. F. Medzihradszky, “Peptide sequence analysis,” *Methods Enzymol.*, vol. 402, no. 2000, pp. 209–244, 2005.

- [249] J. S. Brodbelt, "Ion activation methods for peptides and proteins," *Anal. Chem.*, vol. 88, no. 1, pp. 30–51, 2016.
- [250] Y. Zhang, B. R. Fonslow, B. Shan, M. Baek, and J. R. Yates, "Protein analysis by shotgun/bottom-up proteomics," *Chem. Rev.*, vol. 113, no. 4, pp. 2343–2394, 2013.
- [251] J. R. Yates, "The revolution and evolution of shotgun proteomics for large-scale proteome analysis," *J. Am. Chem. Soc.*, vol. 135, pp. 1629–1640, 2013.
- [252] R. L. Gundry, M. Y. White, C. I. Murray, L. A. Kane, Q. Fu, B. A. Stanley, and J. E. van Eyk, "Preparation of proteins and peptides for mass spectrometry analysis in a bottom-up proteomics workflow," *Curr. Protoc. Mol. Biol.*, no. 88, pp. 1–29, 2009.
- [253] S. L. Volchenbom, K. Kristjansdottir, D. Wolfgeher, and S. J. Kron, "Rapid validation of Mascot search results via stable isotope labeling, pair picking, and deconvolution of fragmentation patterns," *Mol. Cell. Proteomics*, vol. 8, no. 8, pp. 2011–22, 2009.
- [254] A. Bairoch and R. Apweiler, "The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000," *Nucleic Acids Res.*, vol. 28, no. 1, pp. 45–48, 2000.
- [255] J. Reinders, U. Lewandrowski, J. Moebius, Y. Wagner, and A. Sickmann, "Challenges in mass spectrometry-based proteomics," *Proteomics*, vol. 4, no. 12, pp. 3686–3703, 2004.
- [256] D. A. Shepherd, M. T. Marty, K. Giles, A. J. Baldwin, and J. L. P. Benesch, "Combining tandem mass spectrometry with ion mobility separation to determine the architecture of polydisperse proteins," *Int. J. Mass Spectrom.*, vol. 377, pp. 663–671, 2014.
- [257] D. A. Shepherd, K. Holmes, D. J. Rowlands, N. J. Stonehouse, and A. E. Ashcroft, "Using ion mobility spectrometry-mass spectrometry to decipher the conformational and assembly characteristics of the hepatitis B capsid protein," *Biophys. J.*, vol. 105, no. 5, pp. 1258–1267, 2013.
- [258] L. F. A. Santos, A. H. Iglesias, E. J. Pilau, A. F. Gomes, and F. C. Gozzo, "Traveling-wave ion mobility mass spectrometry analysis of isomeric modified peptides arising from chemical cross-linking," *J. Am. Soc. Mass Spectrom.*, vol. 21, no. 12, pp. 2062–2069, 2010.
- [259] A. N. Calabrese and T. L. Pukala, "Chemical cross-linking and mass spectrometry for the structural analysis of protein assemblies," *Aust. J. Chem.*, no. 66, pp. 749–759, 2013.
- [260] Q. Zheng, H. Zhang, L. Tong, S. Wu, and H. Chen, "Cross-linking electrochemical mass spectrometry for probing protein three-dimensional structures," *Anal. Chem.*, vol. 86, pp. 8983–8991, 2014.
- [261] A. Sinz, "The advancement of chemical cross-linking and mass spectrometry for structural proteomics: from single proteins to protein interaction networks," *Expert Rev. Proteomics*, vol. 11, no. 6, pp. 733–743, 2014.

- [262] F. Stengel, R. Aebersold, and C. V. Robinson, "Joining Forces: Integrating Proteomics and cross-linking with the mass spectrometry of intact complexes," *Mol. Cell. Proteomics*, vol. 11, no. 3, pp. 1–14, 2012.
- [263] X. Tang, G. R. Munske, W. F. Siems, and J. E. Bruce, "Mass spectrometry identifiable cross-linking strategy for studying protein-protein interactions," *Anal. Chem.*, vol. 77, no. 1, pp. 311–318, 2005.
- [264] G. T. Hermanson, "Heterobifunctional crosslinkers," in *Bioconjugate Techniques*, 3rd ed., London: Academic Press, 2013, pp. 299–339.
- [265] G. T. Hermanson, "Homobifunctional crosslinkers," in *Bioconjugate Techniques*, 3rd ed., London: Academic Press, 2013, pp. 275–298.
- [266] F. Chu, S. Mahrus, C. S. Craik, and A. L. Burlingame, "Isotope-coded and affinity-tagged cross-linking (ICATXL): an efficient strategy to probe protein interaction surfaces," *J. Am. Chem. Soc.*, vol. 128, no. 32, pp. 10362–10363, Aug. 2006.
- [267] D. Vellucci, A. Kao, R. M. Kaake, S. D. Rychnovsky, and L. Huang, "Selective enrichment and identification of azide-tagged cross-linked peptides using chemical ligation and mass spectrometry," *J. Am. Soc. Mass Spectrom.*, vol. 21, no. 8, pp. 1432–1445, 2010.
- [268] X. Zhang and Y. Zhang, "Applications of azide-based bioorthogonal click chemistry in glycobiology," *Molecules*, vol. 18, no. 6, pp. 7145–7159, 2013.
- [269] J. M. Baskin and C. R. Bertozzi, "Bioorthogonal click chemistry: covalent labeling in living systems," *QSAR Comb. Sci.*, vol. 26, no. 11–12, pp. 1211–1219, 2007.
- [270] S. M. Chowdhury, X. Du, N. Tolic, S. Wu, R. J. Moore, U. M. Mayer, R. D. Smith, and J. N. Adkins, "Identification of cross-linked peptides after click-based enrichment using sequential CID and ETD tandem mass spectrometry," *Anal. Chem.*, vol. 81, no. 13, pp. 1–18, 2009.
- [271] A. Sinz and K. Wang, "Mapping spatial proximities of sulfhydryl groups in proteins using a fluorogenic cross-linker and mass spectrometry," *Anal. Biochem.*, vol. 331, no. 1, pp. 27–32, 2004.
- [272] J. W. Back, A. F. Hartog, H. L. Dekker, A. O. Muijsers, L. J. De Koning, and L. De Jong, "A new crosslinker for mass spectrometric analysis of the quaternary structure of protein complexes," *J. Am. Soc. Mass Spectrom.*, vol. 12, pp. 222–227, 2001.
- [273] M. Götze, J. Pettelkau, R. Fritzsche, C. H. Ihling, M. Schäfer, and A. Sinz, "Automated assignment of MS/MS cleavable cross-links in protein 3D-structure analysis," *J. Am. Soc. Mass Spectrom.*, vol. 26, no. 9, pp. 83–97, 2014.
- [274] M. Q. Müller, F. Dreiocker, C. H. Ihling, M. Schäfer, and A. Sinz, "Cleavable cross-linker for protein structure analysis: reliable identification of cross-linking products by tandem MS,"

- Anal. Chem.*, vol. 82, no. 16, pp. 6958–6968, 2010.
- [275] A. Kao, C.-L. Chiu, D. Vellucci, Y. Yang, V. R. Patel, S. Guan, A. Randall, P. Baldi, S. D. Rychnovsky, and L. Huang, “Development of a novel cross-linking strategy for fast and accurate identification of cross-linked peptides of protein complexes,” *Mol. Cell. Proteomics*, vol. 10, no. 1, pp. 1–17, 2011.
- [276] E. J. Soderblom and M. B. Goshe, “Collision-induced dissociative chemical cross-linking reagents and methodology: applications to protein structural characterization using tandem mass spectrometry analysis,” *Anal. Chem.*, vol. 78, no. 23, pp. 8059–8068, 2006.
- [277] P. Murphy and H. LeVine, “Alzheimer’s disease and the  $\beta$ -amyloid peptide,” *Natl. Inst. Heal.*, vol. 19, no. 1, pp. 311–328, 2010.
- [278] E. Karran, M. Mercken, and B. D. Strooper, “The amyloid cascade hypothesis for Alzheimer’s disease: an appraisal for the development of therapeutics,” *Nat. Rev.*, vol. 10, no. 9, pp. 698–712, 2011.
- [279] C. L. Masters and D. J. Selkoe, “Biochemistry of amyloid  $\beta$ -protein and amyloid deposits in Alzheimer disease,” *Cold Spring Harb. Perspect. Med.*, no. 2, pp. 1–24, 2012.
- [280] M. Nutu, H. Zetterberg, E. Londos, L. Minthon, K. Nägga, K. Blennow, O. Hansson, and A. Öhrfelt, “Evaluation of the cerebrospinal fluid amyloid- $\beta$ 1-42/ amyloid- $\beta$ 1-40 ratio measured by alpha-LISA to distinguish Alzheimer’s disease from other dementia disorders,” *Dement. Geriatr. Cogn. Disord.*, vol. 36, no. 1–2, pp. 99–110, 2013.
- [281] J. Ghanta, C. L. Shen, L. L. Kiessling, and R. M. Murphy, “A strategy for designing inhibitors of  $\beta$ -amyloid toxicity,” *J. Biol. Chem.*, vol. 271, no. 47, pp. 29525–29528, 1996.
- [282] L. M. Young, J. C. Saunders, R. A. Mahood, C. H. Revill, R. J. Foster, L.-H. Tu, D. P. Raleigh, S. E. Radford, and A. E. Ashcroft, “Screening and classifying small-molecule inhibitors of amyloid formation using ion mobility spectrometry-mass spectrometry,” *Nat. Chem.*, vol. 7, no. 1, pp. 73–81, 2015.
- [283] Q. Nie, X. Du, and M. Geng, “Small molecule inhibitors of amyloid  $\beta$  peptide aggregation as a potential therapeutic strategy for Alzheimer’s disease,” *Acta Pharmacol. Sin.*, vol. 32, no. 5, pp. 545–551, 2011.
- [284] M. E. Kennedy, A. W. Stamford, X. Chen, K. Cox, J. N. Cumming, M. F. Dockendorf, M. Egan, L. Ereshefsky, R. A. Hodgson, L. A. Hyde, S. Jhee, H. J. Kleijn, R. Kuvelkar, W. Li, B. A. Mattson, H. Mei, J. Palcza, J. D. Scott, M. Tanen, M. D. Troyer, J. L. Tseng, J. A. Stone, E. M. Parker, and M. S. Forman, “The BACE1 inhibitor verubecestat (MK-8931) reduces CNS  $\beta$ -amyloid in animal models and in Alzheimer’s disease patients,” *Sci. Transl. Med.*, vol. 8, no. 363, pp. 1–13, 2016.
- [285] J. E. Gestwicki, G. R. Crabtree, and I. A. Graef, “Harnessing chaperones to generate small-

- molecule inhibitors of amyloid beta aggregation,” *Sci. Reports*, vol. 306, no. 5697, pp. 865–869, 2004.
- [286] J. Sevigny, P. Chiao, T. Bussière, P. H. Weinreb, L. Williams, M. Maier, R. Dunstan, S. Salloway, T. Chen, Y. Ling, J. O. Gorman, F. Qian, M. Arastu, M. Li, S. Chollate, M. S. Brennan, O. Quintero-monzon, R. H. Scannevin, H. M. Arnold, T. Engber, K. Rhodes, J. Ferrero, Y. Hang, A. Mikulskis, J. Grimm, C. Hock, R. M. Nitsch, and A. Sandrock, “The antibody aducanumab reduces A $\beta$  plaques in Alzheimer’s disease,” *Nature*, vol. 537, pp. 50–56, 2016.
- [287] M. R. Mangione, S. Vilasi, C. Marino, F. Librizzi, C. Canale, D. Spigolon, F. Bucchieri, A. Fucarino, R. Passantino, F. Cappello, D. Bulone, and P. San Biagio, “Hsp60, amateur chaperone in amyloid-beta fibrillogenesis,” *Biochim. Biophys. Acta*, vol. 1860, no. 11, pp. 2474–2483, 2016.
- [288] E. Hermansson, S. Schultz, D. Crowther, S. Linse, B. Winblad, G. Westermark, J. Johansson, and J. Presto, “The chaperone domain BRICHOS prevents CNS toxicity of amyloid- $\beta$  peptide in *Drosophila melanogaster*,” *Dis. Model. Mech.*, vol. 7, no. 6, pp. 659–65, 2014.
- [289] T. Hoshino, T. Nakaya, W. Araki, K. Suzuki, T. Suzuki, and T. Mizushima, “Endoplasmic reticulum chaperones inhibit the production of amyloid- $\beta$  peptides,” *Biochem. J.*, vol. 402, pp. 581–589, 2007.
- [290] C. Månsson, P. Arosio, R. Hussein, H. H. Kampinga, R. M. Hashem, W. C. Boelens, C. M. Dobson, T. P. J. Knowles, S. Linse, and C. Emanuelsson, “Interaction of the molecular chaperone DNAJB6 with growing amyloid-beta 42 (A $\beta$ 42) aggregates leads to sub-stoichiometric inhibition of amyloid formation,” *J. Biol. Chem.*, vol. 289, no. 45, pp. 31066–31076, 2014.
- [291] M. M. M. Wilhelmus, R. M. W. De Waal, and M. M. Verbeek, “Heat shock proteins and amateur chaperones in amyloid-beta accumulation and clearance in Alzheimer’s disease,” *Mol. Neurobiol.*, vol. 35, no. 3, pp. 203–216, 2007.
- [292] S. A. Lewis and N. J. Cowan, “Purification of prefoldin,” *Methods Mol. Biol.*, vol. 140, pp. 179–184, 2000.
- [293] C. L. Anderson, C. L. Strobe, and E. N. Moriyama, “SuiteMSA: visual tools for multiple sequence alignment comparison and molecular sequence simulation,” *Biomed. Cent.*, vol. 12, pp. 184–197, 2011.
- [294] G. J. Barton and M. J. E. Sternberg, “A strategy for the rapid multiple alignment of protein sequences,” *J. Mol. Biol.*, vol. 198, no. 2, pp. 327–337, 1987.
- [295] Q. Le, F. Sievers, and D. G. Higgins, “Protein multiple sequence alignment benchmarking through secondary structure prediction,” *Bioinformatics*, vol. 33, no. 9, pp. 1331–1337, 2017.



- [296] S. B. Needleman and C. D. Wunsch, "BLAST global alignment: Needleman-Wunsch global align protein sequences," *Natl. Libr. Med.* [Online]. Available: [https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&BLAST\\_PROGRAMS=blastp&PAGE\\_TYPE=BlastSearch&BLAST\\_SPEC=GlobalAln&LINK\\_LOC=blasttab&LAST\\_PAGE=blastn&BLAST\\_INIT=GlobalAln](https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&BLAST_PROGRAMS=blastp&PAGE_TYPE=BlastSearch&BLAST_SPEC=GlobalAln&LINK_LOC=blasttab&LAST_PAGE=blastn&BLAST_INIT=GlobalAln). [Accessed: 16-Jun-2016].
- [297] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, "Basic local alignment search tool," *J. Mol. Biol.*, vol. 215, no. 3, pp. 403–410, 1990.
- [298] S. B. Needleman and C. D. Wunsch, "A general method applicable to the search for similarities in the amino acid sequence of two proteins," *J. Mol. Biol.*, vol. 48, no. 3, pp. 443–453, 1970.
- [299] H. Cheng, B. H. Kim, and N. V Grishin, "Discrimination between distant homologs and structural analogs: lessons from manually constructed, reliable data sets," *J. Mol. Biol.*, vol. 377, no. 4, pp. 1265–1278, 2008.
- [300] M. S. Izydorczyk and C. G. Biliaderis, "Gradient ammonium sulphate fractionation of galactomannans," *Food Hydrocoll.*, vol. 10, no. 3, pp. 295–300, 1996.
- [301] W. K. Choong, H. Y. Chang, C. T. Chen, C. F. Tsai, W. L. Hsu, Y. J. Chen, and T. Y. Sung, "Informatics view on the challenges of identifying missing proteins from shotgun Proteomics," *J. Proteome Res.*, vol. 14, no. 12, pp. 5396–5407, 2015.
- [302] K. Chandramouli and P. Qian, "Proteomics: challenges, techniques and possibilities to overcome biological sample complexity," *Hum. Genomics Proteomics*, pp. 1–22, 2009.
- [303] A. Jungbauer and R. Hahn, "Ion-Exchange Chromatography," in *Methods in Enzymology*, 1st ed., vol. 463, Amsterdam: Elsevier, 2009, pp. 349–371.
- [304] J. T. McCue, "Theory and use of hydrophobic interaction chromatography in protein purification applications," in *Methods in Enzymology*, 1st ed., vol. 463, no. C, Elsevier Inc., 2009, pp. 405–414.
- [305] L. J. Cummings, M. A. Snyder, and K. Brisack, "Protein chromatography on hydroxyapatite columns," in *Methods in Enzymology*, 1st ed., vol. 463, Amsterdam: Elsevier, 2009, pp. 387–404.
- [306] P. Hong, S. Koza, and E. S. P. Bouvier, "Size-exclusion chromatography for the analysis of protein biotherapeutics and their aggregates," *J. Liq. Chromatogr. Relat. Technol.*, vol. 35, no. 20, pp. 2923–2950, 2012.
- [307] K. C. Duong-Ly and S. B. Gabelli, "Gel filtration chromatography (size exclusion chromatography) of proteins," in *Methods in Enzymology*, 1st ed., vol. 541, Amsterdam: Elsevier, 2014, pp. 105–114.

- [308] R. Westermeier and R. Marouga, "Protein detection methods in proteomics research," *Biosci. Rep.*, vol. 25, no. 1–2, pp. 19–32, 2005.
- [309] C. Manoil and B. Traxler, "Insertion of in-frame sequence tags into proteins using transposons," *Methods*, vol. 20, no. 1, pp. 55–61, 2000.
- [310] R. B. Kapust, J. Tözsér, T. D. Copeland, and D. S. Waugh, "The P1' specificity of tobacco etch virus protease," *Biochem. Biophys. Res. Commun.*, vol. 294, no. 5, pp. 949–955, 2002.
- [311] J. Phan, A. Zdanov, A. G. Evdokimov, J. E. Tropea, H. K. Peters, R. B. Kapust, M. Li, A. Wlodawer, and D. S. Waugh, "Structural basis for the substrate specificity of tobacco etch virus protease," *J. Biol. Chem.*, vol. 277, no. 52, pp. 50564–50572, 2002.
- [312] M. Urh, D. Simpson, and K. Zhao, "Affinity chromatography: general methods," in *Methods in Enzymology*, 1st ed., vol. 463, Amsterdam: Elsevier, 2009, pp. 417–438.
- [313] J. A. Bornhorst and J. J. Falke, "Purification of proteins using polyhistidine affinity tags," *Methods Enzymol.*, vol. 326, pp. 245–254, 2010.
- [314] Addgene, "Analyze Sequence: pPROEX HTb," *Addgene*, 2013. [Online]. Available: [https://www.addgene.org/browse/sequence\\_vdb/3835/](https://www.addgene.org/browse/sequence_vdb/3835/). [Accessed: 14-Aug-2012].
- [315] H.-G. Zhang, J. Xie, I. Dmitriev, E. Kashentseva, D. T. Curiel, H.-C. Hsu, and J. D. Mountz, "Addition of six-his-tagged peptide to the C terminus of adeno-associated virus VP3 does not affect viral tropism or production," *J. Virol.*, vol. 76, no. 23, pp. 12023–12031, 2002.
- [316] A. Abe, K. Takahashi-Niki, Y. Takekoshi, T. Shimizu, H. Kitaura, H. Maita, S. M. M. Iguchi-Ariga, and H. Ariga, "Prefoldin plays a role as a clearance factor in preventing proteasome inhibitor-induced protein aggregation," *J. Biol. Chem.*, vol. 288, no. 39, pp. 27764–27776, 2013.
- [317] K. Irie, K. Murakami, Y. Masuda, A. Morimoto, H. Ohigashi, R. Ohashi, K. Takegoshi, M. Nagao, T. Shimizu, and T. Shirasawa, "Structure of  $\beta$ -amyloid fibrils and its relevance to their neurotoxicity: implications for the pathogenesis of Alzheimer's disease," *J. Biosci. Bioeng.*, vol. 99, no. 5, pp. 437–447, 2005.
- [318] R. W. Shin, K. Ogino, A. Kondo, T. C. Saido, J. Q. Trojanowski, T. Kitamoto, and J. Tateishi, "Amyloid  $\beta$ -protein (A $\beta$ ) 1-40 but not A $\beta$ 1-42 contributes to the experimental formation of Alzheimer disease amyloid fibrils in rat brain," *J. Neurosci. Off. J. Soc. Neurosci.*, vol. 17, no. 21, pp. 8187–8193, 1997.
- [319] M. Schmidt, C. Sachse, W. Richter, C. Xu, M. Fändrich, and N. Grigorieff, "Comparison of Alzheimer A $\beta$ (1–40) and A $\beta$ (1–42) amyloid fibrils reveals similar protofilament structures," *Proc. Natl. Acad. Sci.*, vol. 106, no. 47, pp. 19813–19818, 2009.
- [320] P. Arosio, T. P. J. Knowles, and S. Linse, "On the lag phase in amyloid fibril formation," *Phys. Chem. Chem. Phys.*, vol. 17, no. 12, pp. 7606–7618, 2015.

- [321] Y. Levites, B. O’Nuallain, R. D. Puligedda, T. Ondrejcek, S. P. Adekar, C. Chen, P. E. Cruz, A. M. Rosario, S. Macy, A. J. Mably, D. M. Walsh, R. Vidal, A. Solomon, D. Brown, M. J. Rowan, T. E. Golde, and S. K. Dessain, “A human monoclonal IgG that binds A $\beta$  assemblies and diverse amyloids exhibits anti-amyloid activities in vitro and in vivo,” *J. Neurosci.*, vol. 35, no. 16, pp. 6265–6276, 2015.
- [322] V. Valls-Comamala, B. Guivernau, J. Bonet, M. Puig, A. Perálvarez-Marín, E. Palomer, X. Fernández-Busquets, X. Altafaj, M. Tajés, A. Puig-Pijoan, R. Vicente, B. Oliva, and F. J. Muñoz, “The antigen-binding fragment of human gamma immunoglobulin prevents amyloid  $\beta$ -peptide folding into  $\beta$ -sheet to form oligomers,” *Oncotarget*, vol. 8, no. 25, pp. 41154–41165, 2017.
- [323] J. R. Ou, M. S. Tan, A. M. Xie, J. T. Yu, and L. Tan, “Heat shock protein 90 in Alzheimer’s disease,” *Biomed Res Int*, vol. 2014, pp. 1–7, 2014.
- [324] C. G. Evans, S. Wisén, and J. E. Gestwicki, “Heat shock proteins 70 and 90 inhibit early stages of amyloid  $\beta$ -(1-42) aggregation in vitro,” *J. Biol. Chem.*, vol. 281, no. 44, pp. 33182–33191, 2006.
- [325] L. B. Shelton, J. D. Baker, D. Zheng, L. E. Sullivan, P. K. Solanki, J. M. Webster, Z. Sun, J. J. Sabbagh, B. A. Nordhues, J. Koren, S. Ghosh, B. S. J. Blagg, L. J. Blair, and C. A. Dickey, “Hsp90 activator Aha1 drives production of pathological tau aggregates,” *Proc. Natl. Acad. Sci.*, vol. 2017, pp. 1–6, 2017.
- [326] Y. Porat, A. Abramowitz, and E. Gazit, “Inhibition of amyloid fibril formation by polyphenols: Structural similarity and aromatic interactions as a common inhibition mechanism,” *Chem. Biol. Drug Des.*, vol. 67, pp. 27–37, 2005.
- [327] Y. Porat, Y. Mazor, S. Efrat, and E. Gazit, “Inhibition of islet amyloid polypeptide fibril formation: A potential role for heteroaromatic interactions,” *Biochemistry*, vol. 43, no. 45, pp. 14454–14462, 2004.
- [328] R. Iizuka, Y. Sugano, N. Ide, A. Ohtaki, T. Yoshida, S. Fujiwara, T. Imanaka, and M. Yohda, “Functional characterization of recombinant prefoldin complexes from a hyperthermophilic archaeon, thermococcus sp. strain KS-1,” *J. Mol. Biol.*, vol. 377, no. 3, pp. 972–983, 2008.
- [329] A. M. Kabir, W. Uddin, A. Narayanan, K. P. Reddy, A. M. Jairajpuri, F. Sherman, and Z. Ahmad, “Functional subunits of eukaryotic chaperonin CCT/TRiC in protein folding,” *J. Amino Acids*, vol. 2011, pp. 1–16, 2011.
- [330] H. Saibil, “Chaperone machines for protein folding, unfolding and disaggregation,” *Nat. Rev. Mol. Cell Biol.*, vol. 14, no. 10, pp. 630–642, 2013.
- [331] F. U. Hartl, “Molecular chaperones in cellular protein folding,” *Nature*, vol. 381, no. 6583, pp. 571–579, 1999.

- 571–579, 1996.
- [332] T. A. Walton, C. M. Sandoval, A. C. Fowler, A. Pardi, and M. C. Sousa, “The cavity-chaperone Skp protects its substrate from aggregation but allows independent folding of substrate domains,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 106, no. 6, pp. 1772–1777, 2009.
  - [333] T. Zako, S. Banba, M. Sahlan, M. Sakono, N. Terada, M. Yohda, and M. Maeda, “Hyperthermophilic archaeal prefoldin shows refolding activity at low temperature,” *Biochem. Biophys. Res. Commun.*, vol. 391, no. 1, pp. 467–470, 2010.
  - [334] P. Laksanalamai, T. A. Whitehead, and F. T. Robb, “Minimal protein-folding systems in hyperthermophilic archaea,” *Nat. Rev. Microbiol.*, vol. 2, no. 4, pp. 315–324, 2004.
  - [335] K. R. Shockley, D. E. Ward, S. R. Chhabra, S. B. Connors, C. I. Montero, and R. M. Kelly, “Heat shock response by the hyperthermophilic archaeon *pyrococcus furiosus*,” *Appl. Environ. Microbiol.*, vol. 69, no. 4, pp. 2365–2371, 2003.
  - [336] B. B. Boonyaratanakornkit, A. J. Simpson, T. A. Whitehead, C. M. Fraser, N. M. A. El-Sayed, and D. S. Clark, “Transcriptional profiling of the hyperthermophilic methanarchaeon *Methanococcus jannaschii* in response to lethal heat and non-lethal cold shock,” *Environ. Microbiol.*, vol. 7, no. 6, pp. 789–797, 2005.
  - [337] B. Mannini and F. Chiti, “Chaperones as suppressors of protein misfolded oligomer toxicity,” *Front. Mol. Neurosci.*, vol. 10, no. 4, pp. 1–8, 2017.
  - [338] D. Cox, D. R. Whiten, J. Brown, M. H. Horrocks, R. San Gil, C. M. Dobson, D. Klenerman, A. M. van Oijen, and H. Ecroyd, “The small heat shock protein Hsp27 binds  $\alpha$ -synuclein fibrils, preventing elongation and cytotoxicity,” *J. Biol. Chem.*, vol. 1, pp. 1–22, 2018.
  - [339] R. Bakthisaran, R. Tangirala, and C. M. Rao, “Small heat shock proteins: role in cellular functions and pathology,” *Biochim. Biophys. Acta - Proteins Proteomics*, vol. 1854, no. 4, pp. 291–319, 2015.
  - [340] C. G. Glabe and R. Kaye, “Common structure and toxic function of amyloid oligomers implies a common mechanism of pathogenesis,” *Neurology*, vol. 66, no. 1, pp. 1–5, 2006.
  - [341] F. C. Dehle, H. Ecroyd, I. F. Musgrave, and J. A. Carver, “ $\alpha$ B-crystallin inhibits the cell toxicity associated with amyloid fibril formation by  $\kappa$ -casein and the amyloid-B peptide,” *Cell Stress Chaperones*, vol. 15, no. 6, pp. 1013–1026, 2010.
  - [342] Y. Wang, P. Rodríguez de Gil, Y.-H. Chen, J. D. Kromrey, E. S. Kim, T. Pham, D. Nguyen, and J. L. Romano, “Comparing the performance of approaches for testing the homogeneity of variance assumption in one-factor ANOVA models,” *Educ. Psychol. Meas.*, vol. 77, no. 2, pp. 305–329, 2017.
  - [343] M. L. Abbot and J. McKinney, “Statistics and data management,” in *Understanding and*

- Applying Research Design*, 1st ed., New York: John Wiley & Sons, 2013, pp. 327–396.
- [344] A. K. Mahalka and P. K. J. Kinnunen, "Binding of amphipathic  $\alpha$ -helical antimicrobial peptides to lipid membranes: Lessons from temporins B and L," *Biochim. Biophys. Acta*, vol. 1788, no. 8, pp. 1600–1609, 2009.
  - [345] F. Harris, S. R. Dennison, and D. A. Phoenix, "Aberrant action of amyloidogenic host defense peptides: a new paradigm to investigate neurodegenerative disorders?," *Fed. Am. Soc. Exp. Biol.*, vol. 26, no. 5, pp. 1776–1781, 2012.
  - [346] P. K. J. Kinnunen, "Amyloid formation on lipid membrane surfaces," *Open Biol. J.*, no. 2, pp. 163–175, 2009.
  - [347] H. A. Lashuel and P. T. L. Jr, "Are amyloid diseases caused by protein aggregates that mimic bacterial pore-forming toxins?," *Q. Rev. Biophys.*, vol. 39, no. 2, pp. 167–201, 2006.
  - [348] S. M. Yatin, M. Aksenova, M. Aksenov, W. R. Markesbery, T. Aulick, and D. A. Butterfield, "Temporal relations among amyloid  $\beta$ -peptide-induced free-radical oxidative stress, neuronal toxicity, and neuronal defensive responses," *J. Mol. Neurosci.*, vol. 11, no. 3, pp. 183–198, 1998.
  - [349] D. A. Butterfield, A. M. Swomley, and R. Sultana, "Amyloid  $\beta$ -peptide (1–42)-induced oxidative stress in Alzheimer disease: importance in disease pathogenesis and progression," *Antioxid. Redox Signal.*, vol. 19, no. 8, pp. 823–835, 2013.
  - [350] R. Kaye and C. Lasagna-Reeves, "Molecular mechanisms of amyloid oligomers toxicity," *J. Alzheimer's Dis.*, vol. 33, pp. 67–78, 2013.
  - [351] V. Cabra, R. Arreguin, R. Vazquez-Duhalt, and A. Farres, "Effect of temperature and pH on the secondary structure and processes of oligomerization of 19 kDa alpha-zein," *Biochim. Biophys. Acta*, vol. 1764, no. 6, pp. 1110–1118, 2006.
  - [352] J. Jandaruang, J. Siritapetawee, K. Thumanu, C. Songsiriritthigul, C. Krittanai, S. Daduang, A. Dhiravisit, and S. Thammasirirak, "The effects of temperature and pH on secondary structure and antioxidant activity of *Crocodylus siamensis* hemoglobin," *Protein J.*, vol. 31, no. 1, pp. 43–50, 2011.
  - [353] R. Busch, Z. Reich, D. M. Zaller, V. Sloan, and E. D. Mellins, "Secondary structure composition and pH-dependent conformational changes of soluble recombinant HLA-DM," *J. Biol. Chem.*, vol. 273, no. 42, pp. 27557–27564, 1998.
  - [354] M. R. Eftink, "Use of multiple spectroscopic methods to monitor equilibrium unfolding of proteins," *Methods Enzymol.*, vol. 259, no. 1989, pp. 487–512, 1995.
  - [355] M. M. Long and D. W. Urry, "Absorption and circular dichroism spectroscopies," in *Molecular Biology*, vol. 31, 1981, pp. 143–171.

- [356] T. Rao, G. Ruiz-Gómez, T. A. Hill, H. N. Hoang, D. P. Fairlie, and J. M. Mason, "Truncated and helix-constrained peptides with high affinity and specificity for the cFos coiled-coil of AP-1," *PLoS One*, vol. 8, no. 3, pp. 1–12, 2013.
- [357] M. G. Wuo, A. B. Mahon, and P. S. Arora, "An effective strategy for stabilizing minimal coiled coil mimetics," *J. Am. Chem. Soc.*, vol. 137, no. 36, pp. 11618–11621, 2015.
- [358] D. Wang, K. Chen, J. L. Kulp III, and P. S. Arora, "Evaluation of biologically relevant short  $\alpha$ -helices stabilized by a main-chain hydrogen-bond surrogate," *J. Am. Chem. Soc.*, vol. 128, no. 28, pp. 9248–9256, 2006.
- [359] J. G. Lees, A. J. Miles, F. Wien, and B. A. Wallace, "A reference database for circular dichroism spectroscopy covering fold and secondary structure space," *Bioinformatics*, vol. 22, no. 16, pp. 1955–1962, 2006.
- [360] A. J. Miles, F. Wien, J. G. Lees, and B. A. Wallace, "Calibration and standardisation of synchrotron radiation and conventional circular dichroism spectrometers. Part 2: factors affecting magnitude and wavelength," *Spectroscopy*, vol. 19, no. 1, pp. 43–51, 2005.
- [361] Y. Feng, A. Yingge, Q. Dezhi, Y. Lin, and S. Lan, "Spectroscopic study on the effect of crystallization of the hydroxyapatite on the secondary structure of bovine serum albumin," *J-Global*, vol. 27, no. 2, pp. 321–324, 2007.
- [362] N. El Kadi, N. Taulier, J. Y. Le Huérou, M. Gindre, W. Urbach, I. Nwigwe, P. C. Kahn, and M. Waks, "Unfolding and refolding of bovine serum albumin at acid pH: Ultrasound and structural studies," *Biophys. J.*, vol. 91, no. 9, pp. 3397–3404, 2006.
- [363] Y. Moriyama, E. Watanabe, K. Kobayashi, H. Harano, E. Inui, and K. Takeda, "Secondary structural change of bovine serum albumin in thermal denaturation up to 130 °C and protective effect of sodium dodecyl sulfate on the change," *J. Phys. Chem. B*, vol. 112, no. 51, pp. 16585–16589, 2008.
- [364] N. V Di Russo, D. A. Estrin, M. A. Martí, and A. E. Roitberg, "pH-dependent conformational changes in proteins and their effect on experimental pK<sub>a</sub>s: the Case of nitrophorin 4," *PLoS Comput. Biol.*, vol. 8, no. 11, pp. 1–9, 2012.
- [365] P. S. Santiago, F. A. O. Carvalho, M. M. Domingues, J. W. P. Carvalho, N. C. Santos, and M. Tabak, "Isoelectric point determination for *Glossoscolex paulistus* extracellular hemoglobin: oligomeric stability in acidic pH and relevance to protein-surfactant interactions," *Am. Chem. Soc. Langmuir*, vol. 26, no. 12, pp. 9794–9801, 2010.
- [366] T. Shimizu, T. Mori, M. Tomita, and K. Tsumoto, "pH switching that crosses over the isoelectric point (pI) can improve the entrapment of proteins within giant liposomes by enhancing protein-membrane interaction," *Am. Chem. Soc. Langmuir*, vol. 30, no. 2, pp. 554–

- 563, 2014.
- [367] M. E. Holtzer and A. Holtzer, "α-helix to random coil transitions: determination of peptide concentration from the CD at the isodichroic point," *Biopolymers*, vol. 32, no. 12, pp. 1675–1677, 1992.
  - [368] V. Consalvi, R. Chiaraluce, L. Giangiacomo, R. Scandurra, P. Christova, A. Karshikoff, S. Knapp, and R. Ladenstein, "Thermal unfolding and conformational stability of the recombinant domain II of glutamate dehydrogenase from the hyperthermophile *Thermotoga maritima*," *Protein Eng.*, vol. 13, no. 7, pp. 501–507, 2000.
  - [369] G. T. Dolphin and L. Baltzer, "The pH-dependent tertiary structure of a designed helix-loop-helix dimer," *Fold. Des.*, vol. 2, no. 5, pp. 319–330, 1997.
  - [370] A. Micsonai, F. Wien, L. Kernya, Y.-H. Lee, Y. Goto, M. Réfrégiers, and J. Kardos, "Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 112, no. 24, pp. 3095–3103, 2015.
  - [371] S. Vivekanandan, J. R. Brender, S. Y. Lee, and R. Ayyalusamy, "A partially folded structure of amyloid-beta(1-40) in an aqueous environment," *Biochem. Biophys. Res. Commun.*, vol. 411, no. 2, pp. 312–316, 2011.
  - [372] C. Bleiholder, N. F. Dupuis, T. Wyttenbach, and M. T. Bowers, "Ion mobility-mass spectrometry reveals a conformational conversion from random assembly to beta-sheet in amyloid fibril formation," *Nat. Chem.*, vol. 3, no. 2, pp. 172–177, 2011.
  - [373] W. Hoyer, C. Gronwall, A. Jonsson, S. Stahl, and T. Hard, "Stabilization of a β-hairpin in monomeric Alzheimer's amyloid-β peptide inhibits amyloid formation," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 105, no. 13, pp. 5099–5104, 2008.
  - [374] E. Sitkiewicz, M. Kłoniecki, J. Poznański, W. Bal, and M. Dadlez, "Factors influencing compact-extended structure equilibrium in oligomers of αβ1-40 peptide—an ion mobility mass spectrometry study," *J. Mol. Biol.*, vol. 426, no. 15, pp. 2871–2885, 2014.
  - [375] T. Härd, "Protein engineering to stabilize soluble amyloid β-protein aggregates for structural and functional studies," *Fed. Eur. Biochem. Soc. J.*, vol. 278, no. 20, pp. 3884–3892, 2011.
  - [376] K. A. DaSilva, J. E. Shaw, and J. McLaurin, "Amyloid-β fibrillogenesis: structural insight and therapeutic intervention," *Exp. Neurol.*, vol. 223, no. 2, pp. 311–321, 2010.
  - [377] M. Ahmed, J. Davis, D. Aucoin, T. Sato, S. Ahuja, S. Aimoto, J. I. Elliott, W. E. Van Nostrand, and S. O. Smith, "Structural conversion of neurotoxic amyloid-β(1-42) oligomers to fibrils," *Nat. Struct. Mol. Biol.*, vol. 17, no. 5, pp. 561–567, 2010.
  - [378] K. Berthelot, C. Cullin, and S. Lecomte, "What does make an amyloid toxic: morphology, structure or interaction with membrane?," *Biochimie*, vol. 95, no. 1, pp. 12–19, 2013.

- [379] O. W. Howarth and D. M. J. Lilley, "Carbon-13-NMR of peptides and proteins," *Prog. Nucl. Magn. Reson. Spectrosc.*, vol. 12, no. 1, pp. 1–40, 1978.
- [380] D. M. LeMaster, "Chiral  $\beta$  and random fractional deuteration for the determination of protein sidechain conformation by NMR," *Fed. Eur. Biochem. Soc. Lett.*, vol. 223, no. 1, pp. 191–196, 1987.
- [381] M. Kainosho, T. Torizawa, Y. Iwashita, T. Terauchi, A. M. Ono, and P. Güntert, "Optimal isotope labelling for NMR protein structure determinations," *Nature*, vol. 440, no. 7080, pp. 52–57, 2006.
- [382] F. W. Muskett, "Sample preparation, data collection and processing," in *Protein NMR Spectroscopy: Practical Techniques and Applications*, 1st ed., L. Y. Lian and G. Roberts, Eds. New York: John Wiley & Sons, 2011, pp. 1–21.
- [383] A. E. Kelly, H. D. Ou, R. Withers, and V. Dötsch, "Low-conductivity buffers for high-sensitivity NMR measurements," *J. Am. Chem. Soc.*, vol. 124, no. 40, pp. 12013–12019, 2002.
- [384] L. C. Robosky, M. D. Reily, and D. Avizonis, "Improving NMR sensitivity by use of salt-tolerant cryogenically cooled probes," *Anal. Bioanal. Chem.*, vol. 387, no. 2, pp. 529–532, 2007.
- [385] A. K. Downing, *Protein NMR Techniques*, 2nd ed., vol. 278. New Jersey: Humana, 2004.
- [386] D. P. Frueh, "Practical aspects of NMR signal assignment in larger and challenging proteins," *Prog. Nucl. Magn. Reson. Spectrosc.*, vol. 78, pp. 47–75, 2014.
- [387] T. T. Nakashima, R. E. McClung, and G. Kotovych, "Information from the water stripe in TOCSY experiments on systems with exchangeable protons," *J. Magn. Reson.*, vol. 133, pp. 222–226, 1998.
- [388] M. Nilges, M. J. Macias, S. I. O'Donoghue, and H. Oschkinat, "Automated NOESY interpretation with ambiguous distance restraints: the refined NMR solution structure of the pleckstrin homology domain from beta-spectrin," *J. Mol. Biol.*, vol. 269, no. 3, pp. 408–422, 1997.
- [389] M. Y. Xiao, T. C. Pochapsky, and S. S. Pochapsky, " $^1\text{H}$  NMR sequential assignments and identification of secondary structural elements in oxidized putidaredoxin, an electron-transfer protein from *Pseudomonas*," *Biochemistry*, vol. 31, no. 7, pp. 1961–1968, 1992.
- [390] E. Kurimoto, Y. Nishi, Y. Yamaguchi, T. Zako, R. Iizuka, N. Ide, M. Yohda, and K. Kato, "Dynamics of group II chaperonin and prefoldin probed by  $^{13}\text{C}$  NMR spectroscopy," *Proteins*, vol. 70, no. 2, pp. 1257–1263, 2007.
- [391] A. H. Kwan, M. Mobli, P. R. Gooley, G. F. King, and J. P. MacKay, "Macromolecular NMR spectroscopy for the non-spectroscopist," *Fed. Eur. Biochem. Soc. J.*, vol. 278, no. 5, pp. 687–703, 2011.



- [392] P. K. Mandal and A. Majumdar, "A comprehensive discussion of HSQC and HMQC pulse sequences," *Concepts Magn. Reson. Part A Bridg. Educ. Res.*, vol. 20, no. 1, pp. 1–23, 2004.
- [393] W. Hu, L. T. Kakalis, L. Jiang, F. Jiang, X. Ye, and a Majumdar, "3D HCCH-COSY-TOCSY experiment for the assignment of ribose and amino acid side chains in <sup>13</sup>C labeled RNA and protein.," *J. Biomol. NMR*, vol. 12, no. 4, pp. 559–64, 1998.
- [394] J. Cavanagh, W. J. Fairbrother, A. G. Palmer, M. Rance, and N. J. Skelton, *Protein NMR Spectroscopy*, 2nd ed. Burlington: Elsevier, 2007.
- [395] C. H. U. K. Liew, R. Gamsjaeger, R. E. Mansfield, and J. P. Mackay, "NMR spectroscopy as a tool for the rapid assessment of the conformation of GST-fusion proteins," *Protein Sci.*, vol. 17, no. 9, pp. 1630–1635, 2008.
- [396] A. A. Yee, A. Savchenko, A. Ignachenko, J. Lukin, X. Xu, T. Skarina, E. Evdokimova, C. S. Liu, A. Semesi, V. Guido, A. M. Edwards, and C. H. Arrowsmith, "NMR and X-ray crystallography, complementary tools in structural proteomics of small proteins," *J. Am. Chem. Soc.*, vol. 127, no. 47, pp. 16512–16517, 2005.
- [397] H. Kida, Y. Sugano, R. Iizuka, M. Fujihashi, M. Yohda, and K. Miki, "Structural and molecular characterization of the prefoldin  $\beta$  subunit from thermococcus strain KS-1," *J. Mol. Biol.*, vol. 383, no. 3, pp. 465–474, 2008.
- [398] R. Baumgartner, C. Fernandez-Catalan, a Winoto, R. Huber, R. a Engh, and T. a Holak, "Structure of human cyclin-dependent kinase inhibitor p19INK4d: comparison to known ankyrin-repeat-containing structures and implications for the dysfunction of tumor suppressor p16INK4a.," *Structure*, vol. 6, no. 10, pp. 1279–1290, 1998.
- [399] W. Kalus, R. Baumgartner, C. Renner, A. Noegel, F. K. M. Chan, A. Winoto, and T. A. Holak, "NMR structural characterization of the CDK inhibitor p19INK4d," *Fed. Eur. Biochem. Soc. Lett.*, vol. 401, no. 2–3, pp. 127–132, 1997.
- [400] D. S. Wishart, C. G. Bigam, J. Yao, F. Abildgaard, H. J. Dyson, E. Oldfield, J. L. Markley, and B. D. Sykes, "<sup>1</sup>H, <sup>13</sup>C and <sup>15</sup>N chemical shift referencing in biomolecular NMR," *J. Biomol. NMR*, vol. 6, no. 2, pp. 135–140, 1995.
- [401] S. Grzesiek and A. Bax, "Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR," *J. Am. Chem. Soc.*, vol. 114, no. 16, pp. 6291–6293, 1992.
- [402] B. K. John, D. Plant, P. Webb, and R. E. Hurd, "Effective combination of gradients and crafted RF pulses for water suppression in biological samples," *J. Magn. Reson.*, vol. 98, no. 1, pp. 200–206, 1992.
- [403] B. Ganem, Y. T. Li, and J. D. Henion, "Detection of noncovalent receptor-ligand complexes by

- mass spectrometry," *J. Am. Chem. Soc.*, vol. 113, no. 16, pp. 6294–6296, 1991.
- [404] A. J. R. Heck, "Native mass spectrometry: a bridge between interactomics and structural biology," *Nat. Methods*, vol. 5, no. 11, pp. 927–933, 2008.
- [405] I. D. G. Campuzano, H. Li, D. Bagal, J. L. Lippens, J. Svitel, R. J. M. Kurzeja, H. Xu, P. D. Schnier, and J. A. Loo, "Native MS analysis of bacteriorhodopsin and an empty nanodisc by orthogonal acceleration time-of-flight, orbitrap and ion cyclotron resonance," *Anal. Chem.*, vol. 88, no. 24, pp. 12427–12436, 2016.
- [406] E. B. Erba and C. Petosa, "The emerging role of native mass spectrometry in characterizing the structure and dynamics of macromolecular complexes," *Protein Sci.*, vol. 24, no. 8, pp. 1176–1192, 2015.
- [407] F. Lermite and F. Sobott, "Electron transfer dissociation provides higher-order structural information of native and partially unfolded protein complexes," *Proteomics*, vol. 15, no. 16, pp. 2813–2822, 2015.
- [408] S. H. Chen and D. H. Russell, "How closely related are conformations of protein ions sampled by IM-MS to native solution structures?," *J. Am. Soc. Mass Spectrom.*, vol. 26, no. 9, pp. 1433–1443, 2015.
- [409] M. F. Bush, Z. Hall, K. Giles, J. Hoyes, C. V. Robinson, and B. T. Ruotolo, "Collision cross sections of proteins and their complexes: a calibration framework and database for gas-phase structural biology," *Anal. Chem.*, vol. 82, no. 22, pp. 9557–9565, 2010.
- [410] Y. Song, M. T. Nelp, V. Bandarian, and V. H. Wysocki, "Refining the structural model of a heterohexameric protein complex: surface induced dissociation and ion mobility provide key connectivity and topology information," *ACS Cent. Sci.*, vol. 1, no. 9, pp. 477–487, 2015.
- [411] N. F. Zinnel, P. J. Pai, and D. H. Russell, "Ion mobility-mass spectrometry (IM-MS) for top-down proteomics: increased dynamic range affords increased sequence coverage," *Anal. Chem.*, vol. 84, no. 7, pp. 3390–3397, 2012.
- [412] L. Konermann and D. J. Douglas, "Unfolding of proteins monitored by electrospray ionization mass spectrometry: A comparison of positive and negative ion modes," *J. Am. Soc. Mass Spectrom.*, vol. 9, no. 12, pp. 1248–1254, 1998.
- [413] H. Zhang, W. Cui, J. Wen, R. E. Blankenship, and M. L. Gross, "Native electrospray and electron-capture dissociation FTICR mass spectrometry for top-down studies of protein assemblies," *Anal. Chem.*, vol. 83, no. 14, pp. 5598–5606, 2011.
- [414] M. Zhou, C. M. Jones, and V. H. Wysocki, "Dissecting the large noncovalent protein complex GroEL with surface-induced dissociation and ion mobility-mass spectrometry," *Anal. Chem.*, vol. 85, no. 17, pp. 8262–8267, 2013.

- [415] A. Konijnenberg, L. Bannwarth, D. Yilmaz, A. Koçer, C. Venien-Bryan, and F. Sobott, "Top-down mass spectrometry of intact membrane protein complexes reveals oligomeric state and sequence information in a single experiment," *Protein Sci.*, vol. 24, no. 8, pp. 1292–1300, 2015.
- [416] J. F. Brandts and L. Hunt, "The thermodynamics of protein denaturation. III. The denaturation of ribonuclease in water and in aqueous urea and aqueous ethanol mixtures," *J. Am. Chem. Soc.*, vol. 89, no. 19, pp. 4826–4838, 1967.
- [417] S. C. Li, N. K. Goto, K. A. Williams, and C. M. Deber, " $\alpha$ -helical, but not  $\beta$ -sheet, propensity of proline is determined by peptide environment," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 93, no. 6, pp. 6676–6681, 1996.
- [418] H. Hernandez and C. V Robinson, "Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry," *Nat. Protoc.*, vol. 2, no. 3, pp. 715–726, 2007.
- [419] P. Tompa, P. Bánki, M. Bokor, P. Kamasa, D. Kovács, G. Lasanda, and K. Tompa, "Protein-water and protein-buffer interactions in the aqueous solution of an intrinsically unstructured plant dehydrin: NMR intensity and DSC aspects," *Biophys. J.*, vol. 91, no. 6, pp. 2243–2249, 2006.
- [420] J. A. Leary, M. R. Schenauer, R. Stefanescu, A. Andaya, B. T. Ruotolo, C. V Robinson, K. Thalassinou, J. H. Scrivens, M. Sokabe, and J. W. B. Hershey, "Methodology for measuring conformation of solvent-disrupted protein subunits using T-WAVE ion mobility MS: an investigation into eukaryotic initiation factors," *J. Am. Soc. Mass Spectrom.*, vol. 20, no. 9, pp. 1699–1706, 2009.
- [421] Z. Hall, A. Politis, and C. V Robinson, "Structural modeling of heteromeric protein complexes from disassembly pathways and ion mobility-mass spectrometry," *Cell Press Struct.*, vol. 20, no. 9, pp. 1596–1609, 2012.
- [422] T. L. Pukala, "Mass spectrometry for structural biology: determining the composition and architecture of protein complexes," *CSIRO Publ.*, vol. 64, pp. 681–691, 2011.
- [423] J. L. P. Benesch, F. Sobott, and C. V Robinson, "Thermal dissociation of multimeric protein complexes by using nanoelectrospray mass spectrometry," *Anal. Chem.*, vol. 75, no. 10, pp. 2208–2214, 2003.
- [424] C. Wang, R. Xu, W. Tian, X. Jiang, Z. Cui, M. Wang, H. Sun, K. Fang, and N. Gu, "Determining intracellular temperature at single-cell level by a novel thermocouple method," *Cell Res.*, vol. 21, no. 10, pp. 1517–1519, 2011.
- [425] G. R. Hilton and J. L. P. Benesch, "Two decades of studying non-covalent biomolecular

- assemblies by means of electrospray ionization mass spectrometry," *J. R. Soc. Interface*, vol. 9, no. 70, pp. 801–816, 2012.
- [426] B. T. Ruotolo, J. L. P. Benesch, A. M. Sandercock, S.-J. Hyung, and C. V Robinson, "Ion mobility-mass spectrometry analysis of large protein complexes," *Nat. Protoc.*, vol. 3, no. 7, pp. 1139–1152, 2008.
- [427] M. F. Bush, I. D. G. Campuzano, and C. V Robinson, "Ion mobility mass spectrometry of peptide ions: effects of drift gas and calibration strategies," *Anal. Chem.*, vol. 84, no. 16, pp. 7124–7130, 2012.
- [428] A. Marchand, S. Livet, F. Rosu, and V. Gabelica, "Drift tube ion mobility: how to reconstruct collision cross section distributions from arrival time distributions?," *Anal. Chem.*, vol. 89, no. 23, pp. 12674–12681, 2017.
- [429] M. Farenc, B. Paupy, S. Marceau, E. Riches, C. Afonso, and P. Giusti, "Effective ion mobility peak width as a new isomeric descriptor for the untargeted analysis of complex mixtures using ion mobility-mass spectrometry," *J. Am. Soc. Mass Spectrom.*, vol. 28, no. 11, pp. 2476–2482, 2017.
- [430] S. J. Valentine and D. E. Clemmer, "Treatise on the measurement of molecular masses with ion mobility spectrometry," *Anal. Chem.*, vol. 81, no. 14, pp. 1–12, 2009.
- [431] M. J. Manard and P. R. Kemper, "Ion mobility mass spectrometry: the design of a new high-resolution ion mobility instrument with applications toward electronic-state characterization of first-row transition metal cations," *Int. J. Mass Spectrom.*, vol. 402, pp. 1–11, 2016.
- [432] J. L. P. Benesch and B. T. Ruotolo, "Mass spectrometry: Come of age for structural and dynamical biology," *Curr. Opin. Struct. Biol.*, vol. 21, no. 5, pp. 641–649, 2011.
- [433] Z. Hall, C. Schmidt, and A. Politis, "Uncovering the early assembly mechanism for amyloidogenic  $\beta$ 2-microglobulin using cross-linking and native mass spectrometry," *J. Biol. Chem.*, vol. 291, no. 9, pp. 4626–4637, 2016.
- [434] S. Sandin, L. G. Öfverstedt, A. C. Wikström, Ö. Wrange, and U. Skoglund, "Structure and flexibility of individual immunoglobulin G molecules in solution," *Structure*, vol. 12, no. 3, pp. 409–415, 2004.
- [435] K. J. Pacholarz, M. Porrini, R. A. Garlish, R. J. Burnley, R. J. Taylor, A. J. Henry, and P. E. Barran, "Dynamics of intact immunoglobulin G explored by drift-tube ion-mobility mass spectrometry and molecular modeling," *Angew. Chemie - Int. Ed.*, vol. 53, no. 30, pp. 7765–7769, 2014.
- [436] K. Leitzgen, M. R. Knittler, and I. G. Haas, "Assembly of immunoglobulin light chains as a prerequisite for secretion," *J. Biol. Chem.*, vol. 272, no. 5, pp. 3117–3123, 1997.
- [437] A. W. P. Vermeer and W. Norde, "The thermal stability of immunoglobulin: unfolding and

- aggregation of a multi-domain protein," *Biophys. J.*, vol. 78, no. 1, pp. 394–404, 2000.
- [438] K. J. Pacholarz, S. J. Peters, R. A. Garlish, A. J. Henry, R. J. Taylor, D. P. Humphreys, and P. E. Barran, "Molecular insights into the thermal stability of mAbs with variable-temperature ion-mobility mass spectrometry," *ChemBioChem*, vol. 17, no. 1, pp. 46–51, 2016.
- [439] J. L. P. Benesch, "Collisional activation of protein complexes: picking up the pieces," *J. Am. Soc. Mass Spectrom.*, vol. 20, no. 3, pp. 341–348, 2009.
- [440] C. M. Jones, "Characterization of macromolecular protein assemblies by collisional-induced and surface-induced dissociation: expanding the role of mass spectrometry in structural biology," The University of Arizona, 2008.
- [441] S. A. McLuckey, "Principles of collisional activation in analytical mass spectrometry," *J. Am. Soc. Mass Spectrom.*, vol. 3, no. 6, pp. 599–614, 1992.
- [442] C. K. Siu, Y. Guo, I. S. Saminathan, A. C. Hopkinson, and K. W. M. Siu, "Optimization of parameters used in algorithms of ion-mobility calculation for conformational analyses," *J. Phys. Chem. B*, vol. 114, no. 2, pp. 1204–1212, 2010.
- [443] J. T. S. Hopper and N. J. Oldham, "Collision induced unfolding of protein ions in the gas phase studied by ion mobility-mass spectrometry: the effect of ligand binding on conformational stability," *J. Am. Soc. Mass Spectrom.*, vol. 20, no. 10, pp. 1851–1858, 2009.
- [444] Y. Sun, S. Vahidi, M. A. Sowole, and L. Konermann, "Protein structural studies by traveling wave ion mobility spectrometry: A critical look at electrospray sources and calibration issues," *J. Am. Soc. Mass Spectrom.*, vol. 27, no. 1, pp. 31–40, 2016.
- [445] Y. Tian, L. Han, A. C. Buckner, and B. T. Ruotolo, "Collision induced unfolding of intact antibodies: rapid characterization of disulfide bonding patterns, glycosylation, and structures," *Anal. Chem.*, vol. 87, no. 22, pp. 11509–11515, 2015.
- [446] Y. Zhong, S.-J. Hyung, and B. T. Ruotolo, "Ion mobility–mass spectrometry for structural proteomics," *Expert Rev. Proteomics*, vol. 9, no. 1, pp. 47–58, 2012.
- [447] E. G. Marklund, M. T. Degiacomi, C. V Robinson, A. J. Baldwin, and J. L. P. Benesch, "Collision cross sections for structural proteomics," *Structure*, vol. 23, no. 4, pp. 791–799, 2015.
- [448] J. Chaloupka, K. R. Pillai, and M. Strnadova, "The complex effect of temperature on protein degradation in sporulating bacillus megaterium," *Fed. Eur. Microbiol. Soc. Lett.*, vol. 9, pp. 107–110, 1980.
- [449] R. Jaenicke, "Protein stability and molecular adaptation to extreme conditons," *Eur. J. Biochem.*, vol. 202, no. 3, pp. 715–728, 1991.
- [450] M. Wilm and M. Mann, "Analytical properties of the nanoelectrospray ion source," *Anal. Chem.*, vol. 68, no. 1, pp. 1–8, 1996.

- [451] K. C. Giese and E. Vierling, "Changes in oligomerization are essential for the chaperone activity of a small heat shock protein in vivo and in vitro," *J. Biol. Chem.*, vol. 277, no. 48, pp. 46310–46318, 2002.
- [452] M. Haslbeck, S. Walke, T. Stromer, M. Ehrnsperger, H. E. White, S. Chen, H. R. Saibil, and J. Buchner, "Hsp26: a temperature-regulated chaperone," *Eur. Mol. Biol. Organ. J.*, vol. 18, no. 23, pp. 6744–6751, 1999.
- [453] F. Narberhaus, "α-crystallin-type heat shock proteins: socializing minichaperones in the context of a multichaperone network," *Microbiol. Mol. Biol. Rev.*, vol. 66, no. 1, pp. 64–93, 2002.
- [454] D. E. Clemmer, "Cross section database," *Indiana University*, 2017. .
- [455] M. M. Young, N. Tang, J. C. Hempel, C. M. Oshiro, E. W. Taylor, I. D. Kuntz, B. W. Gibson, and G. Dollinger, "High throughput protein fold identification by using experimental constraints derived from intramolecular cross-links and mass spectrometry," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 97, no. 11, pp. 5802–5806, 2000.
- [456] A. Leitner, T. Walzthoeni, A. Kahraman, F. Herzog, O. Rinner, M. Beck, and R. Aebersold, "Probing native protein structures by chemical cross-linking, mass spectrometry, and bioinformatics," *Mol. Cell. Proteomics*, vol. 9, no. 8, pp. 1634–1649, 2010.
- [457] C. Arlt, C. H. Ihling, and A. Sinz, "Structure of full-length p53 tumor suppressor probed by chemical cross-linking and mass spectrometry," *Proteomics*, vol. 15, no. 16, pp. 2746–2755, 2015.
- [458] D. Mouradov, G. King, I. L. Ross, J. K. Forwood, D. A. Hume, A. Sinz, J. L. Martin, B. Kobe, and T. Huber, "Protein structure determination using a combination of cross-linking, mass spectrometry, and molecular modeling," *Methods Mol. Biol.*, vol. 426, pp. 459–474, 2008.
- [459] C. Schmidt and C. V Robinson, "A comparative cross-linking strategy to probe conformational changes in protein complexes," *Nat. Protoc.*, vol. 9, no. 9, pp. 2224–2236, 2014.
- [460] A. Sinz, "Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein-protein interactions," *Mass Spectrom. Rev.*, vol. 25, no. 4, pp. 663–682, 2006.
- [461] A. Sinz, C. Arlt, D. Chorev, and M. Sharon, "Chemical cross-linking and native mass spectrometry: a fruitful combination for structural biology," *Protein Sci.*, vol. 24, no. 8, pp. 1193–1209, 2015.
- [462] A. N. Calabrese, T. Wang, J. H. Bowie, and T. L. Pukala, "Negative ion fragmentations of disulfide-containing crosslinking reagents are competitive with aspartic acid side-chain-induced cleavages," *Rapid Commun. Mass Spectrom.*, vol. 27, no. 1, pp. 238–248, 2013.

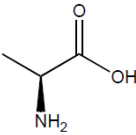
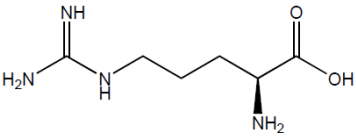
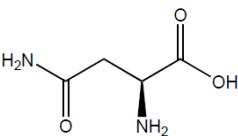
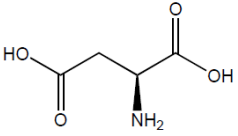
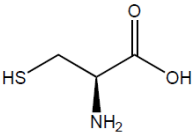
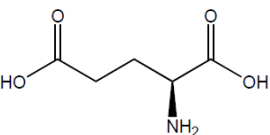
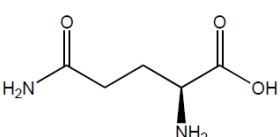
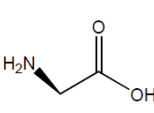
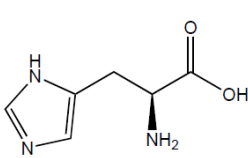
- [463] A. N. Calabrese, N. J. Good, T. Wang, J. He, J. H. Bowie, and T. L. Pukala, "A negative ion mass spectrometry approach to identify cross-linked peptides utilizing characteristic disulfide fragmentations," *J. Am. Soc. Mass Spectrom.*, vol. 23, no. 8, pp. 1364–1375, 2012.
- [464] M. W. Gardner, L. A. Vasicek, S. Shabbir, E. V. Anslyn, and J. S. Brodbelt, "Chromogenic cross-linker for the characterization of protein structure by infrared multiphoton dissociation mass spectrometry," *Anal. Chem.*, vol. 80, no. 13, pp. 4807–4819, 2008.
- [465] G. J. King, A. Jones, B. Kobe, T. Huber, D. Mouradov, D. A. Hume, and I. L. Ross, "Identification of disulfide-containing chemical cross-links in proteins using MALDI-TOF/TOF-mass spectrometry," *Anal. Chem.*, vol. 80, no. 13, pp. 5036–5043, 2008.
- [466] Y. Lu, M. Tanasova, B. Borhan, and G. E. Reid, "Ionic reagent for controlling the gas-phase fragmentation reactions of cross-linked peptides," *Anal. Chem.*, vol. 80, no. 23, pp. 9279–9287, 2008.
- [467] B. Schilling, R. H. Row, B. W. Gibson, X. Guo, and M. M. Young, "MS2Assign, automated assignment and nomenclature of tandem mass spectra of chemically crosslinked peptides," *J. Am. Soc. Mass Spectrom.*, vol. 14, no. 8, pp. 834–850, 2003.
- [468] A. Leitner, T. Walzthoeni, and R. Aebersold, "Lysine-specific chemical cross-linking of protein complexes and identification of cross-linking sites using LC-MS/MS and the xQuest/xProphet software pipeline," *Nat. Protoc.*, vol. 9, no. 1, pp. 120–137, 2014.
- [469] M. Götze, J. Pettelkau, S. Schaks, K. Bosse, C. H. Ihling, F. Krauth, R. Fritzsche, U. Kühn, and A. Sinz, "StavroX- a software for analyzing crosslinked products in protein interaction studies," *J. Am. Soc. Mass Spectrom.*, vol. 23, no. 1, pp. 76–87, 2012.
- [470] M. Grimm, T. Zimniak, A. Kahraman, and F. Herzog, "XVis: A web server for the schematic visualization and interpretation of crosslink-derived spatial restraints," *Nucleic Acids Res.*, vol. 43, no. W1, pp. 362–369, 2015.
- [471] S. P. Gaucher, M. Z. Hadi, and M. M. Young, "Influence of crosslinker identity and position on gas-phase dissociation of lys-lys crosslinked peptides," *J. Am. Soc. Mass Spectrom.*, vol. 17, no. 3, pp. 395–405, 2006.
- [472] R. J. Waugh, J. H. Bowie, and R. N. Hayes, "Collision-induced dissociations of deprotonated peptides. Dipeptides containing aspartic or glutamic acids," *Org. Mass Spectrom.*, vol. 26, no. 10, pp. 250–256, 1991.
- [473] D. Bilusich and J. H. Bowie, "Fragmentations of (M-H)<sup>-</sup> anions of underivatized peptides. Part 2: characteristic cleavages of Ser and Cys and of disulfides and other post-translational modifications, together with some unusual internal processes," *Mass Spectrom. Rev.*, vol. 28, pp. 20–34, 2009.

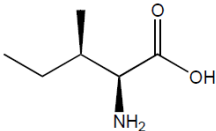
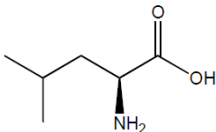
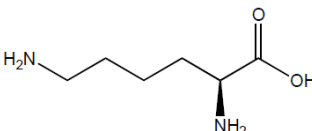
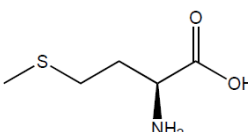
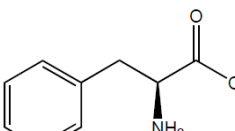
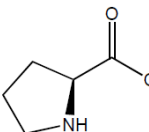
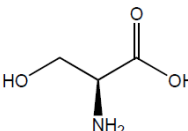
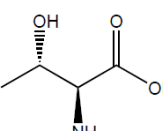
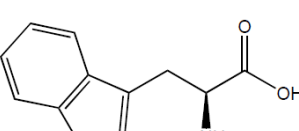
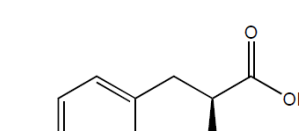
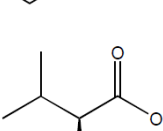
- [474] S. T. Steinborner and J. H. Bowie, "The negative ion mass spectra of [M-H]<sup>-</sup> ions derived from caeridin and dynastin peptides. Internal backbone cleavages directed through Asp and Asn residues," *Rapid Commun. Mass Spectrom.*, vol. 11, no. 3, pp. 253–258, 1997.
- [475] C. S. Brinkworth, S. Dua, A. M. McAnoy, and J. H. Bowie, "Negative ion fragmentations of deprotonated peptides: Backbone cleavages directed through both Asp and Glu," *Rapid Commun. Mass Spectrom.*, vol. 15, no. 20, pp. 1965–1973, 2001.
- [476] G. H. Kruppa, J. Schoeniger, and M. M. Young, "A top down approach to protein structural studies using chemical cross-linking and fourier transform mass spectrometry," *Rapid Commun. Mass Spectrom.*, vol. 17, no. 2, pp. 155–162, 2003.
- [477] C. S. Brinkworth, J. A. Carver, K. L. Wegener, J. Doyle, L. E. Llewellyn, and J. H. Bowie, "The solution structure of frenatin 3, a neuronal nitric oxide synthase inhibitor from the giant tree frog, *Litoria infrafronata*," *Biopolymers*, vol. 70, no. 3, pp. 424–434, 2003.
- [478] M. D. Leavell, P. Novak, C. R. Behrens, J. S. Schoeniger, and G. H. Kruppa, "Strategy for selective chemical cross-linking of tyrosine and lysine residues," *J. Am. Soc. Mass Spectrom.*, vol. 15, no. 11, pp. 1604–1611, 2004.
- [479] J. Luo, S. K. T. S. Wärmländer, A. Gräslund, and J. P. Abrahams, "Cross interactions between the Alzheimer's disease amyloid- $\beta$  peptide and other amyloid proteins: a further aspect of the amyloid cascade hypothesis," *J. Biol. Chem.*, vol. 292, no. 5, pp. 2028–2046, 2016.
- [480] J. Luo, S. K. T. S. Wärmländer, A. Gräslund, and J. P. Abrahams, "Human lysozyme inhibits the in vitro aggregation of A $\beta$  peptides, which in vivo are associated with Alzheimer's disease," *Chem. Commun.*, vol. 49, pp. 6507–6509, 2013.



# Appendix

## Appendix A: The 20 Common Amino Acid Residues

Amino Acid	Structure	Residue Integral Mass
Alanine (Ala/A)		71
Arginine (Arg/R)		156
Asparagine (Asn/N)		114
Aspartic acid (Asp/D)		115
Cysteine (Cys/C)		103
Glutamic acid (Glu/E)		129
Glutamine (Gln/Q)		128
Glycine (Gly/G)		57
Histidine (His/H)		137

Isoleucine (Ileu/I)		113
Leucine (Leu/L)		113
Lysine (Lys/K)		128
Methionine (Met/M)		131
Phenylalanine (Phe/F)		147
Proline (Pro/P)		97
Serine (Ser/S)		87
Threonine (Thr/T)		101
Tryptophan (Trp/W)		186
Tyrosine (Tyr/Y)		163
Valine (Val/V)		99

## Appendix B: Amino Acid Sequences of Different Species of PFD

Species	Sequence	PI	Molecular mass (mono isotopic)
<b>Prefoldin Subunit 1 (beta)</b>			
Human (hPFD)	MAAPVDLELKKAFTELQAKVIDTQQKVKLADIQIEQLNRTKKHAHLTDTEIMTLVDETNNMYEGGRMFIL QSKEAIHSQLEKQKIAEEKIKELEQKKSYSERSVKEAEDNIREMLMARRAQ	6.32	14102.44
Bovine (bPFD)	MAAPVDLELKKAFTELQAKVIDTQQKVKLADVQIEQLNRTKKHAHLTDTEIMTLVDETNNMYEGVGRM FILQSKEAIHNQLLEKQKIAEEKIKELEQKKSYSERSVKEAEDNIREMLMARRAQ	6.32	14214.50
Yeast (yPFD)	MSQIAQEMTVSLRNARTQLDMVNNQQLAYLDRQEKLAEELTKKELESYPTDKVWVRSCGKSFILQDKSKYV NDLSHDETLLDQRKTLKIKKNYLETTEKTIDNLKALMKN	8.97	12773.68
phPFD	MQNIPPQVQAMLGQLDTYQQQLQLVIQKQKQVQADLNEAKKALEEIEIETLPDDAQIYKTVGTIVKTTK EKAVQELKEKIEIETLVRLNALNRQEQKINEKVKELTQKIQAAALRPPTAG	7.88	13308.26
mtPFD	MELPQNVQHQLAQFQQQLQQQAQAIQVQKQTVEMQINETQKALEELSRAADDAEVYKSSGNILIRVAK DELTEELQEKELETLQREKTIERQEERVMKKLQEQMVNIQEQAMKGAGINPGMGN	4.87	13854.06
mjPFD	MELPPQIAQALMQLQQLQQLQMLMQKQSVETELKECKKALEELEKSSSDEVYKLVGGFLVKKRKED VKKELEEKVETELRVKTELEKQEEKLSRLKELQEKIQKMIPTAQ	6.41	13366.70
Random	TKAMAMQTKHEKLNMMKTRLDIIQLLEATILTQEIQIDTAVSEHLLREEGEALENIEKPKNQEKRFVKKRL QLFEGEELADLDSYVHAADKSYLKKKKSVMTIMVAAMDKEKEQAIEQRQI		
<b>Prefoldin Subunit 2 (beta)</b>			
Human (hPFD)	MAENSGRAGKSSGSGAGKAVSAEQVIAGFNLRLRQEQRLASKAAELEMELNEHSLVIDTLKEVDETR KCYRMVGGVGLVERTVKEVLPALENNKEQIQKIETLTQQLQAKGKELNEFREKHNIRLMGEDEKPAAKE NSEGAGAKASSAGVLVS	6.20	16637.61
Bovine (bPFD)	MAENSGRAGKSSGSGTGKAVSAEQVIAGFNLRLRQEQRLASKAAELEMELNEHSLVIDTLKEVDETR KCYRMVGGVGLVERTVKEVLPALENNKEQIQKIETLTQQLQAKGKELNEFREKHNIRLMGEDEKPAAKE NSEGAGAKASSAGVLVS	6.20	16637.61
Yeast (yPFD)	MEQRNNVFQAKYNEYKQILEELQTKIIELGHDKDEHTIVIKTLKDAEPTKCYRMIGGALVESDVQTSPLI LETKENIEGTISKMKETLIQTAKEFEKWKKDNKIQVVKNN	7.79	12985.85
phPFD	MQNIPPQVQAMLGQLDTYQQQLQLVIQKQKQVQADLNEAKKALEEIEIETLPDDAQIYKTVGTIVKTTK EKAVQELKEKIEIETLVRLNALNRQEQKINEKVKELTQKIQAAALRPPTAG	7.88	13308.26
mtPFD	MELPQNVQHQLAQFQQQLQQQAQAIQVQKQTVEMQINETQKALEELSRAADDAEVYKSSGNILIRVAK DELTEELQEKELETLQREKTIERQEERVMKKLQEQMVNIQEQAMKGAGINPGMGN	4.87	13854.06
mjPFD	MELPPQIAQALMQLQQLQQLQMLMQKQSVETELKECKKALEELEKSSSDEVYKLVGGFLVKKRKED VKKELEEKVETELRVKTELEKQEEKLSRLKELQEKIQKMIPTAQ	6.41	13366.70
Random	ENNKVKKESILQQESGANSVETGELARKGETAFKEEFKMTNYNVAGLDVGKCGKAMIRAHKKGE RTLRLTILIVAPSKVLLQEAERAEPVSLLAGNLLGAGIAGSSMGQSDAGELEQGAIREEVVARKSESEELN HSQQKDARKLQGNMV		
<b>Prefoldin Subunit 3 (alpha)</b>			
Human (hPFD)	MAAVKDCGKGEMATNGRRRLHLGIPEAVFVEDVDSFMKQPGNETADTVLKKLDEQYQYKFMELN LAQKKRRRLKGQIPEIKQTEILKYMQKKKESTNSMETRFLADNLYCKASVPPTDKMCLWLGANVMLEY DIDEAQALLEKNLSTATKNLDSLEEDLDFLRDQFTTEVNMAVYNWVVKRRNKDDSTKNKA	6.64	22643.44
Bovine (bPFD)	MAASKDGCVGVEAAGNGRRRLHLGIPEAVFVEDVDSFMKQPGNETADIVLKKLDEQYQYKFMELNL AQKKRRRLGQIPEIKQTEILKYMQKKKESTSSLETFLADNLYCKASVPPTDKVCLWLGANVMLEYDIDE AQALLEKNLLTATKNLDSLEEDLDFLRDQFTTEVNMAVYNWVVKRRNKDDSTKNKA	6.01	22340.50
Yeast (yPFD)	MDTLFNSTEKNARGIPQAPFIENVNIEIKDPSDFELCFNKFQERLSKYKFMQESKATIKQLKTRIPDLENT LKCQSLRNHSDGDESDPILLHYQLNDTLYTKAQVDIPEDRADLKVGLWLGADVMLEYPIDEAIELLKK KLADSEQLTVSTEDVEFLRENITTMEVNCARLYNWDVQRRQDLKQAQEGTKNLKI	4.78	22987.64
phPFD	MAQNNKELEKLAYEYQVLQAQAQILAQNLELLNLAKAEVQTVRETLENLKKIEEEKPEILVPIGAGSFLK GVIVDKNNIAIVSVSGYAVERSIDEAISFLEKRLKEYDEAIKKTQGALAELEKRIGEVARKAQEVQQKQS MTSFKVKK	6.43	16592.97
mtPFD	MEDQQRLEEIVNQLNIYQSQVELIQQQMEAVRATISEILEIKTSLDIQKGDSETLVPVGAGSFIKAEK DTSEVIMSVGAGVAIKKNFEDAMESIKSQKNELESTLQKMGENLRKITDIMMKLSPQAEELLKKVRGSG E	4.70	15689.09
mjPFD1	MENMAEDLRQKAMALEIYNQQLQMIQSEITSIRALKSEIMNSIKTIENIKADEETLIPVGPVFLKAKIVD DKALIGVKSDIYVEKSFNEVIEDLKKSVEDLDKAEKEGMKKAELAKAITALRKELQTEIQKAQQAQDKK Q	5.11	16079.60
mjPFD2	MVNEVIDINEAVRAYIAQIEGLRAEIGRLDATIATLRQSLATLKSRLTLEGKTVLVPVGSIAQVEMKVEK MDKVVVSVGQNNISAELEYEEALKYIEDEIKLLTFRVLVEQAIAELYAKIEDLIAEAQQTSEEEKAEENE KAE	4.45	16393.71
Random	TVKENLRSTDVAIVDKASEDQCDSCAFMPVLKGLQTEQNSLKLKRSMRDGGYLDLEIGLKAEEKVVN YDGTKETDLVIEQEAEEGLMDTALKGRIRKYTLPMQGNLYLRAEFYATNWEALAEVKKNTDTLAPD KEFSKLYADFTLLSDRTNLMKRPFGLKLMWNTNDEPNQKIHLMVQKMMFNKCDNRNKAQKD		
<b>Prefoldin Subunit 4 (beta)</b>			

Human ( <i>h</i> PFD)	MAATMKKAAAEDVNVTFEDQQKINKFARNTSRITELKEEIEVKKKQLQNLLEDACDDIMLADDDCLMIP YQIGDVFISHSQEETQEMLEEAQKKNLQEEIDALESRVESIQRVLADLVQLYAKFGSNINLEADE	4.42	15217.55
Bovine ( <i>b</i> PFD)	MAATMKKAAAEDVNVTFEDQQKINKFARNTSRITELKEEIEVKKKQLQNLLEDACEDIMLADDDCLMIP YQIGDVFISHSQEETQEMLEEAQKKNLQEEIDALESRVESIQRVLADLVQLYAKFGSNINLEADES	4.43	15318.60
Yeast ( <i>y</i> PFD)	MELLPPQQRNNTQVTFEDQQKINEFSKLIMRKDAIAQELSLQREEKEYLDDVSLEIELIDEDEPVQYKVG DLFIFMKQSKVTAQLEKDAERLDNKIETLEDKQRDIDSRDLAKAILYAKFGDNINLER	4.25	15170.78
<i>ph</i> PFD	MQNIPPQVQAMLGQLDTYQQQLQLVQKQKQVQADLNEAKKALEEIEITLPDDAQYKTVGTIVKTTK EKAVQELKEKIEITLEVRLNALNRQEQKINEKVKELTQKIQAAALRPPTAG	7.88	13308.26
<i>mt</i> PFD	MELPQNVQHQLAQFQQQLQQAQAIQSVQKQTVEMQINETQKALEELSRADDAEVYKSSGNILIRVAK DELTEELQEKELETQLREKTIERQEERVMKKLQEMQVNIQEAAMKGAGINPGMGN	4.87	13854.06
<i>mj</i> PFD	MELPPQIQAAQLMQLQQLQQLQMLMQKQSVETELKECKKALEELEKSSSDEVYKLVGGFLVKKRKED VKKELEEKVETELRVKTELEKQEEKLQSRKELQEIKQKMIPTAQ	6.41	13366.70
Random	RFNLVLLADQTCMVLQVQKASQMMAQNIIRETSFKEVARANEENDIGLEKDNKMDAEPAAIDAS LKGTDQLFESNESKDLKFYALCKIHDQVAEQEIEENITREDLDIELVELIDSKEAEDKMVTAKYE		
<b>Prefoldin Subunit 5 (alpha)</b>			
Human ( <i>h</i> PFD)	MAQSINITELNLPQLEMLKNQLDQVEFLSTSIQQLKVVTQKYVEAKDCLNVLNKSNEGKELLVPLTSSM YVPGKLHDVEHVLIDVGTGYVEKTAEDAKDFFRKIDFLTKQMEKIQPALQEKHAMKQAVMEMMS QKIQQLTALGAAQATAKA	5.94	17317.00
Bovine ( <i>b</i> PFD)	MAQSVNITELNLPQLEMLKNQLDQVEFLSTSIQQLKVVTQKYVEAKDCLNVLNKSNEGKELLVPLTSS MYVPGKLHDVEHVLIDVGTGYVEKTAEDAKDFFRKIDFLTKQMEKIQPALQEKHAMKQAVMEMM SQKIQQLTTLGAAQATAKA	6.31	17374.05
Yeast ( <i>y</i> PFD)	MSSQKIDLTCLNPEQLNAVKKQFDQELQHFTQSLQALTMAGKFTCEIDDIKTVSQAGNEGQKLLVPA SASLYIPGKIVDNKKFMVDIGTGYVEKSAEAAIAFYQKQVDKLNKESVQIQDIIEKKTQYSLIEAQIRQA AIRQHEAMSKQQQQQKESSTA	8.84	18344.51
<i>ph</i> PFD	MAQNNKELEKLAYEQVLQAQAQILAQNLLELLNAKAEVQTVRETENLKKIEEEKPEILVPIGAGSFLK GVIVDKNNNAIVSGGYAVERSIDEAISFLEKRLKEYDEAIKKTQGALAELEKRIGEVARKAQEVQQKQS MTSFKVKK	6.43	16592.97
<i>mt</i> PFD	MEDQQRLEEIVNQLNIYQSQVELIQQQMEAVRATISEILEIKTLDSDIQGKDGSETLVPVGAGSFIKAEK DTSEVIMSVGAGVAIKKNFEDAMESIKSQKNELESTLQKMGENLRKITDIMMKLSPQAEELLKKVRGSG E	4.70	15689.09
<i>mj</i> PFD1	MENMAEDLRQKAMALEIYNQQLMIQSEITSIRALKSEIMNSIKTIENIKADEETLIPVGPVFLKAKIVD DKALIGVKSIDIYVEKSFNEVIEDLKKSVEDLDKAEKEGMMKAEELAKAITALRKELQTEIQKAQQAQDKK Q	5.11	16079.60
<i>mj</i> PFD2	MVNEVIDINEAVRAYIAQIEGLRAEIGRLDATIATLRQSLATLKSRLTLEGKTVLVPVGSIAQVEMKVEK MDKVVVSVGGQNISAELEYEEALKYIEDEIKKLLTFRVLVEQAIAELYAKIEDLIAEAQQTSEEEKAEENE KAE	4.45	16393.71
Random	EEVSLGLEVKAAAHHSKGPMNVEIPYGGQAAMKLYQKLDMLLVILKKVGGKALNERTNLFEFKLKELQE TQQKSQTHALDMSQMNLNLPVTKQSAIYLFIMCNIEAQEKVDVSDIQTEEEADQQTSDIPAFADVQ TQGVKLVMTVAMTAKKVL		
<b>Prefoldin Subunit 6 (beta)</b>			
Human ( <i>h</i> PFD)	MAELIQKKLQGEVEKYQQLQKDLKSKMSGRQKLEAQLTENNIVKEELALLDGSNVVFKLLGPVLVKQEL GEARATVGKRLDYITAEIKRYESQLRDLERQSEQQRETLAQQLQEFQRAQAAGAPGKA	8.83	14573.80
Bovine ( <i>b</i> PFD)	MAELIQKKLQGEVEKYQQLQKDLKSKMSGRQKLEAQLTENNIVKEELALLDGSNVVFKLLGPVLVKQEL GEARATVGKRLDYITAEIKRYESQLRDLERQSEQQRETLAQQLQEFQRAQAAGAPGKA	7.88	14545.76
Yeast ( <i>y</i> PFD)	MSELGAKYQQLQNELEEFIVARQKLETQLQENKIVNEEFDQLEEDTPVYKLTGNVLLPVEQSEARTNVD KRLEFIETITRCEKNIRDQKEELEKMRSELKLNNTAASGPGR	4.70	13275.77
<i>ph</i> PFD	MQNIPPQVQAMLGQLDTYQQQLQLVQKQKQVQADLNEAKKALEEIEITLPDDAQYKTVGTIVKTTK EKAVQELKEKIEITLEVRLNALNRQEQKINEKVKELTQKIQAAALRPPTAG	7.88	13308.26
<i>mt</i> PFD	MELPQNVQHQLAQFQQQLQQAQAIQSVQKQTVEMQINETQKALEELSRADDAEVYKSSGNILIRVAK DELTEELQEKELETQLREKTIERQEERVMKKLQEMQVNIQEAAMKGAGINPGMGN	4.87	13854.06
<i>mj</i> PFD	MELPPQIQAAQLMQLQQLQQLQMLMQKQSVETELKECKKALEELEKSSSDEVYKLVGGFLVKKRKED VKKELEEKVETELRVKTELEKQEEKLQSRKELQEIKQKMIPTAQ	6.41	13366.70
Random	EQMYRKKTLLALLAESSTLQKKQVQKEGLRAQYNEVPKEEDGRSLAQQNGQLTALLGEQIQQLQGESDI KKSREGKFLVRLKIQAGVEQIAEALGQKRRARVQEVLLQNLALREMPDLYEKQDESFAKV		

Pyrococcus horikoshii (*ph*), Methanobacterium thermoautotrophicum (*mt*), Methanocaldococcus jannaschii (*mj*)

*ph* and *mt* only have alpha and beta subunits. *mj* has 2 alpha subunits and 1 beta subunit. Therefore, the sequence of *ph*, *mt*, *mj1* and *mj2*  $\alpha$  subunits are compared against all corresponding *h*PFD  $\alpha$  subunits (subunits 3 and 5). The sequence of *ph*, *mt* and *mj*  $\beta$  subunits were compared against all corresponding *h*PFD  $\alpha$  subunits (subunits 1, 2, 4, 6).

## Appendix C: Ions Identified by Proteomic Analysis Correlating to Endoplasmin

Observed <i>m/z</i>	Mass Expected	Mass Calculated	Error (ppm)	Miss	Sequence	Ion Score
460.2758	918.5371	918.5287	9	1	K.NKEIFLR.E	12
473.3086	944.6026	944.5906	13	0	R.GTTITLVLK.E	2
482.3007	962.5868	962.5800	7	0	K.LIINSLYK.N	5
541.2785	1080.5425	1080.5352	7	0	K.FAFQAEVNR.M	3
380.5354	1138.5845	1138.5731	10	0	K.LGVIEDHSNR.T	10
575.7777	1149.5409	1149.5302	9	0	K.EAESSPFVER.L	22
594.3522	1186.6898	1186.6710	16	0	K.SILFVPTSAPR.G	4
638.3306	1274.6467	1274.6354	9	0	R.ELISNASDALDK.I	8
743.3898	1484.7650	1484.7471	12	0	K.GVVDSDDLPLNVSR.E	40
515.6183	1543.8332	1543.8205	8	1	R.ELISNASDALDKIR.L	13
772.9248	1543.8351	1543.8205	9	1	R.ELISNASDALDKIR.L	18
537.9611	1610.8614	1610.8887	-17	2	R.MMKLIINSLYKNK.E + Oxidation (M)	3
1015.5478	2029.0810	2029.0579	11	0	R.LISLTDENALAGNEELTVK.I	14

Error (ppm): refers to the error window of the mass in parts per million. Miss: Refers to the b or y ions of a particular amino acid that were not identified in the MS/MS frame. Ion score: refers to the significance threshold, based on the calculated probability P,  $-10\log(P)$ . N.O.= not observed

## Appendix D: Genes Encoding the *h*PFD subunits 5 and 6

	<i>h</i> PFD Subunit 5
Gene	<p><b>atg</b>tcgtactaccatcaccatcaccatcacgattacgatatccaacgaccgaaaacgtgtattttcagggcgccatgggatccatggcccagagcatcaatattaccga  actgaacctgccgagctggagatgctgaagaatcagctggatcaagagggtggagtttctgagcaccagcattgccagctgaaagtggcgagaccaagtatgttg  aggccaaagactgcctgaacgtgctgaacaagagcaacgagggcaaagaactgctggtgccgctgaccagcagcatgtatgtccgggcaaactgcatgatgtgga  gcacgtgctgatcgtgtgggtacaggctattacgtggagaagaccgccgaagacgccaagacttcttaaacgtaaaattgattttctgaccaaacagatggaga  aaattcagcggccctgcaggaagcatgccatgaaacaggccgtgatggagatgatgagccagaagatccagcagctgacagcactgggtgccgcagggcaa  ccgccaagcagcggccgctttcgaatctagagcctgcagtctcaggcatgcggtaccaagcttggtgttttggcgga<b>tga</b></p>
Sequence	<p>MSYYHHHHHDYDIPTTENLYFQGAMGSMASINITELNLPQLEMLKNQLDQVEFLSTSIAQLKVVQTKYVEAKDCLNVLNKSNE  GKELLVPLTSSMYVPGLKLDVEHVLIDVGTGYVEKTAEDAKDFFKRKIDFLTQKMEIKPALQEKHAMKQAVMEMMSQIKIQL  TALGAAQATAKAAAFESRACSLACGTLKLCFGG</p>

	<i>h</i> PFD Subunit 6
Gene	<p><b>atg</b>tcgtactaccatcaccatcaccatcacgattacgatatccaacgaccgaaaacgtgtattttcagggcgccatgggatccatggccgaactgattcagaaaaagc  tgcagggcggaagtggagaaataccagcagctgcagaaggatctgagcaaaagcatgagcggtcgccagaaactggaggcccagctgaccgagaacaacatcgtg  aaagaggagctggccctgctggacggtagcaatgtggtgttcaaactgctgggtccggtgctggtgaagcaggaactgggcgaagcccggtgaacctgggtaaac  gcctggactacatcaccgccgaaatcaaacgtacgaagccagctgcgcatctggaacgtcagagcgaacagcagctgagaccctggcccagctgcagcagga  atttcagcgtgccaggcagccaaagcaggtgcaccgggtaaagccgcccgcgctttcgaatctagagcctgcagtctcaggcatgcggtaccaagcttggtgtttt  ggcgga<b>tga</b></p>
Sequence	<p>MSYYHHHHHDYDIPTTENLYFQGAMGSMALIQKKLQGEVEKYQLQKDLKSMSGRQKLEAQLTENNIVKEELALLDGSNNVF  KLLGPVLVKQELGEARATVGKRLDYITAEIKRYESQLRDLERQSEQRETLAQLQEFQRAQAAGAPGKAAAFESRACSLAC  GTKLGCFFG</p>

## Appendix E: Computationally Calculated Properties of *b*PFD and *h*PFD Subunits 5 and 6

	<i>b</i> PFD complex	<i>h</i> PFD Subunit 5	<i>h</i> PFD Subunit 6
Molecular Weight (Ave) Da	100403.87	22931.1 w/o m 22842.147 (av.) 22827.369 (mono.)	20185.83 w/o m 20096.641 (av.) 20084.176 (mono.)
Number of amino acids	889	205	180
Theoretical pI	5.51	5.98	6.60
Extinction coefficient	38195*	12170*	10555*
	37820**	11920**	10430**
Abs 0.1% (=1 g/l)	0.380*	0.531*	0.523*
	0.377**	0.520**	0.517**
Instability index	43.48 (unstable)	41.22 (unstable)	52.51 (unstable)
Alipathic index	88.68	81.90	78.17
Grand average of hydropathicity	-0.573	-0.301	-0.626

Notes: Open sources available on <http://web.expasy.org/protparam/> were used to calculate these parameters. \*assuming all pairs of Cys residues form cysteines, \*\* assuming all Cys residues are reduced

## Appendix F: Post-hoc Test Following 1 Way ANOVA Summary of Cell Viability Assays for A $\beta$ (1-40) Co-Incubated with *h*PFD Subunit 5

	Mean Difference	q	Significance: P < 0.05
<b>Buffer</b>			
<i>h</i> PFD subunit 5	149.2	2.495	No
A $\beta$ (1-40) fibrils	532.0	11.24	Yes
A $\beta$ (1-40) monomer	774.1	17.12	Yes
1:0.02	269.8	6.427	Yes
1:0.1	484.6	10.99	Yes
1:0.2	542.0	12.56	Yes
1:0.5	562.4	12.89	Yes
<b><i>h</i>PFD subunit 5</b>			
A $\beta$ (1-40) fibrils	-382.7	6.125	Yes
A $\beta$ (1-40) monomer	-624.9	10.25	Yes
1:0.02	-120.5	2.654	Yes
1:0.1	-335.4	5.573	Yes
1:0.2	-392.8	6.601	Yes
1:0.5	-413.2	6.906	Yes
<b>A<math>\beta</math>(1-40) fibrils</b>			
A $\beta$ (1-40) monomer	-242.1	4.962	Yes
1:0.02	-262.2	5.726	No
1:0.1	-47.39	0.9922	No
1:0.2	10.05	0.2143	No
1:0.5	30.45	0.6436	No
<b>A<math>\beta</math>(1-40) monomer</b>			
1:0.02	-504.4	11.56	Yes
1:0.1	-289.5	6.335	Yes
1:0.2	-232.1	5.182	Yes
1:0.5	-211.7	4.681	Yes
<b>1:0.02</b>			
1:0.1	214.8	5.057	No
1:0.2	272.3	6.562	No
1:0.5	292.7	6.973	No
<b>1:0.1</b>			
1:0.2	57.44	1.316	No
1:0.5	77.84	1.765	No
<b>1:0.2</b>			
1:0.5	20.40	0.4726	No

Buffer, A $\beta$ (1-40) fibrils and A $\beta$ (1-40) monomer incubated independently. Ratio: A $\beta$ (1-40) monomer: *h*PFD subunit 5. Mean Difference: difference between the means of the data sets. q: false discovery rate (FDR) adjusted p values. Significance: if probability (P) is calculated to be smaller than 0.05, than the sets of data are significantly different.



## Appendix G: Post-hoc Test Following 1 Way ANOVA Summary of Cell Viability Assays for A $\beta$ (1-40) Co-Incubated with *h*PFD Subunit 6

	Mean Difference	q	Significance: P < 0.05
<b>Buffer</b>			
<i>h</i> PFD subunit 6	75.04	1.688	No
A $\beta$ (1-40) fibrils	649.1	17.90	Yes
A $\beta$ (1-40) monomer	846.0	24.71	Yes
1:0.02	669.4	20.58	Yes
1:0.1	840.4	26.05	Yes
1:0.2	896.3	28.21	Yes
1:0.5	967.5	30.45	Yes
<b><i>h</i>PFD subunit 6</b>			
A $\beta$ (1-40) fibrils	-574.1	12.57	Yes
A $\beta$ (1-40) monomer	-771.0	17.48	Yes
1:0.02	-594.4	13.88	Yes
1:0.1	-765.3	17.45	Yes
1:0.2	-821.3	19.43	Yes
1:0.5	-892.5	21.11	Yes
<b>A<math>\beta</math>(1-40) fibrils</b>			
A $\beta$ (1-40) monomer	-196.9	5.493	Yes
1:0.02	20.29	0.5928	No
1:0.1	191.3	5.631	Yes
1:0.2	247.2	7.379	Yes
1:0.5	318.4	9.505	Yes
<b>A<math>\beta</math>(1-40) monomer</b>			
1:0.02	-176.7	5.509	Yes
1:0.1	-5.673	0.1784	No
1:0.2	50.25	1.606	No
1:0.5	121.5	3.881	No
<b>1:0.02</b>			
1:0.1	171.0	5.711	Yes
1:0.2	226.9	7.715	Yes
1:0.5	298.1	10.14	Yes
<b>1:0.1</b>			
1:0.2	55.92	1.921	No
1:0.5	127.1	4.367	Yes
<b>1:0.2</b>			
1:0.5	71.21	2.493	No

Buffer, A $\beta$ (1-40) fibrils and A $\beta$ (1-40) monomer incubated independently. Ratio: A $\beta$ (1-40) monomer: *h*PFD subunit 6. Mean Difference: difference between the means of the data sets. q: false discovery rate (FDR) adjusted p values. Significance: if probability (P) is calculated to be smaller than 0.05, than the sets of data are significantly different.