Structural Studies of Cystatin B Amyloid Fibre and Oligomer

Peter John Davis

PhD Thesis

July 2013

University of Sheffield

Department of Molecular Biology and Biotechnology

Abstract

Amyloid fibres are characteristic of over 25 degenerative human diseases including Alzheimer's and Parkinson's disease. Amyloid fibres are insoluble, highly stable, ordered cross-β sheet structures, which form as a result of conformational change and aggregation in a range of unrelated soluble proteins and peptides. High molecular weight oligomers, potentially on-pathway to amyloid formation, have been posited as the toxicity agent in amyloid-associated disease. Heterogeneity of amyloid fibres and oligomers precludes the use of standard structural biology techniques.

The work presented utilised recombinant human cystatin B as a model system for structural analysis of amyloid fibres and oligomers, particularly for the structurally homologous human cystatin C which directly causes a form of amyloid angiopathy. Limited proteolysis of cystatin B amyloid fibre shows that the core structure consists maximally of residues 27-80 out of a total 98. Electron microscopy techniques, including mass per unit length measurements, reveal an average fibre width of 8.6 nm and formation of four fibre classes, composed of 4, 8 and 16 molecules per 4 β-strand rise. These data are incompatible with the previous native-like model, therefore a new working fibre model is proposed where native β-strands 2 and 3 are extended into a single strand with adjacent β-strands 4 and 5 forming the other half of a β-sheet arc, which is stacked in parallel, perpendicular to the fibre axis.

Stable proteinase-K resistant G4R mutant cystatin B oligomers were generated and preliminary characterisation presented, with identification of five structural classes ranging in diameter from 8-88 nm. Furthermore, investigation of a potential amyloid therapeutic in *Salvia* plant extracts was carried out with dye binding kinetic assays and electron microscopy. The work herein primarily constrains a new non-native β-sheet cystatin B amyloid fibre model and has initiated new avenues of research into cystatin oligomer structure and novel therapeutic agents.

Acknowledgements

The process of completing a PhD has been described to me as a strange beast (Morgan, 2006) or living in a dark tunnel for several years (Peter Artymiuk, personal communication). I certainly understand the sentiment, with its combination of independence and collaboration, working and learning, earning and yet not doing a 'real' job. I suppose it is more of a calling. I've found the best metaphor for describing a PhD being that of mining, in which a senior geologist (supervisor) suggests a direction and you then go off and dig using the right tools and a combination of basic knowledge, extra supervisor suggestions, help from colleagues and using your own growing knowledge to choose a sensible direction. Sometimes you come across unyielding rock, broken tools and complete darkness where it is difficult to keep going. Other times things go well, the going is easier and you find iron, silver, gold and haematite. The joy of finding something interesting and good is amazing, yet equally, the lows when nothing seems to work can be pretty deep. When looking at the process overall I have enjoyed my PhD and had I not done it, would probably have regretted the decision. I have developed a deeper inquisitiveness, cynicism and analytical ability. I would recommend doing a PhD to others, but it would come with a warning; it ain't easy, nor the most sensible thing to do. Anyway, without help and support at University and outside in the form of supervision, advice and friendship, it would not have been possible to have produced this body of work. So thank you to everyone who has helped, whether named here or not.

First, I must thank Rosie Staniforth for providing me the opportunity to follow my interests, suggesting avenues of study, encouragement, advice and discussions. Thank you for supervising me and my work, answering a barrage of questions every so often and 11th hour reviewing of my thesis and helping work out what some of the data means. Plus providing a friendly insight into life and living around the world.

I must also thank my advisors, Jon Waltho and Per Bullough for critical appraisal and insight. Helping direct me along suitable avenues of study, advising a course of action and keeping me on the straight and narrow when I was plunging into a wide sea of nothing. I must also thank Jon for getting me involved in the therapeutics studies and offering me some interesting work with Aβ and phage proteins, which has reinvigorated my interest in research.

I am grateful for the members of the cystatin group past and present for the work they have done, of which I based many of my experiments. Robert Paramore and Carrie-Anne Sharma, were very helpful in the early days, by explaining much of what there is to know about working with cystatin and HPLC machines. A special thanks to Rob who introduced me to many of the quirky features and techniques required when working with aggregated protein and helped me design many of my experiments in the first few years, plus helped keep me sane through friendship and ensured I got some

exercise during the weekly football matches. Thank you to Abigail Williams as well for putting up with my aired frustration, listening, discussions about cake, life etc. Thanks go to Adham Shawaihde for interesting discussions about Libya and Amy Irvine for sharing her recent results on Salvia extracts. Thanks also goes to the seven undergraduate students whom have worked with me. Some of the data is presented and they are named accordingly.

The NMR group collective have been great, I won't name everyone, but thanks for enduring lab meetings about amyloid, engaging in interesting critical discussion and more importantly, helping in the labs and general banter. A special thank you to Hugh Dannatt for collective rants, discussions about the world, science, music, NMR etc. Plus cheers for the amusing distractions, pizza, falafel Tuesdays and always intriguing conversations, sometimes involving Bernard's watch and lots of liquid nitrogen. Matthew Cliff is a star and can usually answer every question in a sensible, easy to understand way and could usually solve the many persistent little problems. Moreover, he is a fun chap to chat to and has a great thoughtful view on most things. Andrea Hounslow is super for sorting out and aiding in all the NMR experiments and dealing with hand wavy experimental descriptions. Cheers to everyone, and thanks for all the excellent cake, it makes Fridays so much better.

One of the tremendous things about working at a University is the excellent people you meet and work with everyday. Outside of the lab group I have to be grateful to Ralph Moyle for fixing most of the equipment in the wet lab. Svetomir Tzokov for keeping the electron microscope functioning, advice on getting good images and ensuring the grids keep coming. Simon Thorpe in the Chemistry department for letting me use one of the mass spectrometers, showing me how to use it and putting up with regular questions and calibrations. Arthur Moir has helped with N-terminal sequencing, peptide preparation and for his knowledge of all things peptide and explaining size exclusion chromatography in the best way possible. Thank you also to David Holmes who has spent several long days with me measuring the mass of amyloid fibres and showing me how everything works, without even blinking an eye, despite taking up so much time.

Outside University, the 'old' High Peakers have been amazing friends with lots of lovely meals, wine, walks, games, biking, climbing, skiing, trips etc. Thank you to my family for supporting me, and to my parents for providing the inspiration and materials when as a child I took an interest in science. Thank you to my wife Kathryn for being supportive, putting up with me, especially when I say, I should be home in 20 minutes (2 hours), saying yes, travelling and companionship.

Finally, I hope you the reader find this thesis informative and useful. Enjoy.

Contents

List of Figures

Chapter 2

Chapter 3

Chapter 4

Appendices

List of Tables

Abbreviations:

ADF	Annular dark field
ALS	Amyloid lateral sclerosis
AFM	Atomic force microscopy
AUC	Analytical ultracentrifugation
β_2m	β_2-microglobulin
C-terminal	Carboxylate end protein region
CD	Circular dichroism spectroscopy
EDTA	Ethylenediaminetetraacetic acid
EGCG	Epigallocatechin-gallate
EM	Electron microscopy
EPM1	Myoclonus epilepsy of type I (Unverricht-Lundborg Disease)
EPR	Electron paramagnetic resonance spectroscopy
ESI-MS	Electrospray-ionisation mass spectroscopy
FAP	Familial amyloidotic polyneuropathy
FLM	Fluourescence light microscopy
FTIR	Fourier transform infrared spectroscopy
H/D Exchange	Hydrogen/deuterium exchange of protein backbone amides
HO	High molecular weight oligomer
HPLC-MS	Reverse-phase high-Pressure-liquid-chromatography coupled with electrospray-ionisation mass spectroscopy
HSQC	Heteronuclear single quantum coherence NMR spectroscopy
LO	Low molecular weight oligomer
LP	Limited proteolysis
MPL	Mass per unit length
MS	Mass spectroscopy
MWCO	Molecular weight cut-off
N-terminal	Amide end protein region
NMR	Nuclear magnetic resonance
OD	Optical density
PES	Polyethersulfone
PMSF	phenylmethanesulfonyl fluoride

RP-HPLC	Reverse-phase high-pressure-liquid-chromatography
SDM	Site-directed mutagenesis
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SEC HPLC	Size exclusion chromatography high-pressure-liquid-chromatography
SOD-1	Superoxide dismutase-1
SSNMR	Solid state nuclear magnetic resonance spectroscopy
STEM	Scanning transmission electron microscopy
TEM	Transmission electron microscopy
TFE	2,2,2, trifluoroethanol
ThT	Thioflavin-T
TLCK	N_{α}-Tosyl-L-lysine chloromethyl ketone hydrochloride
TROSY	Transverse relaxation optimised NMR spectroscopy
TTR	Transthyretin
UV	Ultraviolet

Chapter 1: Introduction

1.1 Background in Brief

Amyloid-related diseases, or amyloidoses, include over 25 familial, sporadic or transmissible pathological conditions, some with high prevalence in the human population (Stefani, 2010), and include Alzheimer's disease (AD) (Selkoe and Schenk, 2003), Parkinson's disease (PD) (Lang and Lozano, 1998) and Huntington's disease (Perutz, 1999) (see table 1.1.1). Many of these diseases are late-onset chronic diseases which have become increasingly important as the populations in developed nations age. For example, Alzheimer's disease (AD) is the fourth most widespread illness in the economically developed world, with an estimated 24 million sufferers (Ballard *et al.*, 2011) and therefore understanding the cause and developing a treatment for amyloidoses is essential.

Amyloidoses are characterised by amyloid fibre deposits in samples of affected tissue. Amyloid fibres are insoluble, highly ordered and typically highly stable protein deposits, formed as a result of conformational change and aggregation of normally soluble protein. Amyloid fibres are highly stable due to extensive hydrogen bonding, and thus may accumulate and persist within cells for long periods of time (Dobson, 2003). However to complicate the issue, long mature amyloid fibres appear to be relatively inert and the quantities of aggregate in disease can range from nearly undetectable to kilograms of protein in affected organs (Ye and Godzik, 2003).

Instead of amyloid fibres directly causing disease, prefibrillar protein complex intermediates either on or off pathway to amyloid formation, broadly termed 'oligomers', have been identified by electron microscopy (EM) (Lashuel *et al.*, 2002) and with conformation specific antibodies as the possible toxicity agent in amyloid-associated disease (Glabe, 2004). Although conclusive proof of direct cytotoxicity of these structures remains elusive, overexpression, duplication and sequence mutations of amyloid forming proteins such as Aβ in Alzheimer's disease or huntingtin in Huntington's disease can lead to early-onset and/or increased risk of disease, thus proving these proteins are involved in disease aetiology (Benilova *et al.*, 2012) .

Protein/Peptide	**Disease(s)**	**Notes**
Aβ	Alzheimer's disease	Accumulation of Aβ plaques
α-synuclein	Parkinson's disease, dementia with Lewy bodies and others	
β2-Microglobulin	Haemodialysis-related amyloidosis	Systemic amyloidosis
Huntingtin	Huntington's disease	Accretion of polyglutamine-containing aggregates (Perutz, 1999).
Insulin	Injection-localised amyloidosis	
IAPP (amylin)	Type II diabetes	Accumulation of islet amyloid polypeptide (IAPP).
Lysozyme	Systemic amyloidosis	
PrP (prion)	Spongiform encephalopathies	Includes Creutzfeldt-Jakob disease (CJD) – amplification of misfolded prion protein (PrP^{sc}).
Tau	Alzheimer's disease, frontotemporal dementia and others	Forms tau neurofibrillary tangles.
TTR (Transthyretin)	Familial amyloidotic polyneuropathy (FAP), senile systemic amyloidosis	FAP is caused by amyloid fibre composed largely or entirely of TTR (Costa *et al.*, 1978).

Table 1.1.1 Selected Amyloid Fibre Forming Proteins and Associated Diseases

Based upon a table from (Hard and Lendel, 2012).

Despite limited evidence supporting toxicity of amyloid fibre itself, these structures must still be studied because they are robust with extensive hydrogen bonding, thus may accumulate and persist within cells for a long period of time and are proposed to act as a reservoir species for toxic protofibrils and oligomers (Benilova et al., 2012, Harper et al., 1997, Walsh et al., 1997). The potential of amyloid fibres as a reservoir species is highlighted by the dynamic nature of amyloid fibres which exhibit extensive, but slow molecular recycling whereby the protein components of a fibre dissociate and reattach over a period of days (Sanchez et al., 2011, Carulla et al., 2005). Understanding the structure of amyloid fibres is therefore essential for the research field as it is one of the potential therapeutic targets whereby stabilisation or ablation of fibre could remove damaging concentrations of toxic species.

The protein chemist also asks; how does a broad range of soluble proteins lacking sequence or structural homology form amyloid fibre with a common morphology shown in electron microscopy and a suggested universal, cross-β core structure from X-ray fibre diffraction (Blake and Serpell, 1996). For example, amyloid found in Alzheimer's disease is composed of a 40-43 residue peptide (Selkoe and Schenk, 2003), whilst amyloid associated with Creutzfeldt-Jacob disease (CJD) is generated from a 30 kDa glycoprotein (Zahn *et al.*, 2000). This notion therefore produces the suggestion that amyloid fibre structures are generic and formed from short regions of relatively homologous protein sequence producing an extensive β-sheet structure (Chiti *et al.*, 2003). Therefore, even without the medical importance, amyloid formation is a fascinating aspect of protein folding. Furthermore, the high prevalence and growing requirement for developing therapeutics is driving the importance of research into amyloidoses. This includes research on the protein structures formed, the mechanism of formation, toxicity and the search for effective treatments

1.2 Definitions

To be designated an amyloid fibre forming protein by the Nomenclature Committee of the International Society of Amyloidosis, the protein must occur in tissue deposits and exhibit affinity for Congo red and green birefringence when viewed by polarisation microscopy. Furthermore, the protein must have been unambiguously characterised by protein sequence analysis (DNA sequencing in the case of familial diseases). Current nomenclature lists include 27 human and nine animal amyloid fibre forming proteins together with a list of eight inclusion bodies that exhibit some of the properties of amyloid fibres (Sipe *et al.*, 2010).

Nomenclature in the literature is diverse and controversial which provides poor clarity in descriptions as no consensus has been reached, particularly in the study of 'oligomers'. Described below is my personally suggested nomenclature consensus which is developed from a variety of sources from different fields of amyloid protein study and is similar to that described in (Kodali and Wetzel, 2007, Fandrich, 2012). The terminology defined here will be used throughout this work.

Amyloid Fibre

Primarily straight, morphologically regular, unbranched protein fibres with diameters typically in the 2-20 nm range and several micrometers long (Fandrich, 2012, Kodali and Wetzel, 2007) (see figure 1.2.1). Amyloid fibres are typically composed of 2-6 protofilaments and can be flat or twisted, with the latter leading to apparent fibre constrictions at regular distances, termed crossovers. Helicity of fibres is usually left-handed, with some rare exceptions to this rule (Fandrich, 2012). Most amyloid fibres exhibit; (i) cross-β x-ray fibre diffraction patterns (ii) β-sheet rich circular dichroism (CD) and Fourier Transform Infrared spectroscopy (FTIR) spectra (iii) high affinity for dyes such as thioflavin T (ThT) and Congo red (CR) (iv) core structures highly resistant to hydrogen/deuterium exchange and proteolysis. Amyloid fibres often consist of multiple protofilaments twisted around a fibril axis (Kodali and Wetzel, 2007). Section 1.5 presents a review of current structural data on amyloid fibre.

Protofilament

Individual strand of which a mature amyloid fibre is composed.

Protofibrils

Non-spherical, filamentous structures absent of periodic substructure which are thought to represent late-stage intermediate species in amyloid fibre formation. Compared to mature amyloid fibre, protofibrils are characteristically smaller in diameter (< 10 nm), shorter (< 400 nm in length) and more curvilinear. Protofibrils typically have a weaker affinity for ThT and CR dyes than mature fibres but still contain high levels of β-sheet structure identified by CD, FTIR, and X-ray fibre diffraction (Fandrich, 2012).

Oligomer

In amyloid research, it is a blanket term used for structurally distinct species larger than monomer, typically metastable, and should only be used for structures without a fibrous morphology to define the difference between oligomers and protofibrils (figure 1.2.1). The nomenclature is perhaps unclear due to little structural information, diverse sizes in single preparations and a variety of morphologies exhibited. The noun oligomer is derived from the Greek word oligos, meaning "a few", which arguably in usage can depend upon the total sum of entities. It has been suggested that aggregate is a better descriptive term (Fandrich, 2012), however I feel that an aggregate, as used in geology, is a collection of entities with different sizes, shapes, no ordered structure and formed from various materials. The latter two features poorly describe these oligomers which individually have an ordered structure, are made of the same material and neither are the single entities bound together, therefore I do not feel it is not a suitable description.

Oligomers have distinct morphologies in which the structures and composition of individual monomers are probably homologous. In the literature oligomers of varied morphology have been described as amorphous aggregates, micelles, prefibrillar aggregates, soluble oligomers, globular oligomers, Aβ-derived diffusible ligands (ADDLs) (Giasson et al., 2003) and protofibrils (Nelson et al., 2005). However, due to common use in current literature, this body of work uses nomenclature based on oligomer size classification described herein as low molecular weight oligomers (LOs) and high molecular weight oligomers (HOs) which for convenience are described respectively as less than and greater than 10 protein molecules. Additional detailed morphology classifications such as annular and spherical oligomers will be discussed when necessary. Some reviews on the subject include (Lashuel and Lansbury, 2006, Benilova et al., 2012, Fandrich, 2012).

Intermediate oligomer

Transitional oligomeric species on-pathway to forming amyloid fibre

Amorphous aggregate

Accumulation of protein molecules absent of clearly defined structural characteristics.

Amyloid Fibre Seed

Entity that when added may induce kinetically stable molecules to aggregate, often through growth of the seed itself (Kodali and Wetzel, 2007).

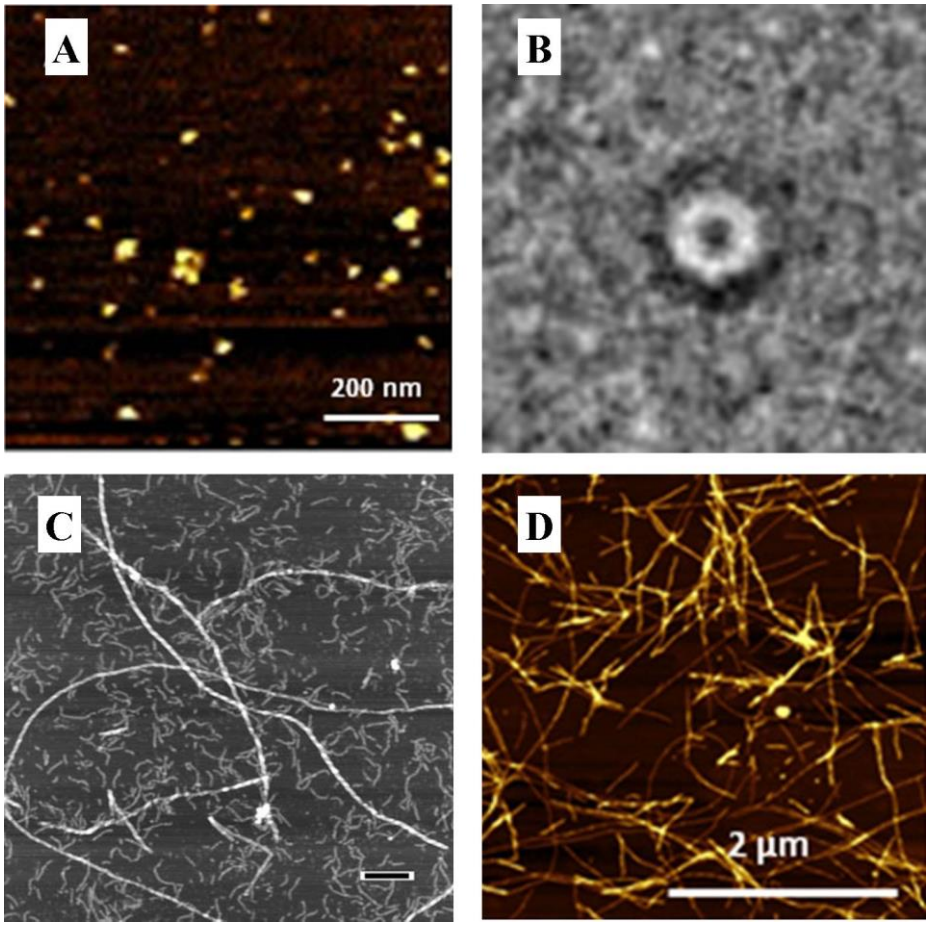


Figure 1.2.1 Amyloid Fibres, Protofibrils and Oligomeric Structures

AFM and EM images of structures formed by amyloid forming proteins. **(A)** AFM of globular Aβ oligomers. **(B)** EM of annular Aβ oligomer. **(C)** AFM of small curly α-synuclein protofibrils and thicker mature fibres. **(D)** AFM of mature Aβ amyloid fibre. Figure adapted and reprint permission obtained (Bartolini *et al.*, 2011) (A, D), Macmillan Publishers Ltd: [Nature] (Lashuel *et al.*, 2002) (B) and (Koo *et al.*, 1999) (C).

1.3 Amyloid Associated Disease

Amyloidoses, include over 25 familial, sporadic or transmissible pathological conditions, some with high prevalence in the human population (Stefani, 2010). Table 1.1.1 contains some of the most common and well-studied examples. Many age-related degenerative disorders are recognised to be amyloid-associated diseases with the accumulation of misfolded proteins and are often termed 'protein misfolding diseases. Aggregate clearance is often slowed down by disease and ageing, leading to the accumulation of misfolded and misassembled proteins which then leads to amyloid associated diseases (Smith, 2003).

A further link to amyloid forming proteins being causally related to the aetiology of amyloidoses is the finding that familial mutations which increase protein concentration or aggregation potential often cause early disease onset and premature death. Amyloid diseases predominantly involve the aggregation of a specific protein; however a range of other components including additional proteins and carbohydrates may become incorporated into deposits when formed *in vivo* (Dobson, 1999).

The effects of amyloidosis *in vivo* vary widely depending on the nature of the misfolded protein and the location of the aggregation. Systemic amyloidosis, the deposition of protein throughout the body can cause medical problems due to the volume of material in tissues (Falk *et al.*, 1997). Two important examples of this are human cystatin C amyloid angiopathy, in which deposition of amyloid around blood vessels can lead to circulation problems and stroke (Palsdottir *et al.*, 1988) and haemodialysis-related amyloidosis in which β2-microglobulin forms plaques in the skeletal joints of individuals on dialysis. In most neurodegenerative diseases however, the primary symptoms appear to be the result of 'toxic gain of function', which is poorly understood (Dobson, 2003).

1.4 Amyloid Assembly

1.4.1 Protein Conformation Change

Virtually any peptide sequence can form amyloid fibre in appropriate conditions (Fandrich *et al.*, 2001). However, peptides and proteins which aggregate readily in neurodegenerative disease predominantly adopt unstructured conformations in their normal biological state (Chiti *et al.*, 2003), and are thus more likely to sample conformational states suitable for amyloidosis. It is suggested that amyloid forming proteins, if not natively unstructured become structurally destabilised before forming amyloid. Natively folded proteins therefore form amyloid through an unfolded, or at least transient locally misfolded state. Local misfolding can be self-catalysed so that a misfolded conformer induces similar misfolding in the protein population in a manner identical to prion propagation (Dhulesia et al., 2010, Eichner et al., 2011, Jahn et al., 2006). Amyloid fibres then elongate from seeds by self-templated growth.

Within a protein or peptide, short sequences may form 'Aggregation prone regions' which share similar physical characteristics responsible for amyloidogenesis. These 'amyloid hotspots' are often regions of instability frequently containing a high percentage of aromatic and hydrophobic side chains (Maurer-Stroh *et al.*, 2010), negatively charged amino acids and a high propensity to convert from α-helices to β-sheets (Chiti *et al.*, 2003).

It has also been suggested that there are 'aggregation gatekeepers', which reduce aggregation by opposing nucleation of aggregates. These 'gatekeeper' peptides disrupt with the repulsive charge effect (arginine, lysine, aspartate, and glutamate residues), the entropic penalty of arginine and lysine or incompatibility with β-sheet structure in the case of proline (Reumers *et al.*, 2009). Proline is highly effective as a β-sheet 'breaker' and proline isomerisation, a typically slow process, is suggested be required for amyloidogenesis in several proteins, including β2-microglobulin (Eichner *et al.*, 2011) and cystatin B (Kokalj et al., 2007). Additionally, chaperone specificity for certain 'aggregation gatekeepers' has been shown whereby chaperone proteins recognise patterns of charged residues followed by hydrophobic regions (Reumers *et al.*, 2009).

Despite the potential that amyloid fibre is a structure that can be assumed by any polypeptide, amyloidogenesis is extremely specific with only very similar proteins forming fibres with one another. In some cases, a single residue change is enough to prevent the co-fibrilisation of two proteins (Hill et al., 1997), although in larger proteins only a small amyloidogenic region needs to be conserved (Sambashivan et al., 2005). It has been determined that six residues peptides represent the minimum length required for amyloid formation (Fernandez-Escamilla et al., 2004).

1.4.2 Amyloid Formation Events

Typical 'Visible' Microscopic Pathway Determined by EM and AFM

1. Formation of spherical particles (HOs) which can range in diameter from 3 nm to 25 nm or the production of amorphous aggregates (Glabe, 2004). This structural rearrangement may in some cases utilise specific structural transitions such as domain swapping.
2. Small 'bead-like' structures are sometimes observed and lead to formation of protofibrils. These structures are commonly short, thin, and sometimes curly (Dobson, 2003).
3. Mature amyloid fibres formation

1.4.3 Kinetics and Thermodynamics of Amyloid Formation

In protein misfolding diseases linked with amyloid fibre and oligomer formation, it is important to consider the kinetic and thermodynamic underpinnings (see figure 1.4.1). Most proteins in physiological conditions fold into their native state because the free energy of activation ($\Delta G^{\ddagger}$ which dictates the folding rate) separates the unfolded and transition state and is surmountable. Therefore the folded state is substantially more stable than the unfolded state. The difference in free energy, ΔG, dictates the relative populations of folded and unfolded proteins. For aggregation-prone proteins, the misfolded oligomeric state can however be more stable than the native folded/unfolded conformation in certain conditions and thus the oligomeric state becomes populated. $\Delta G^{\ddagger *}$ determines the rate of conversion to oligomeric forms, whilst ΔG^{*} dictates the relative populations in competition with the intrinsic clearance mechanisms which remove the misfolded protein (Cohen and Kelly, 2003).

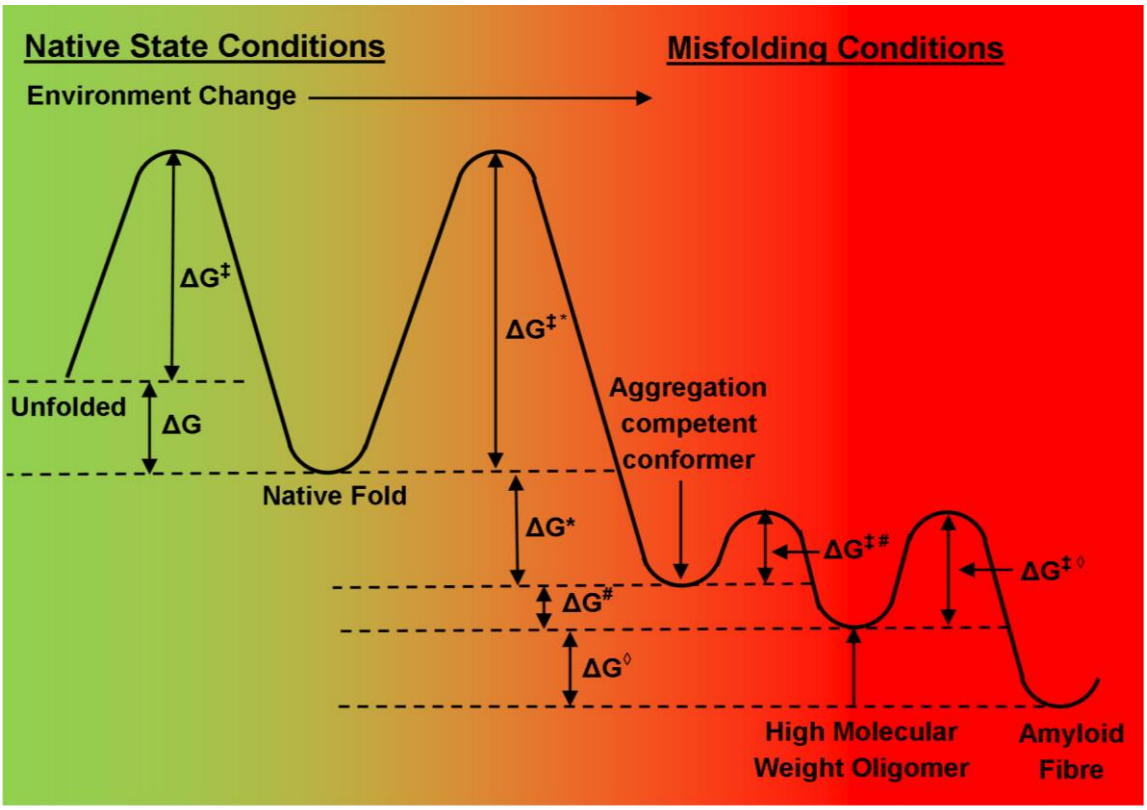


Figure 1.4.1 Thermodynamics of Protein Folding and Misfolding

In physiological conditions the native state is favoured, however in 'misfolding conditions', aggregation/multimeric/fibrous states may be energetically favoured over the natively folded state. $\Delta G^{\ddagger}$ dictates the folding rate whilst the difference in free energy, ΔG, dictates the relative populations of folded and unfolded proteins. $\Delta G^{\ddagger *}$ determines the rate of conversion to aggregation competent forms, whilst ΔG^{*} dictates the relative populations of folded and unfolded states. $\Delta G^{\ddagger\#}$ dictates the rate of high molecular weight oligomeric structure formation, whilst the $\Delta G^{\#}$ dictates the relative population. $\Delta G^{\ddagger \Diamond}$ determines the rate of conversion to amyloid fibre structures, whilst $\Delta G^{* \Diamond}$ dictates the quantity of amyloid fibre (Cohen and Kelly, 2003).

Amyloidogenesis typically exhibits a lag phase, whereby little fibre formation occurs, which is subsequently followed by an exponential growth phase until a thermodynamic equilibrium is reached (see figure 1.4.2). The amyloidogenesis lag phase can be reduced or ablated by the addition of 'seeds' which when observed by EM, are mostly small fragments of amyloid fibre usually generated by sonication of mature fibres. In addition, the fibre seed conformation controls product conformation in a cross-seeding reaction. If a fibre seed contains a small percentage of a second conformation, and if that conformation has a favourable elongation rate, then the minor fibre conformation can dominate the population after a few cycles of seeded growth (Antzutkin *et al.*, 2002)

The seeding phenomena strongly suggests a nucleated growth mechanism whereby the rate limiting step depends upon the existence of a rare perhaps stochastically generated nucleus (Hortschansky *et al.*, 2005). The rarity of the nucleus could be explained if nucleus formation is thermodynamically unfavourable where the resultant intermolecular interactions do not outweigh the entropic cost of association. However, the exponential growth phase, in the simplest case, could be due to further addition of monomer to the nucleus becoming thermodynamically favourable by increased intermolecular interactions and therefore overcoming the entropic barrier (Jarrett and Lansbury Jr., 1993).

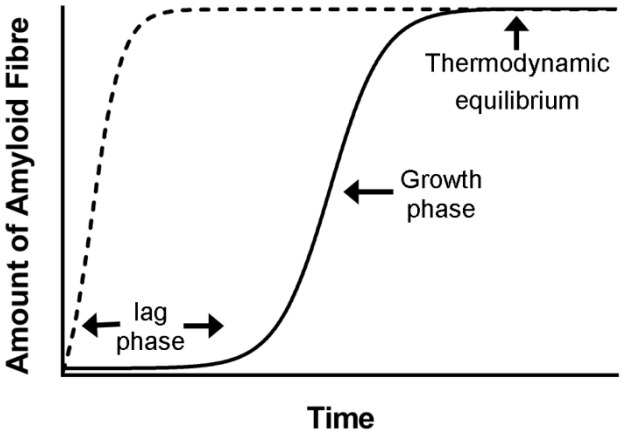


Figure 1.4.2 Typical Fibrilisation Kinetics

Amyloid fibre formation typically follows a kinetic profile of a nucleation-dependent process which exhibits a lag phase followed by an exponential growth period and subsequent thermodynamic equilibrium **(solid line)**. At a high concentration or very favourable environmental conditions, nucleation is so rapid no lag phase is observed **(dashed line)**. The addition of a nucleus or seed eliminates the nucleation step and therefore also ablates the lag phase (Jarrett and Lansbury Jr., 1993).

Amyloidogenesis is complex with several posited steps, figure 1.4.3 shows a suggested basic scheme of the $A\beta_{1-42}$ amyloidosis assembly process as an example. The protein species inhabiting the lag and elongation phase are not clearly defined and may vary for different proteins. High-resolution atomic force microscopy (AFM), low-energy electron-spray ionisation ion mobility tandem mass spectroscopy (ESI-IMS-MS) and analytical ultracentrifugation (AUC) identifies a continuum of aggregated protein states starting as low as dimers for both Aβ and β2-Microglobulin (β2M), although equally other studies have found oligomers to be relatively homogeneous (Kodali and Wetzel, 2007, Ostner et al., 2013).

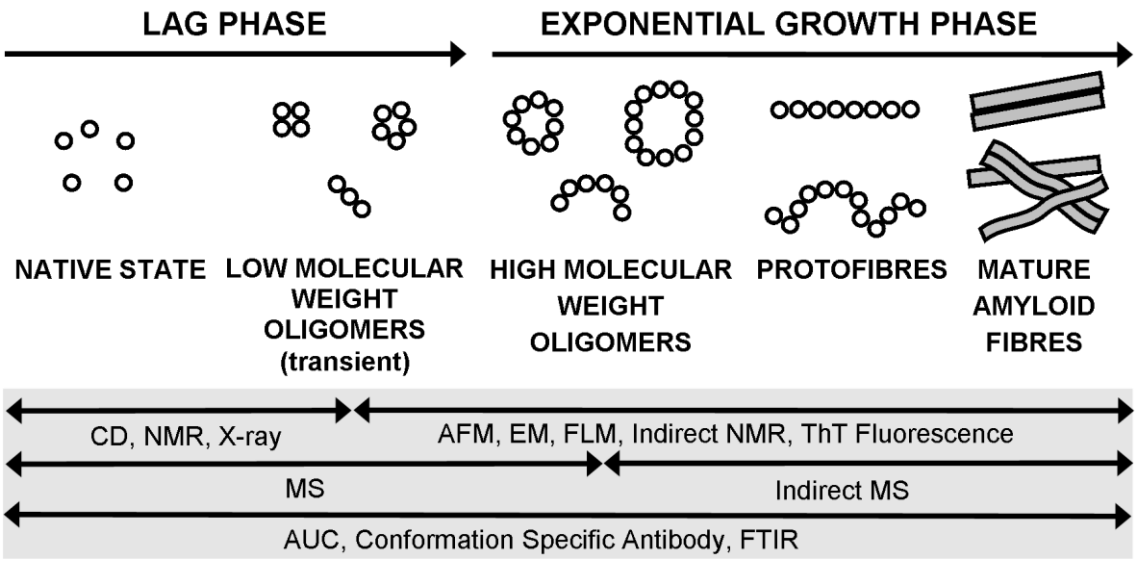


Figure 1.4.3 Schematic of the Amyloidosis Assembly Process

This suggested basic scheme describes the classes of structure possibly on pathway to amyloid fibre formation. Suitable technique ranges for morphological and/or quantitative evaluation of the prevailing species are shown in the grey box. X-ray crystallography (X-ray) and nuclear magnetic resonance spectroscopy (NMR) can be used to determine the native folded conformation, with the latter reporting on structural dynamics. Circular dichroism (CD) and fourier transform infra-red spectroscopy (FTIR) is used to determine the secondary structure content, whilst mass spectroscopy (MS) can determine the stoichiometry and overall morphology of monomers to soluble oligomers. Atomic force microscopy (AFM) and electron microscopy (EM) can be used to observe small oligomeric and fibrous structures, whilst fluorescence light microscopy (FLM) can be used to observe larger oligomers and fibres in real time. Analytical ultracentrifugation (AUC) can be used to observe the stoichiometry of the structures formed assuming specific structural states. Conformation specific antibodies can be used to report on the type of structures present when compared with other techniques. Figure based on one presented in (Bartolini *et al.*, 2011) with permission.

1.4.4 Models of Amyloid Formation

There are at least four prevalent mechanisms proposed for amyloid formation which are shown in figure 1.4.4 and discussed in (Kelly, 2000, Serio *et al.*, 2000, Murphy, 2007), with no mechanism being conclusively proven, although all follow the requirement for a nucleus and/or conformational conversion process. These mechanisms include the template assembly model (TA) (Griffith, 1967), Monomer Directed Conversion (MDC) (Prusiner, 1982), the nucleated polymerisation model (NP) (Jarrett and Lansbury Jr., 1993) and nucleated conformational conversion (NCC) (Serio *et al.*, 2000).

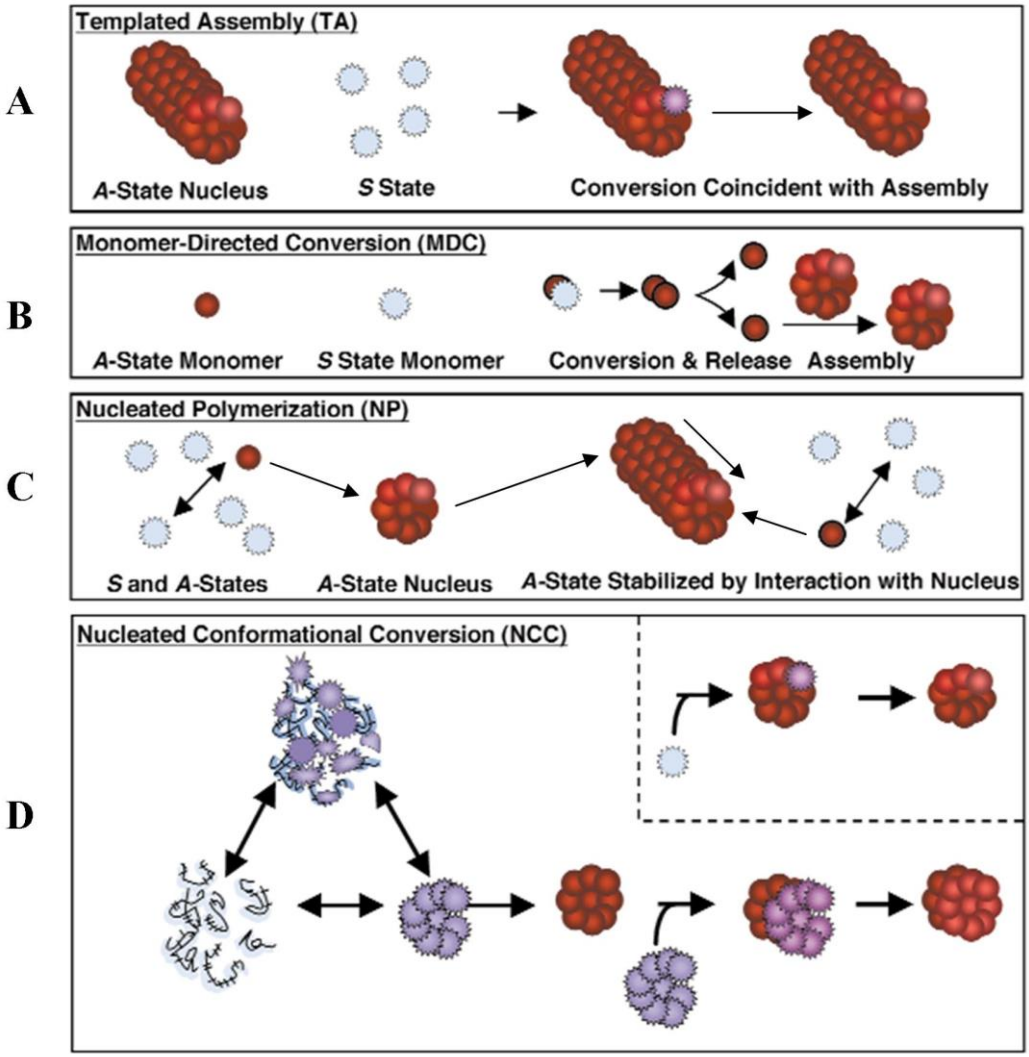


Figure 1.4.4 Prevalent Models of Amyloidogenesis

Smooth circles represent the A-state protein, which is analogous to the conformation adopted in amyloid fibre; jagged circles identify the soluble S-state protein; and open circles denote possible conformational heterogeneity. **(A)** Templated assembly. **(B)** Monomer-directed conversion. **(C)** Nucleated polymerisation **(D)** Nucleated conformational conversion. Figure modified and reprint permission obtained from Macmillan Publishers Ltd: [Nature Structural Biology] (Kelly, 2000).

The templated assembly model involves the soluble S-state binding to a pre-assembled nucleus followed by a rate-determining structural conversion and addition to the end of the fibre (Griffith, 1967). This model predicts that the lag phase should decrease proportionally to an increase in soluble protein concentration whilst the rate of fibre elongation should remain unaffected. The lag phase should also be reduced upon the addition of a seed.

The monomer-directed conversion model proposes that a monomeric peptide adopts a conformation analogous to that of the fibre state, termed the A-state. In the rate-determining step, the A and S-state bind and the S-state monomer is converted to the A-state. The resultant dimer then dissociates and the A-state dimers then rapidly combine with the end of the growing fibre. This model predicts that the presence of a seed should not affect the rate or reduce the lag phase of the reaction since the rate limited step occurs in the soluble protein (Prusiner, 1982).

In the nucleated polymerisation model, an equilibrium between the A and S-states occurs and with the assembly competent A-state forming the rate-limiting A-state nucleus. This A-state nucleus then rapidly forms a fibre by the addition of A-state monomers being stabilised and incorporated into the growing end of the fibre. In this model an increase in soluble protein concentration should accelerate the rate of fibre assembly and decrease the lag phase. The addition of a seed should abolish the time required for seed formation and so remove the lag phase (Jarrett and Lansbury Jr., 1993). Amyloid fibre assembly *in vitro* commonly displays a lag phase prior to rapid assembly. In many cases the addition of preformed aggregates eliminates the lag phase in a process known as ‘seeding’. In addition, this mechanism is common in nature with examples including actin polymerisation as well as microtubule and flagellum assembly and is therefore one of the current preferred assembly models (Jarrett and Lansbury Jr., 1993).

The nucleated conformation conversion model posits that conformational rearrangements within structurally dynamic oligomers induces the formation of a nucleus, which once formed interacts with other structurally flexible oligomers composed of a distribution of subunits and forms a group of subunits which can associate with similar structures or fibre to form extended fibrous structures. This model

rationalises low level concentration dependence, because at higher protein concentrations, oligomers form higher molecular weight complexes which are assembly incompetent. This mechanism can furthermore describe why an increase in seed concentration may produce minimal rate enhancements, because the limitation is oligomer concentration, not fibre ends (Serio *et al.*, 2000). At the time of writing no consensus has been achieved with amyloid formation models, in part because the intermediate and oligomeric structures remain elusive and poorly understood kinetically and structurally.

1.5 Amyloid Fibre Structure

1.5.1 Introduction

Most amyloid fibres by EM and AFM exhibit similar morphologies; long, unbranched and often twisted structures with diameters typically between 4-20 nm and several micrometers long (Fandrich, 2012, Kodali and Wetzel, 2007). Amyloid fibres also exhibit a characteristic 'cross-β' x-ray fibre diffraction pattern revealing that the organised core structure is composed of β-sheets whose strands run perpendicular to the fibre axis (see figure 1.5.8) (Blake and Serpell, 1996). It has been proposed that the fibre core is primarily stabilised with a large number of hydrogen bonds between the polypeptide backbones (Dobson, 1999). However, the side chains must also contribute significantly to the specificity of aggregation and stability of the fibre (Huff *et al.*, 2003). This latter idea forms the basis of prediction programmes that can estimate a sequence-based propensity to form amyloid. This in turn feeds into the design of biomolecules with improved solubilities.

Twisted right handed β-sheets are in a lower energy conformation than left handed and flat β-sheets (Chothia, 1973), thus are likely to be a preferred morphology. Furthermore, evidence from Fourier transform infrared spectroscopy (FTIR) of numerous amyloid fibres (including cystatins) along with EM reconstructions and AFM measurements suggest that amyloid fibre β-sheets are flattened (Zandomeneghi *et al.*, 2004).

Parallel in-register β-sheets have been observed by solid state NMR (SSNMR) in $A\beta_{1-40}$ (Petkova *et al.*, 2002) and $A\beta_{1-42}$ fibres (Antzutkin *et al.*, 2002), amylin fibres (Luca *et al.*, 2007), various yeast prion fibres (Baxa et al., 2007, Shewmaker et al., 2006), HET-s (218-289) fibres (Wasmer *et al.*, 2008), and PrP peptide fibres (Walsh *et al.*, 2009). Parallel in-register β-sheet aligns hydrophobic residues plus glutamine and asparagine residues with themselves, potentially maximising favourable hydrophobic interactions and polar-zipper interactions, respectively, for any amino acid sequence (Perutz *et al.*, 1994). Anti-parallel β-sheets have also been found in amyloid fibres but, until recently, only in fibres formed by short peptides that contain a single hydrophobic segment in a single β-strand (Tycko, 2011, Kammerer et al., 2004).

A further complication is observed structural polymorphism of amyloid fibres, in which the same fibre preparation or subtle changes in growth condition produces distinct structural variation, observable in EM and SSNMR (Petkova *et al.*, 2005). As an example, $A\beta_{1-40}$ quiescently generated striated-ribbon and agitated twisted pair fibre structures are reviewed in section 1.5.3 and figure 1.5.1. Amyloid fibre polymorphism may exhibit itself in varying quaternary structure, protofilament self-association and substructure (e.g. side-chain packing within the protofilament). The kinetic origins of this polymorphism remain unknown (Kodali and Wetzel, 2007, Eichner *et al.*, 2011).

A basic common structure is however, likely adopted because universal amyloid specific antibodies react with generic epitopes on all types of amyloid fibre and soluble oligomer. These antibodies specifically bind amyloid but not native protein, thus providing further proof of non-native structure. It therefore seems likely that this class of antibody recognises a specific conformational epitope; a turn or hydrogen bonding motif on the polypeptide backbone (Glabe, 2004). This antibody conformation-specificity is further supported by consistent amyloid binding of accessory molecules, including serum amyloid P component, proteoglycans/glycosaminoglycans and apolipoprotein E (Huff *et al.*, 2003).

1.5.2 Methods of Amyloid Fibre Structure Determination

Standard high resolution experimental techniques such as X-ray crystallography and traditional solution NMR fails to yield structural information on the large heterogeneous fibre structures. Therefore a combination of biophysical and biochemical techniques is required to develop structural constraints which lead to refinement of the structural models. These techniques include X-ray fibre diffraction, solid-state NMR (SSNMR), with complementary information provided by hydrogen-deuterium (H/D) exchange, electron paramagnetic resonance spectroscopy (EPR), limited proteolysis (LP), site-directed mutagenesis, transmission electron microscopy (TEM) and atomic force microscopy (AFM). The high degree of secondary structure found in amyloid fibres also permits study by fourier transform infra-red spectroscopy (FTIR) and circular dichroism spectroscopy (CD). As of writing however, no atomic resolution fibre model has been reliably constrained for a natively folded globular protein such as the all-β-sheet proteins β2-microglobulin and transthyretin.

Dye Binding Assays

When stained with Congo-red, amyloid fibres exhibit green birefringence when viewed under cross-polarised light (Sipe and Cohen, 2000). Amyloid fibre also binds thioflavin T (ThT) and thioflavin S, which when bound exhibit a shift in fluorescence emission maxima. Congo-red birefringence and ThT fluorescence shift most likely results from alignment of the dye molecules when bound to the fibre. The fibre must display a common epitope and the dye binding most likely occurs through specific associative binding through the extended β-sheet or intercalation (Glenner *et al.*, 1972, Wolfe *et al.*, 2010).

Microscopy Techniques

TEM and AFM are often used singly and in conjunction to provide complementary information on morphology and dimensions of amyloid fibre and their intermediates. TEM and AFM have also been used to follow the formation of oligomeric and fibrous structures and fibrilisation by discontinuous and continuous methods (Goldsbury *et al.*, 1999). A specialised scanning transmission electron microscopy (STEM) can be used to make mass per unit length (MPL) measurements, e.g. χ kDa/nm and is used to determine the spacing of repeating units in amyloid fibre (Wall *et al.*, 2008). Therefore this measurement has been performed on several amyloid fibre forming proteins and forms a major part in amyloid fibre structural modelling (Petkova *et al.*, 2002). Cryo-electron microscopy has been used to produce 3D reconstructions of amyloid fibres by particle averaging and although it doesn't provide atomic resolution it strongly aids structural modelling (White et al., 2009, Serpell et al., 2000, Jimenez et al., 1999).

The crystal structure of a designed peptide incorporating motifs found in a large number of known amyloidogenic peptides has been solved from small needle-like crystals by electron diffraction. In this instance X-ray diffraction was poor and the subsidiary technique was then used (Makin *et al.*, 2005). This KFFEAAAKKFFE peptide was shown to be stabilised in an anti-parallel configuration by π-π stacking interactions between phenylalanine side chains, charge balancing and inter-chain hydrogen bonding. The crystallisation was presumably made possible due to the compact, homogeneous nature of the fibres which have no large mobile regions which would impede crystal packing. However, the interpretation of crystal structures relies on a careful evaluation of how closely the crystals reflect the structural organisation within the fibre states (Marshall *et al.*, 2010).

X-Ray Fibre Diffraction

Most of the initial detailed amyloid fibre structural information was determined by X-ray fibre diffraction (Blake and Serpell, 1996). Amyloid fibre produces diagnostic reflections in two positions: meridional reflections of 4.7Å from repeating β-sheet along the fibre axis and 10 Å equatorial bands generated from electron density perpendicular to the fibre axis, which is indicative of cross-β structure (Geddes *et al.*, 1968, Blake and Serpell, 1996). The meridional reflections are produced by the repeated interstrand hydrogen bond spacing between the β-sheets whilst the orthogonal equatorial reflections

are generated by intersheet interactions. Well aligned fibre preparations can also be used to determine further structural information, such as helical pitch/helical repeating unit, accurate overall shapes, dimensions and even the presence of a hydrated fibre core (Blake and Serpell, 1996, McDonald *et al.*, 2012). Fibre diffraction can however, only image these molecular details in exceptionally favourable cases (Chandrasekaran and Stubbs, 2006, Stubbs, 1999).

X-Ray Crystallography

A range of X-ray microcrystal structures have been determined from short known amyloidogenic peptide sequences. Sequences such as GNNQQNY form both fibres in the traditional sense and microcrystals which form in identical conditions, exhibit small fibrous appendages and in some cases appear composed of several aligned, nanometer-sized blocks which may be the result of individual mature fibre aggregation into the crystalline form. Furthermore, these crystals are primarily composed of β-strands stacked perpendicular to a single axis and produce similar inter-sheet/inter-strand spacings to that of X-ray fibre diffraction patterns (Nelson and Eisenberg, 2006, Marshall et al., 2011).

The resultant structures are cross-β spines of two β-sheet amide stacks tightly packed complementary to one another in a dry interface with wet interfaces between adjacent cross-β spines. A study with 30 peptides derived from Aβ, tau, PrP prion, insulin, islet amyloid polypeptide, lysozyme, myoglobin, α-synuclein and β_2-microglobulin all showed this similar overall topology. 5 out of 8 structural arrangements were observed distinguished by (i) strand orientation in parallel or anti-parallel, (ii) β-sheet packing with the same surface adjacent to one another (‘face-face’) or opposing (‘face-back’) and (iii) β-sheet orientation parallel (‘up-up’) or anti-parallel (‘up-down’) (Serpell, 2000). In my view these structures are unlikely in many cases to directly match the structures formed by the full length polypeptide due to the array of feasible interactions available to a larger molecule. However, these structures highlight the possible variations in orientation and provide a clear indication of potential polymorphism.

Spectroscopic Techniques

Circular dichroism (CD) and fourier transform infra-red (FTIR) spectroscopy techniques have been used to estimate the secondary structure content of amyloid fibres and using these techniques it has been shown that a wide variety of proteins adopt a high β-sheet content in amyloid fibres. FTIR spectra in particular show a strong peak characteristic of large-scale β-structures, which appears as fibres are formed (Zandomeneghi *et al.*, 2004).

Limited Proteolysis

To date, limited proteolysis has been exploited to analyse amyloid fibre formation in Aβ peptide (Kheterpal *et al.*, 2001), α-synuclein (Miake *et al.*, 2002), HET-s (Balguerie *et al.*, 2003a), Ure2p (Baxa *et al.*, 2003), PI3-SH3 (de Laureto *et al.*, 2003), bovine α-lactalbumin (de Laureto *et al.*, 2005), β_2-microglobulin (Myers *et al.*, 2006) and lysozyme (Frare *et al.*, 2006). Limited proteolysis occurs primarily at sites with high chain mobility and rarely within regular secondary structure. Therefore, conformational parameters such as accessibility, mobility and protrusion correlate with limited proteolysis data (Hubbard, 1998).

Hydrogen/Deuterium Exchange

Quenched hydrogen/deuterium (H/D) exchange (Hoshino et al., 2002, Alexandrescu, 2001) experiments on amyloid fibres have been used to map the solvent-accessible regions, thus inferring some structural information. This method observes the rate at which amide protons exchange with deuterons of the peptide backbone (or vice versa). This rate is dependent on solvent accessibility of the amide protons; amides that are solvent exposed exchange rapidly, whilst those that are incorporated within structured regions of the protein or involved in stable hydrogen bonding networks exchange at a slower rate and are thus determined to be 'protected' from exchange. The exchange process is quenched by freezing and drying the samples followed by rapidly dissociating the fibres in organic solvent such as dimethyl-sulphoxide (DMSO) and analysis by NMR or MS. NMR is the typically preferred method of observation because it provides residue specific information from an assigned 2D transverse relaxation optimised spectroscopy or heteronuclear single quantum coherence (TROSY or HSQC) NMR spectrum (Pervushin *et al.*, 1997).

MS can be a complementary approach as this method provides information on the populations of partially and fully exchanged species, however cannot provide residue specific information. MS therefore provides information on the homogeneity of the sample. H/D exchange has also been used to demonstrate the dynamic nature of amyloid fibre, whereby molecules were shown to dissociate and re-associate with the fibres (Carulla *et al.*, 2005, Sanchez *et al.*, 2011). H/D exchange could also be used to ascertain a change in amyloid fibre dissociation rate or binding event and monitor how the structure is perturbed by the binding of novel therapeutic agents.

Solid-State Nuclear Magnetic Resonance

SSNMR is the only technique which can obtain atomic resolution in non-crystalline solids and has therefore been used to measure specific inter-residue distance, torsion angles and dynamics of peptide and protein molecules within amyloid fibres. To generate structural models of the amyloid fibres the SSNMR constraints must be combined with data from other techniques such as X-ray fibre diffraction, MPL measurements and EM (Petkova et al., 2002, Antzutkin et al., 2002, Shewmaker et al., 2006, Baxa et al., 2007, Luca et al., 2007, Paravastu et al., 2008, Fawzi et al., 2011, McDonald et al., 2012, Wasmer et al., 2008). The contribution of this technique to current understanding of amyloid fibre structure is discussed further in the following sections.

1.5.3 Amyloid β Peptide

Amyloid β $_{1\text{-}40}$

SSNMR combined with other techniques such as mass-per-unit-length (MPL) measurements has produced structural constraints for $A\beta_{1\text{-}40}$, sufficient to produce molecular level modelling, with only small unresolved variations in packing interface alignment (Petkova *et al.*, 2002, Petkova *et al.*, 2005, Petkova *et al.*, 2006, Paravastu *et al.*, 2008). In both the quiescently generated striated-ribbon and agitated twisted pair $A\beta_{1\text{-}40}$ fibre models the first 9 residues are structurally disordered, whilst residues 10-22 and 30-40 form in-register parallel β-sheets only through intermolecular hydrogen bonding. Residues 23-29 form a tight loop, linking the two β-sheets into a β-arc, often described as a 'hairpin'. The twisted pair $A\beta_{1\text{-}40}$ fibre is composed of three $A\beta_{1\text{-}40}$ monomer arcs constructed around a central hydrated hydrophobic channel in a triangular cross section with three-fold symmetry (see figure 1.5.1) (Paravastu *et al.*, 2008, McDonald *et al.*, 2012). The striated-ribbon $A\beta_{1\text{-}40}$ fibre instead has a two-fold symmetry and is composed of two $A\beta_{1\text{-}40}$ monomer arcs per protofibril (Petkova *et al.*, 2002). Limited proteolysis with trypsin and chymotrypsin proved relatively inconclusive with hydrolysis sites throughout the peptide. This is perhaps a result of an observed equilibrium state between monomer/soluble peptide and amyloid fibre, thus the results are possibly skewed by soluble unfolded peptide hydrolysis (Kheterpal *et al.*, 2001).

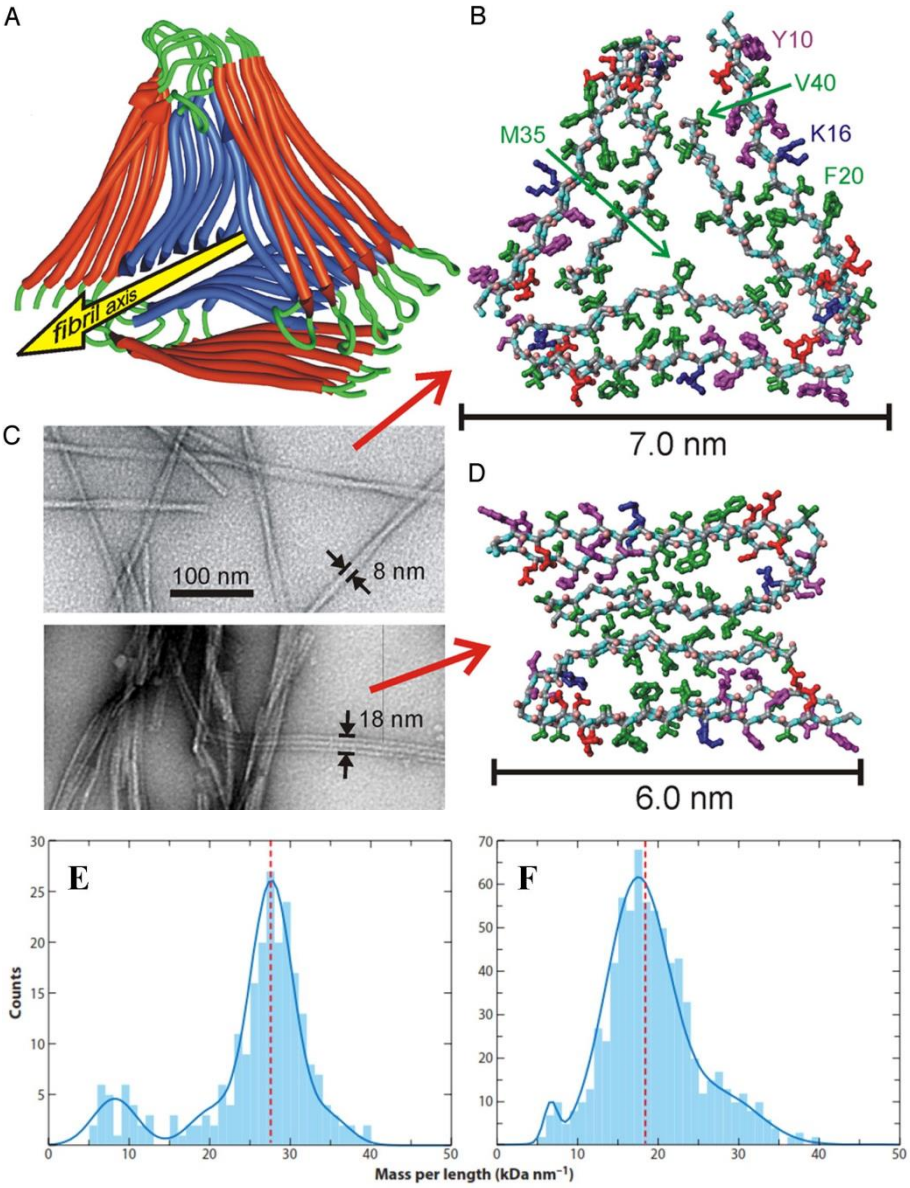


Figure 1.5.1 $A\beta_{1\text{-}40}$ Amyloid Fibre Structural Models

Experimentally based structural models of striated-ribbon and agitated twisted pair $A\beta_{1\text{-}40}$ fibre. Unstructured N-terminal residues 1–8 are omitted. **(A)** Ribbon representation of the lowest-energy model for fibres with the twisted morphology. Modelling calculations assume three-fold symmetry, consistent with MPL data and SSNMR spectra, and are constrained by secondary, tertiary, and quaternary structural data from solid state NMR. **(B)** Atomic representation, viewed down the fibre axis. Hydrophobic, polar, negatively-charged, and positively charged amino acid sidechains are green, magenta, red, and blue, respectively. Backbone nitrogen and carbonyl oxygen atoms are cyan and pink. **(C)** Comparison of twisted (Upper) and striated ribbon (Lower) fibre morphologies from negatively stained TEM. **(D)** Atomic representation of a striated ribbon fibre (Paravastu *et al.*, 2008). **(E and F)** Mass-per-unit-length (MPL) histograms of twisted fibre (E) and striated fibre (F). The red dashed lines indicated predicted MPL values from the models. The solid blue lines are from multiple-Gaussian fits. Figure modified and reprinted from (Paravastu *et al.*, 2008, Tycko, 2011) with permission from Elsevier.

Amyloid β $_{1-42}$

Modelling of $A\beta_{1-42}$ amyloid fibre has been conducted using the (Petkova *et al.*, 2002) $A\beta_{1-40}$ SSNMR study β-sheet arrangement and combining structural constraints of $A\beta_{1-42}$ side-chain packing pairwise mutagenesis studies and H/D exchange NMR data (Luhrs *et al.*, 2005). The parallel in-register β-arc (hairpin) structure remains analogous to $A\beta_{1-40}$, although individual residue involvement is not identical. $A\beta_{1-42}$ residues 1-17 are disordered whilst residues 18-26 and 31-41 form parallel β-sheets linked with a loop formed from the intervening amino acids (see figure 1.5.2). This structure is stabilised by a proposed salt-bridge between K28 and D23 and hydrophobic interactions mediated by residues L17, F19 and A21 in β-sheet 1 and even-numbered residues of β-sheet 2. The model closely fits other experimental data such as H/D exchange NMR in which residues 12-25 and 28-42 are protected for 120 minutes. AFM of $A\beta_{1-42}$ fibres suggests an average height between nodules (generated by fibre twist) to be 3.5 nm, a width of ~4 nm and twist nodules 110 nm apart (Olofsson *et al.*, 2006).

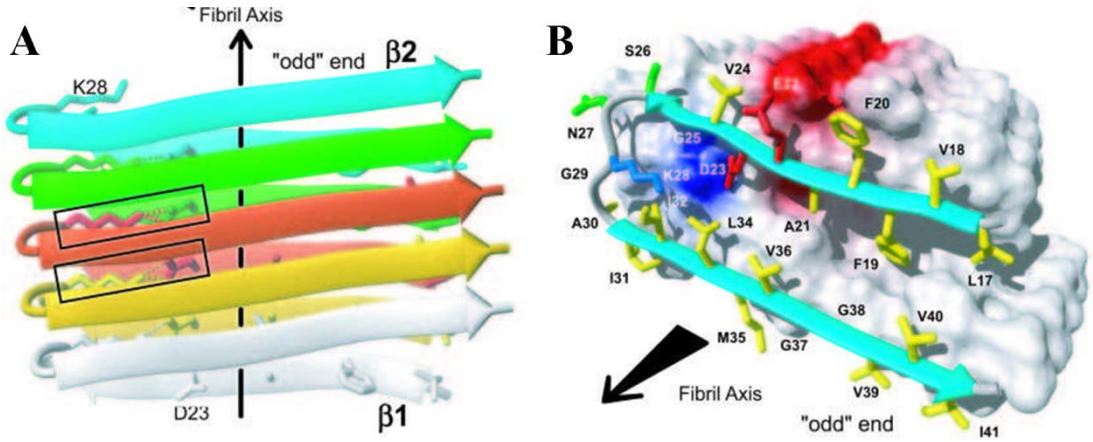


Figure 1.5.2 $A\beta_{1-42}$ Amyloid Fibre Structural Model

Ribbon and space filling model showing the amyloid fibre core β-arc structure of $A\beta_{1-42}$ residues 17-42. **(A)** Side view of individual hairpin ribbons with illustrated inter-β-strand salt bridge between residues D23 and K28, indicated by dotted lines and framed by rectangles. **(B)** Top view of β-sheet arc (hairpin) motif. Hydrophobic, polar, negatively charged, and positively charged amino acid side chains are shown in yellow, green, red and blue, respectively and all others are shown in white (Luhrs *et al.*, 2005).

1.5.4 β_2-Microglobulin

The cryo-EM reconstruction of β_2-microglobulin (β_2m) revealed a high degree of complexity and two structural polymorphs shown in figure 1.5.3 (White *et al.*, 2009). Type A exhibits polarity, whereby both 'crescent half-fibres' (protofilaments) are orientated in the same direction along the fibre axis and type B is bipolar, with opposing protofilaments orientations. Scanning transmission EM (STEM) based mass-per-unit-length measurements (MPL) were used to deduce the molecular packing of β_2m fibres with type A and B structures consistent with 53±3 kDa/nm fibres and fibres with half the width (type C) matching a smaller population of fibres with an MPL of 27±3 kDa/nm. An interpretation of these values could be that type C are protofilaments, whilst type A and B are more populous mature fibres as this twisted morphology is more intimate than simple random lateral association.

An atomic level model which fits the EM density and biochemical and biophysics data has yet to be conclusively constrained, despite attempts at fitting domain swapped dimers, staggered native monomers and other models, all of which neither entirely fit the density nor match previously determined structural constraints. A model with pairs of protein subunits threaded through each 'crescent' with flexible hinge regions between the protofilaments is possible. However, a proffered model for the type A and B forms of ß$_2$m amyloid fibre is based on a dimer-of-dimers subunit packing, with each elliptical cylinder corresponding to two ß$_2$m monomers (White *et al.*, 2009). This model is perhaps favoured due to H/D exchange kinetics which suggests multiple environments for individual residues within the fibre structure (Yamaguchi *et al.*, 2004).

H/D exchange data from β_2m suggests that the N terminal β-strands (A and G) are less protected in mature amyloid fibres than the natively folded protein, thus extraneous from the fibre core. β-strands B-E on the other hand show a high degree of H/D exchange protection and are thus incorporated into the fibre core. The H/D exchange protection of β-strand F fibre was not reported for experimental reasons (Hoshino *et al.*, 2002). Subsequent experimental work identified that the C-terminal β-strand was protected by an amplitude of ~50%, and thus incorporated within the amyloid fibre core in at least half of the fibre composing molecules. This varied protection is also indicative of residue protection throughout the fibre. When H/D exchange was probed from hours up to 13 days, a biphasic exchange mechanism was observed with rapid and subsequent exceptionally slow exchange. This biphasic nature suggests multiple amino acid environments, perhaps due to inherent structural differences. These differences could be due to the helical twist observed in many fibres leading to a mixture of exposed and buried residues or lateral association/assembly of fibres/protofilaments which could also reduce solvent exposure (Yamaguchi *et al.*, 2004), or indeed varied structural motifs within structural 'sub-units'.

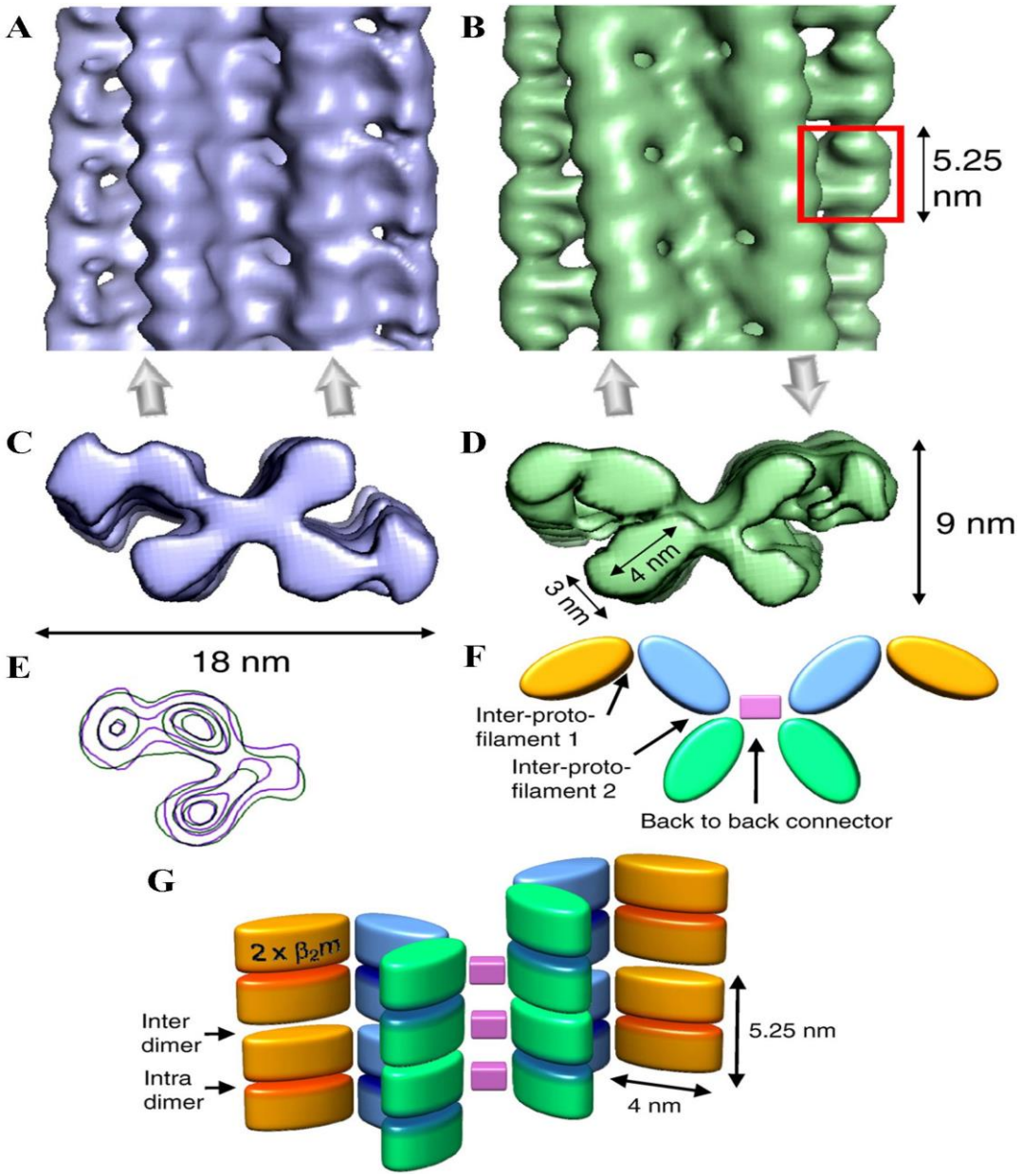


Figure 1.5.3 β2-Microglobulin EM Structural Reconstruction and Model

(A-D) Three-dimensional reconstructions of type A (left) and type B (right) forms of ß$_2$m amyloid fibres. A and B are side views down the fibre axis and C and D are equatorial views. One dimeric density unit is indicated by a red box in B. The directions of the half-fibril crescents are indicated by the arrows below A and B. **(E)** Superposed contour plots of the A type (lilac) and B type (green) fibre repeat units, showing that the two fibril types have the same underlying organisation that differs only in the orientation of the two stacks, either parallel or anti-parallel. **(F and G)** Schematic representation of the dimer-of-dimers subunit packing for the type B form of ß$_2$m amyloid fibre; each elliptical cylinder corresponds to two ß$_2$m monomers. The outermost (orange) protofilaments in the model are disordered in the map, so their full density is not reconstructed. Reprinted from (White *et al.*, 2009) with permission from Elsevier.

1.5.5 Cystatin B

The current cystatin B amyloid fibre model is based on EM fibre width measurements (Jensson *et al.*, 1987), H/D exchange data (Morgan *et al.*, 2008) and application of the simplest model based on the assumption of near-native β-sheet conformation, which is supported but not implicitly constrained by the current data available (see figure 1.5.4). From H/D exchange data, the cystatin B amyloid fibre core is likely to be comprised of β-strands 2, 3, 4 and 5, the C-terminal region and includes the loop between strands 2 and 3. Protection in this loop is indicative of the retention of the 3D domain-swap interface and hence the topology of the dimeric form is used in figure 1.5.4. The α-helix and the first ß-strand of the native structure are excluded from the fibre core.

The current cystatin B model, shown in figure 1.5.5, also requires the removal of β-bulges in β-strand 2 and 5 which then allows amide four NH and carbonyl groups to form intermolecular hydrogen bonds in the direction of the fibre axis. From H/D exchange protection it is suggested that β-strand 2 residues K39, V41, F43, S45 and partnered β-strand 5 residues T81, S83, Y85 and T87 contribute to hydrogen bonding in this interface. In this model the β-sheet face from which the central native α-helix has been removed is hydrophobic and the alternate face has a net positive charge at neutral pH due to lysine or arginine residues at position 40, 44, 68 and 89 (Morgan *et al.*, 2008). The current proposed model from the H/D exchange data is based upon the model posited in (Staniforth *et al.*, 2001), whereby the fibre composes a dimer of domain-swapped dimers stacked in the fibre. The model in figure 1.5.5, panel E and F is historical with the H/D exchange data not completely satisfied by this model and the AFM and EM measurements of cystatin B amyloid fibre (Zerovnik *et al.*, 2002a) suggest a ‘flat tape’ morphology which the model in panel F does not entirely fulfil.

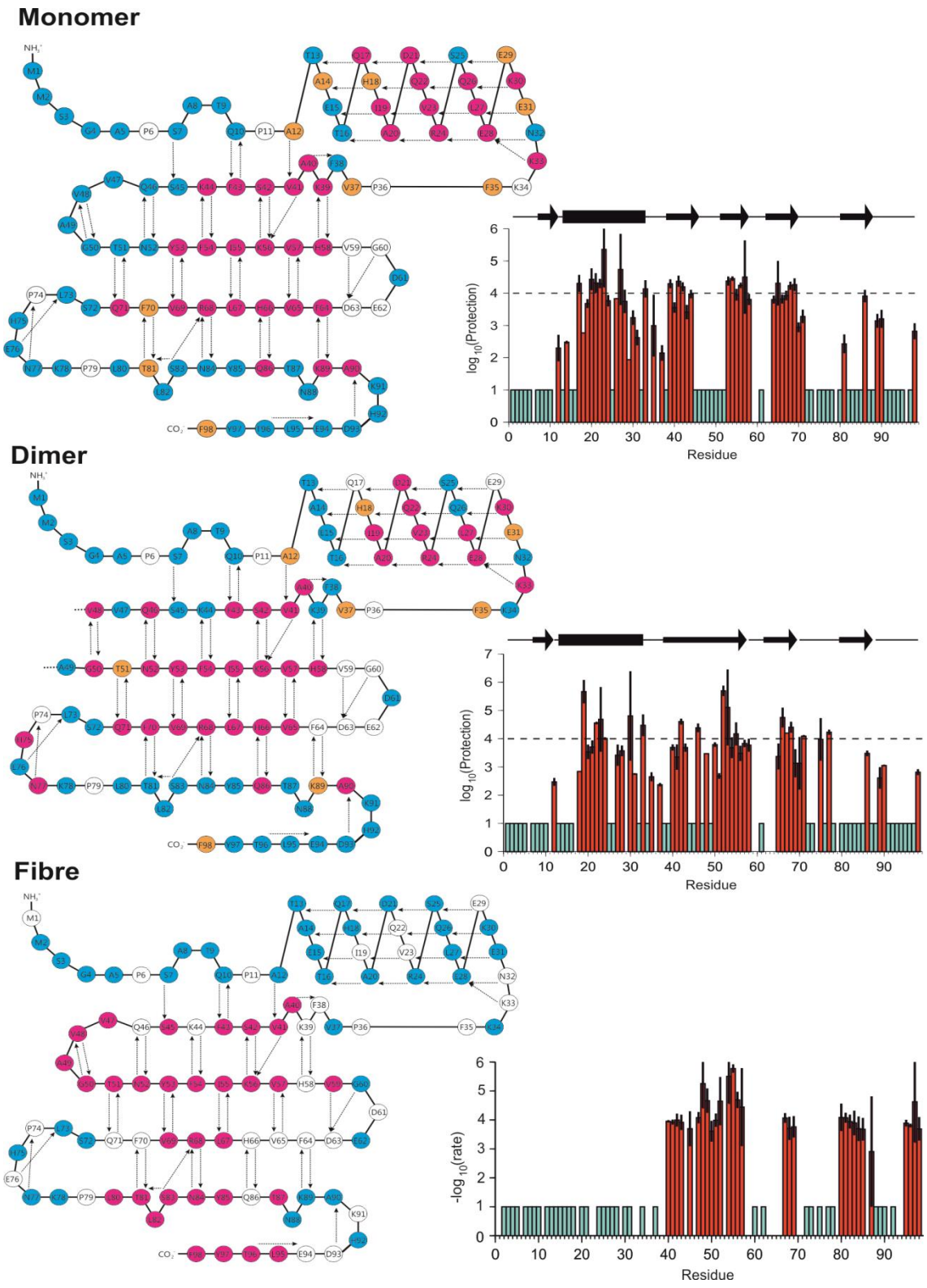


Figure 1.5.4 Hydrogen/Deuterium Exchange Topology of Cystatin B

Hydrogen/deuterium exchange protection mapped onto the native state topology diagrams of cystatin B (left) and protection values plot (right). Protected residues are coloured red, those in orange show partial protection, blue are unprotected and white residues are undetermined. The protection value plots includes above them rectangles denoting α-helices, β-sheets by arrows and lines for loops whilst the dotted line shows the arbitrary threshold required for protection to be assigned . Adapted from (Morgan *et al.*, 2008) with permission from Elsevier.

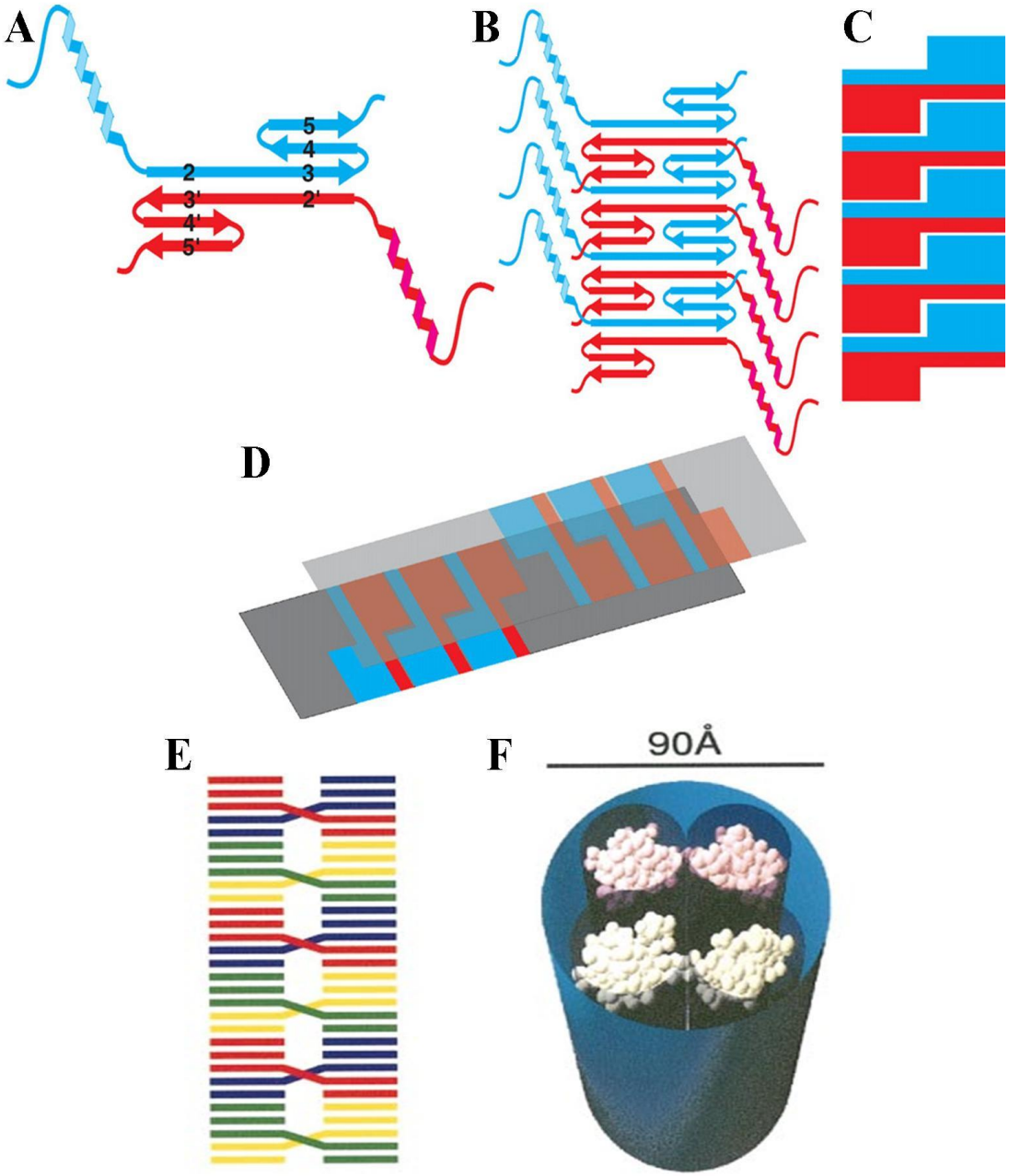


Figure 1.5.5 Cystatin B Amyloid Fibre Model

(A-D) Cartoon representation of the simplest model consistent with H/D exchange data. The α-helix represented is actually unfolded within the fibre and is only shown for clarity. **(A)** Secondary structure representation of the repeating structural moiety within the fibre core. **(B)** Stacking of the dimer-like amyloid core units to create a ribbon, consistent with the observed protection of strand 5 of the globular fold. **(C)** Simplified view of (B), showing the shape of the building blocks. **(D)** Two ribbons, represented as in (C), packed through the hydrophobic faces generated by the displacement of the α-helix to create a fibre model consistent with the overall EM cross-section dimensions of cystatin fibres from (Jensson *et al.*, 1987). **(E)** Schematic representation of cystatin B fibre produced from native-like 3D domain swapped dimers. **(F)** Space filling model of two domain-swapped dimers fitted into the EM cross-section of a 'generic amyloid fibre'. Images A-D are from (Morgan *et al.*, 2008) and reprinted with permission from Elsevier and panels E and F are reprinted by permission from Macmillan Publishers Ltd: [EMBO] (Staniforth *et al.*, 2001).

1.5.6 HET-s (218-289) Prion

The full amyloid fibre structural model formed by residues 218–289 of the *Podospora anserina* HET-s prion protein has been determined by SSNMR structural restraints and is shown in figure 1.5.6 (Wasmer *et al.,* 2008). HET-s (218-289) forms a parallel in-register left-handed β-solenoid with each molecule forming two β-strand helical windings from four β-strands, three of which contribute to form a triangular hydrophobic core. The triangular hydrophobic core is tightly packed, dry interface containing almost exclusively hydrophobic residues (Ala, Leu, Ile and Val). The only polar residues in the core are T233 and S273, which are suggested to form hydrogen bonds and stabilise the formation of a turn between β-strands 1 and 2. In contrast, all the charged residues face outwards and are located in highly solvent accessible β-arc regions. Three of the charged residues are stacked in the direction of the fibre axis, which compensates for the charge and allows salt-bridge formation. In this structure at least 23 hydrogen bonds, three salt bridges and two asparagines ladders have been determined. The cross-section of the structure is approximately circular, thus potentially explaining why no helical twist has been observed in electron micrographs (Wasmer *et al.*, 2008). The SSNMR model matches data from H/D exchange, water accessibility and mutant studies (Ritter *et al.,* 2005). This C-terminal proteinase K resistant core of prion fibres is unstructured in solution and forms infectious fibres *in vitro* (Balguerie *et al.*, 2003b).

The HET-s fibres have a triangular cross-section that is superficially similar to the threefold-symmetric $A\beta_{1\text{-}40}$ model, the details of the two models are however quite different. Specifically, (*i*) β-sheets in HET-s fibres contain both intramolecular and intermolecular backbone hydrogen bonds, whereas β-sheets in $A\beta_{1\text{-}40}$ fibres contain only intermolecular hydrogen bonds; (*ii*) each side of the triangular cross-section in HET-s fibres is formed by a single β-sheet layer, whereas each side of the triangular cross-section in threefold-symmetric $A\beta_{1\text{-}40}$ fibres is formed by two β-sheet layers; (*iii*) the MPL of HET-s fibres (40) is 0.5 monomer per 0.47 nm, compared with the 3 monomers per 0.47 nm of threefold-symmetric $A\beta_{1\text{-}40}$ fibres; (*iv*) HET-s fibres also have no long axis symmetry (Tycko, 2011).

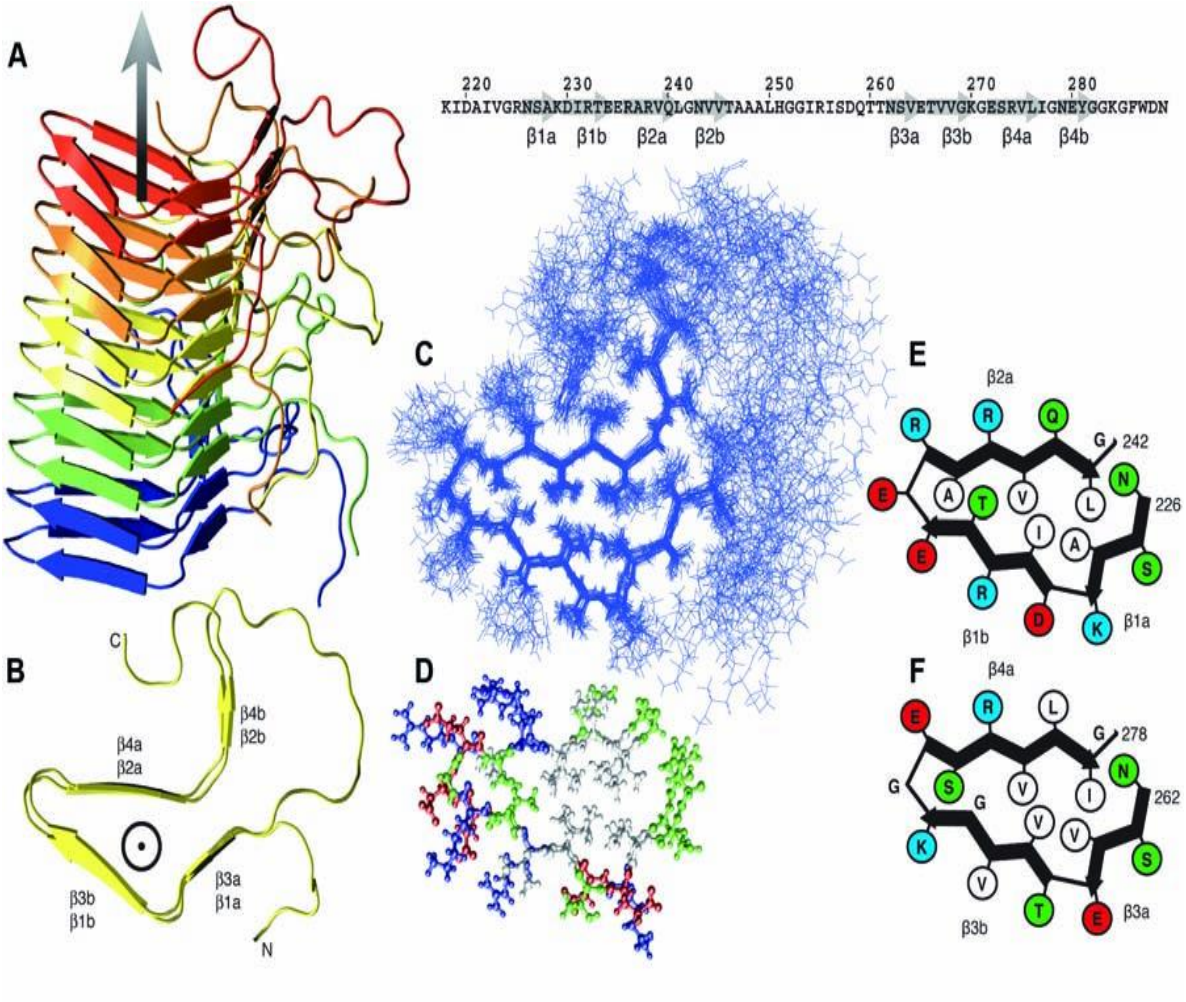


Fig. 1.5.6 Structure of the HET-s (218–289) Fibre

(A) Side view of the five central molecules of the lowest energy structure of HET-s (218–289) calculated from the NMR restraints. The fibre axis is indicated by an arrow. **(B)** Top view of the central molecule from (A). β-sheet β3 and β4 lie on top of β1 and β2, respectively. **(C)** NMR bundle: superposition on residues N226 to G242, N262 to G278 of the 20 lowest-energy structures of a total of 200 calculated HET-s structures. Only the solenoid core residues are shown. **(D)** Representation of the well-defined central core of the fibre (N226 to G242, N262 to G278). Hydrophobic residues are coloured white, acidic residues red, basic residues blue, and others green (lowest-energy structure). **(E and F)** Schematic representations of the two windings in (D): the first winding [N226 to G242, displayed in (E)] of the β-solenoid is located beneath the second one [N262 to G278, displayed in (F)]. Abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr. Figure from (Wasmer *et al.*, 2008). Reprinted with permission from AAAS.

1.5.7 Human Prion Protein PrP

One of the suggested human prion protein (hPrP) models, which is similar to the SSNMR model of HET-s (218-289), suggests the formation of a parallel left handed β-helix with α-helical regions located in extended loop regions which are excluded from the fibre core (see the figure below) (Govaerts *et al.*, 2004).

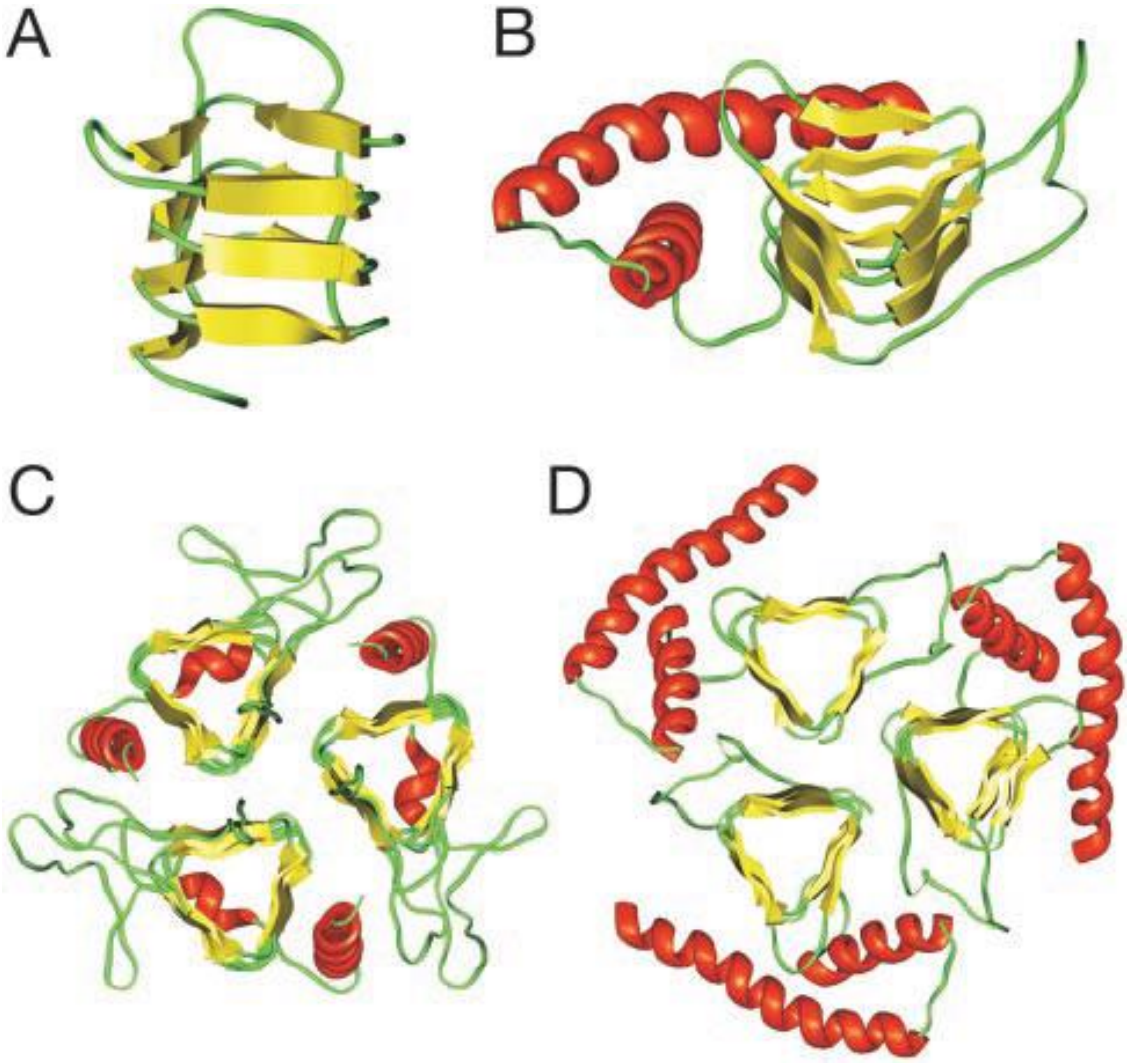


Figure 1.5.7 Modelling of Human Prion Protein Amyloid Fibre

***(*A)** β-helical model of the N-terminal hPrP. **(B)** Model of the hPrP monomer. **(C)** Crystal structure of trimeric carbonic anhydrase from *Methanosarcina thermophila* (PDB ID code 1THJ) which shows a similar structure thus proving the feasibility of such a model. (*D*) Trimeric complete model of hPrP. Models from (Govaerts *et al.*, 2004).

1.5.8 Transthyretin

A Transthyretin (TTR) fibre model from X-ray fibre diffraction and EM dimensions suggest a fibre composed of four parallel protofilaments of 50-60 Å which combine to produce a uniform fibre width of ~130Å. The protofibril structure shown in figure 1.5.8 is composed of four β-sheets, arranged perpendicular to the fibre axis, with a helical pitch of 115.5Å which relates to 24 β-strands per helical turn (Blake and Serpell, 1996).

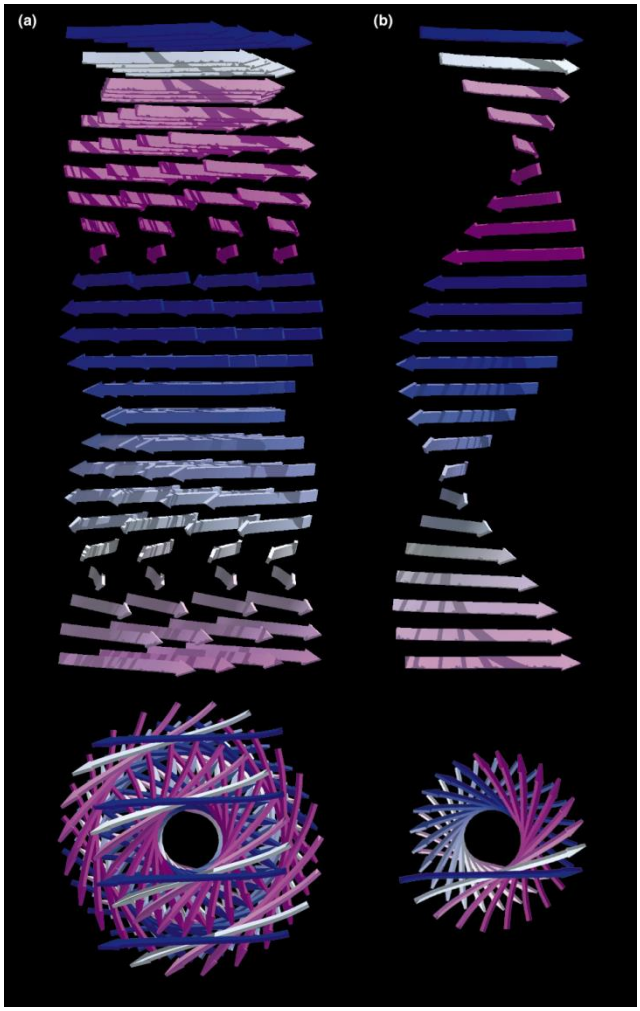


Figure 1.5.8 Transthyretin Protofilament X-Ray Diffraction Structural Model

(a) Model of the transthyretin protofilaments core structure viewed (top) normal to the filament axis and (below) along the filament axis. The arrows represent the paths, but not necessarily the directions, of the β-strands. The β-sheets are arranged in parallel to the filament axis. **(b)** An isolated β-sheet shown for clarity. Figure reprinted from (Blake and Serpell, 1996) with permission from Elsevier.

1.6 High Molecular Weight Oligomeric Species

Soluble oligomers can form on or off pathway to amyloid formation and range in size from dimers to large spherical or fibre-like aggregates (protofibrils) and large annular aggregates (figure 1.6.1) (Lashuel *et al.*, 2002). Aβ peptide oligomers for example, range in size from dimer (4.5 kDa) to dodecamer (56 kDa), and to species measured in nanometre diameters (Haass and Selkoe, 2007).

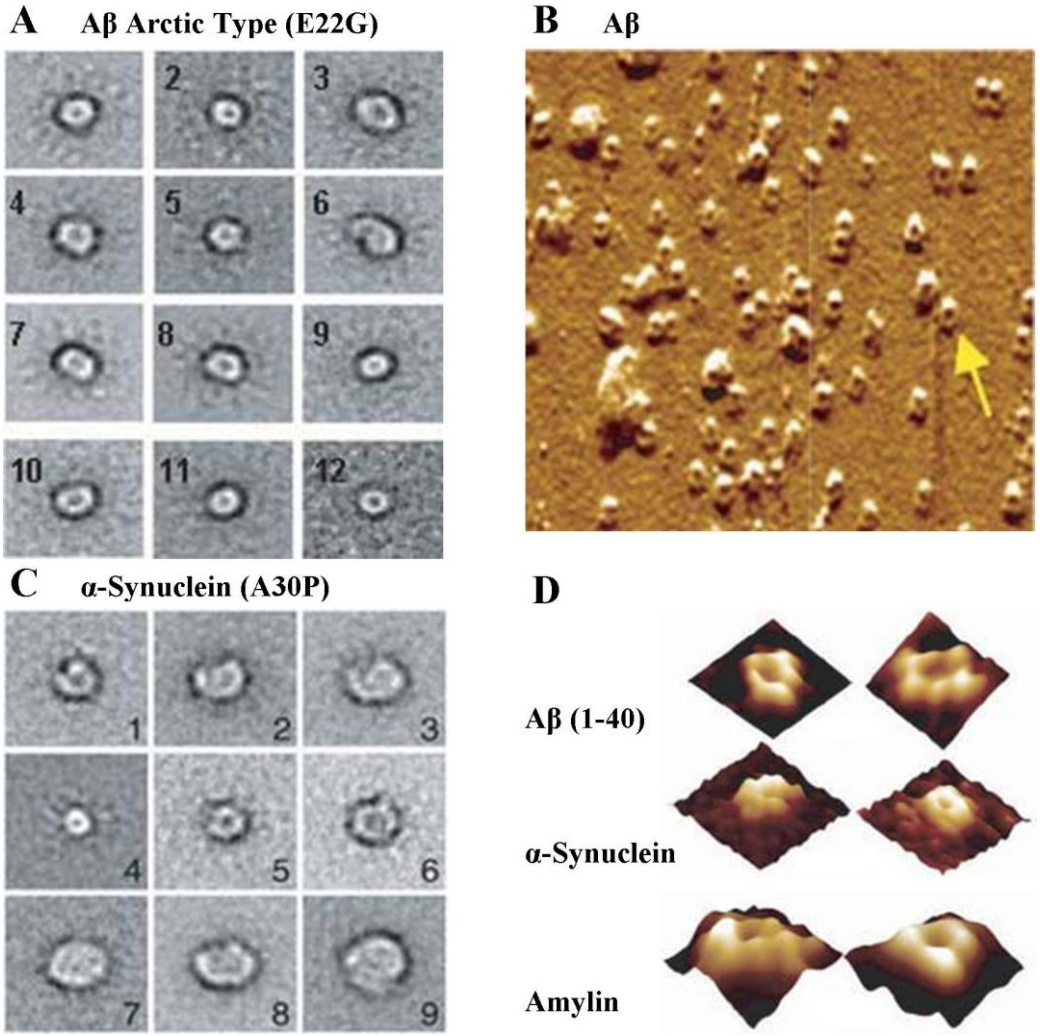


Figure 1.6.1 Annular Oligomer Images

Representative images of the many amyloid forming proteins which produce these variable diameter, channel- or pore-like properties *in vitro*. Images (A and C) are from EM, whilst the AFM images in B and D are from pore structures reconstituted in lipid bilayers. Figure adapted from (Lashuel and Lansbury, 2006) and reprinted with permission.

Oligomers generated from several proteins such as Aβ and lysozyme interacts and alters the fluorescence spectra of the dye 1-anilino-8-naphalene sulfonate (ANS). This dye has been used in protein folding studies as a marker for surface-exposed hydrophobic patches and molten-globule-like characteristics. Thus strong ANS interactions indicate incomplete burial of hydrophobic groups within these oligomeric assemblies (Fandrich, 2012). Oligomers bind specific antibodies such as A11, and some bind the OC antibody which was raised against protofibrils and also recognises amyloid fibre structures, thus suggesting a similar structure (Glabe, 2008).

Oligomer-specific antibodies typically display a common epitope distinct from those present on mature fibres. This epitope is therefore likely to be a common peptide backbone motif, such as an array of hydrogen bond donors and acceptors at the edge of a β-sheet or a turn motif. Due to antibody specificity for oligomers, the epitope must be either conformationally distinct or be greatly reduced in the mature fibre structure. The latter may explain binding affinity of oligomer-specific antibodies; during fibre elongation the length increases but the amount of β-sheet backbone exposed at the terminal end is reduced (Glabe, 2004).

1.7 Amyloid and Oligomer Toxicity

Early prefibrillar protein aggregates or intermediate oligomers appear highly damaging, whereas the mature fibres appear relatively inert and correlate poorly with disease progression (Dobson, 2003). Short amyloid fibres may in fact be responsible for toxicity (see figure 1.7.1) (Xue *et al.*, 2009), or these fibres could be less stable and allow rapid exchange between toxic oligomer species. Annular oligomers, have little known about the detailed molecular structure although bare coarse similarity to pore-forming toxins (see figure 1.6.1), thus suggesting an ability to perturb the integrity of cellular membranes (Fandrich, 2012).

A large body of evidence suggests that certain oligomeric structures are the principal pathogenic species and form unregulated membrane pores or disrupt homeostasis (Glabe, 2004). For instance, nerve biopsies of individuals exhibiting early signs of peripheral neuropathy in familial amyloid polyneuropathy have shown the presence of a non-fibrous transthyretin aggregate (Huff *et al.*, 2003). Oligomer specific antibody binding has been found to block toxicity of a wide range of oligomeric amyloid precursors. This 'universal' blocking of toxicity suggests soluble amyloid oligomers share a primary mechanism of pathogenesis, likely based on a common structure (Glabe, 2004). A further example of this is observed with soluble oligomers of amyloid forming non-pathogenic proteins such as HypF and the SH3 domain inducing significant toxicity in cell culture. A disulphide bonded Aβ β-sheet arc determined to spontaneously form only soluble oligomeric species and protofibrils has been shown to yield a higher degree of toxicity than WT $A\beta_{1-42}$ preparations (Sandberg *et al.*, 2010). A number of studies have also indicated that the presence of amyloid oligomers increases the permeability of membranes, possibly by forming pores or channels (Lashuel *et al.*, 2002, Lashuel and Lansbury, 2006, Milanesi *et al.*, 2012).

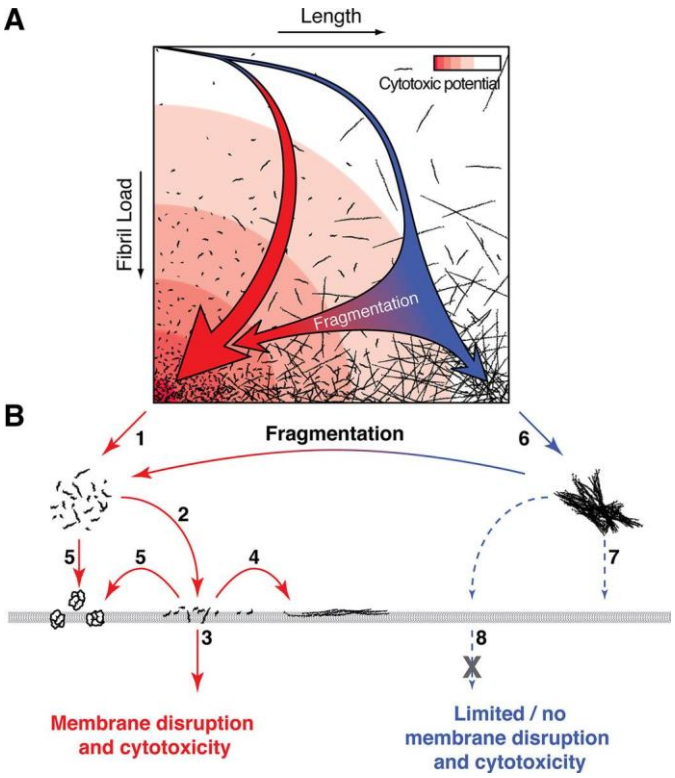


Figure 1.7.1 Amyloid Fibre Assembly and Fragmentation Landscape and Toxicity Mechanism

Amyloid fibre assembly and fragmentation landscape with fibre load plotted against fibre length. The intensity of the red background represents the toxicity potential. **The thick red arrow in (A)** illustrates a representative fibre assembly pathway that would occur in the presence of repeated fibre fragmentation or where nucleation is fast, relative to elongation, and results in rapid amyloidogenesis with short fibre length distributions. **(B1 and B2)** The short fibre may lead to enhanced toxicity through decreased fibre-fibre interaction or increased fibre-membrane interaction. **(B2-5)** Increased interaction between short fibres and membranes could result in membrane damage and/or a cytotoxic response by fibres puncturing through the membrane, extending on the membrane surface or producing cytotoxic species such as a pore. **The thin blue arrow in (A)** illustrates a fibre assembly pathway whereby little fibre fragmentation occurs, nucleation is slow relative to fibre elongation and results in a slow increase of fibre load and the formation of long fibres. **(B6-8)** These longer fibres are less likely to be biologically available due to increases in fibre-fibre interaction, a suggested decrease in membrane interaction and reduced frequency of fibre passing through the membrane, when compared to the short fibres. **Fragmentation** after assembly (either mechanical or chaperone mediated) shortens the average fibre length and thereby enhances the cytotoxic potential without apparent morphology changes. Figure from (Xue *et al.*, 2009).

1.8 Amyloid Disease Therapeutics

1.8.1 Introduction

The mechanism by which amyloid forming proteins and peptides cause disease is currently controversial, leading to substantial difficulties in drug development. However, significant efforts have been made in the search for novel therapeutics, with many in late stage clinical trials such as those for Alzheimer's disease described in table 1.8.1 (Palmer, 2011). Current treatment methods only alleviate symptoms and do not slow disease progression, therefore it is vital to develop effective treatment (Hard and Lendel, 2012).

COMPOUND	MOLECULE TYPE	MECHANISM OF ACTION	STATUS
Ginko biloba	Herbal Extract	Unknown	No efficacy
Bapineuzumab	Humanised Monoclonal Antibody against Aβ	Removal of Aβ peptide	No efficacy, dangerous to human health
Solanezumab	Humanised Monoclonal Antibody against Aβ	Removal of Aβ peptide	Questionable efficacy
Docosahexanoic acid	Long chain fatty acid	Membrane precursor	No efficacy
Alpha-tocopherol	Small molecule	Antioxidant	No efficacy
Arovastin	Small molecule	HMG CoA-reductase inhibitor (cholesterol production pathway)	No efficacy
Dimebon	Small molecule	H1-histamine receptor with neuroprotective activity	No efficacy
(AZD-103) (scyllo-inositol)	Small molecule	Prevents and reverses Aβ amyloid fibre formation	No efficacy
Non-steroidal anti-inflammatory drugs (NSAIDS)	Small molecule	Anti-inflammatory	No efficacy
Rosiglitazone	Small molecule	Peroxisome proliferator-activated receptor γ agonist	No efficacy
Semagacestat	Small molecule	γ-secretase inhibitor	No efficacy
Simvastatin	Small molecule	HMG CoA-reductase inhibitor (cholesterol production pathway)	No efficacy
Tarenflurbil	Small molecule	γ-secretase modulator (reduces Aβ concentration)	No efficacy
Tramiprosate	Small molecule	Reduces Aβ concentration	No efficacy
Valproate	Small molecule	$GABA_A$ receptor agonist and Na channel blocker	No efficacy
Vyndaqel	Small molecule	Stabilises transthyretin tetramer	Approved in Europe
Xaliproden	Small molecule	$5HT_{1A}$ serotonin receptor agonist	No efficacy

Table 1.8.1 Completed Phase III Clinical Trials of Alzheimer's Disease Neuroprotective Therapeutics

Table based on one in (Palmer, 2011) and recent reports on Solanezumab and Bapineuzumab (Panza *et al.*, 2012).

1.8.2 Therapeutic Modes of Activity

Therapeutic interventions for amyloid related disease must do one of three things: (i) manipulate the intrinsic stability of and/or the height of kinetic barriers to the disease causing protein state (see figure 1.8.1) (Cohen and Kelly, 2003), (ii) diminish the amyloid disease protein concentration, e.g. β-secretase inhibition to reduce Aβ concentration (iii) block the downstream pathological mechanism which currently, is poorly understood and controversial.

Direct Interventions could be considered to target five protein states: native, amorphous aggregate, oligomer and protofibril/short amyloid fibre and mature amyloid fibre (see figure 1.8.2). Determination of the toxic species is however, unclear and complicates therapeutic discovery with oligomer and protofibril/short amyloid fibre being primarily instigated (Hard and Lendel, 2012). With this in mind, an ideal therapeutic method would be to stabilise the native state, amorphous aggregate or amyloid fibre, in this order.

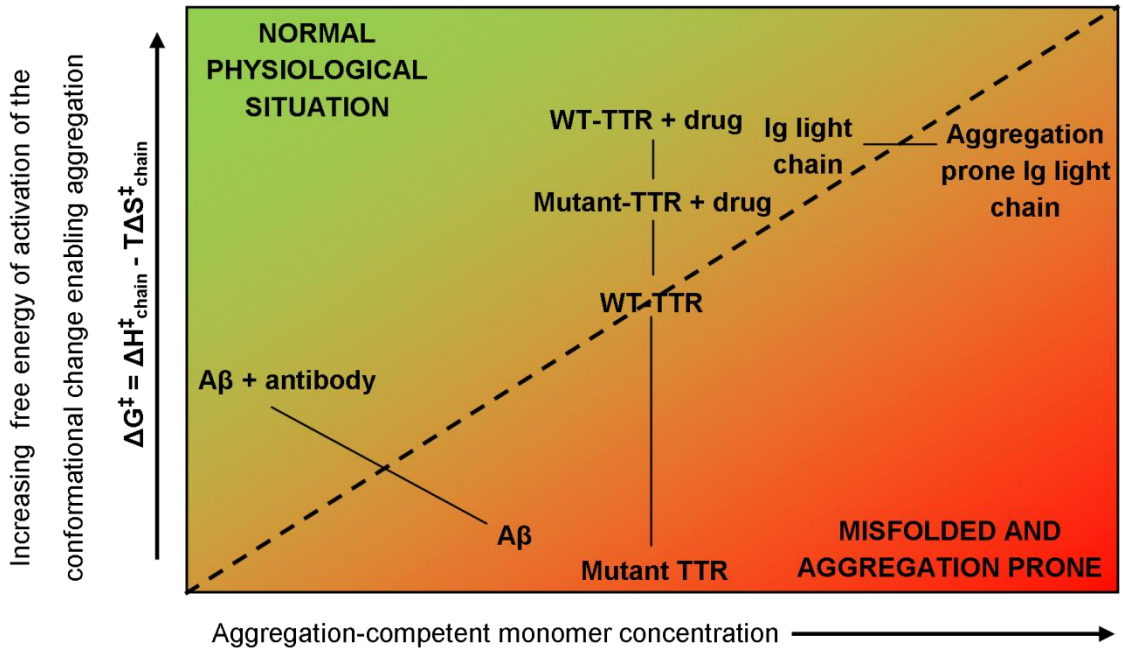


Figure 1.8.1 Misfolding and Aggregation Factors in Amyloid Disease

Analysis of the interplay between protein/peptide sequence, monomer concentration, and the presence of therapeutic in altering the free energy of activation ($\Delta G^{\ddagger} = \Delta H^{\ddagger} - T\Delta S^{\ddagger}$), for amyloidogenesis. Situations above the diagonal resist protein misassembly. Figure based on one present in (Cohen and Kelly, 2003).

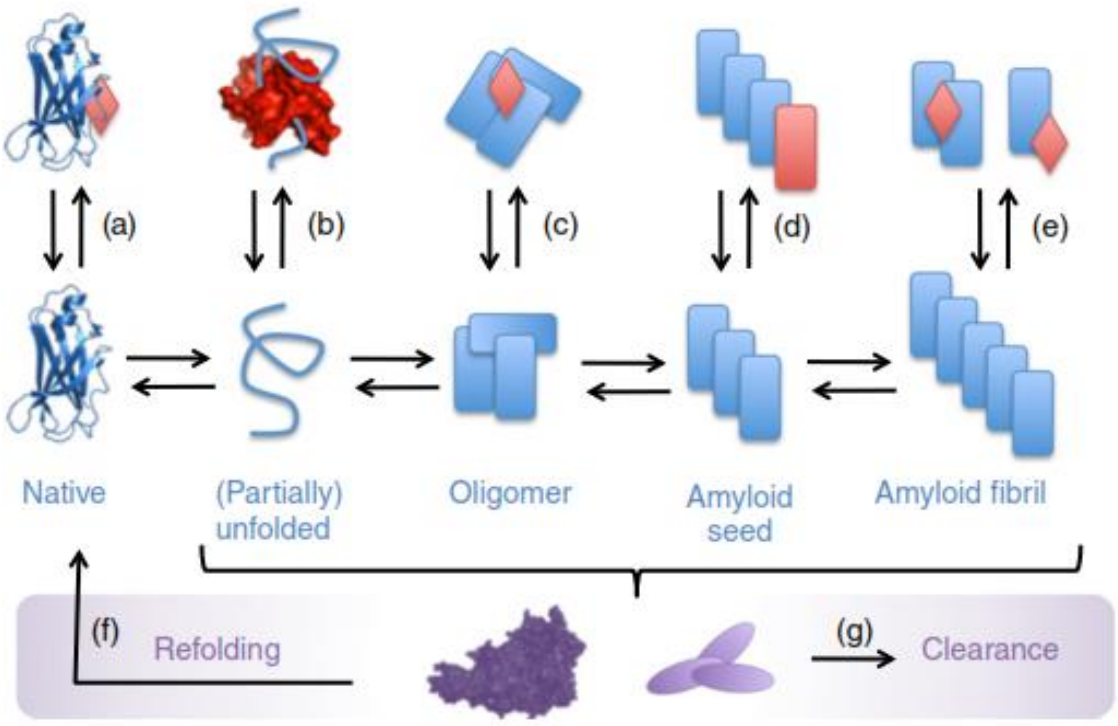


Fig. 1.8.2 Strategies to Inhibit the Formation of Amyloid Fibres

(a)–(e) depict biochemical mechanisms of action that could be employed by small or large ligands (red) binding to different states of amyloid-forming peptides (blue). **(a)** Stabilisation of the native state. **(b)** Sequestering of monomeric (partially) unfolded peptide. **(c)** Stabilisation or promotion of off-pathway oligomers. **(d)** ß-Sheet breakers that terminate fibre elongation. **(e)** Disassembly of amyloid fibres. In addition, refolding **(f)** and clearance **(g)** mechanisms (purple) mediated by chaperones or antibodies. The mechanism of aggregate/amyloid fibre clearance if intracellular, is mediated by lysosomes or the proteasome, or alternatively extracellular secretion and subsequent uptake by phagocytes. Figure reproduced from (Hard and Lendel, 2012) with permission from Elsevier.

Stabilising the native state is the simplest in terms of reducing protein toxicity, however, obtaining a sufficient degree of specificity in vivo can be difficult and may introduce side-effects or suffer delivery issues. An alternative strategy would be to stabilise and enhance non-fibrillar amorphous aggregation, thus allowing intra/extracellular clearance mechanisms to remove the misfolded protein. In a similar manner, stabilising non-toxic oligomer species may be a suitable approach, although ensuring safety by avoiding conversion to toxic forms and accurate *in vivo* toxicity determination may prove more challenging and risky.

Finally stabilising amyloid fibre could also be a suitable treatment for non-systemic amyloid diseases, with the primary aim of reducing the time toxic intermediates such as oligomer and protofibril/short amyloid fibre exist, thus reducing damage (Hard and Lendel, 2012). The limitations to these latter approaches include the increase of intermediate populations, which depending on the mode of toxicity may increase cytotoxicity. For example, once formed evidence suggests that amyloid fibres are dynamic and may act as a reservoir for toxic misfolded protein species, due to the dynamic nature of amyloid fibre (Carulla *et al.*, 2005, Sanchez *et al.*, 2011). In conclusion, no methodology is ideal with many novel therapeutics targeting different states, with varying success *in vitro* and *in vivo*. However, there is hope with several novel therapeutics in late stage clinical trials and the recent licensing in Europe of Vyndaqel, the first anti-aggregate/amyloid pharmaceutical.

1.8.3 Therapeutic Discovery Strategies

Screening for reduced amyloid fibre formation is typically performed by monitoring amyloid binding compound fluorescence, e.g. thioflavin T and S. Relying on external fluorescent probes is potentially troublesome due to potential binding competition and interference between the reporter and novel therapeutic molecules (Lendel et al., 2009, Buell et al., 2010, Hudson et al., 2009). However, the technique remains a starting point for compound screening, and the best molecules can then be followed up by an array of biophysical, biochemical and *in vivo* assays. Large screening projects such as the one performed by the Mandelkow group have identified some general structural features such as hydrophobic patches and hydrogen bonding groups (Bulic *et al.*, 2009). Other screening methods include surface plasmon resonance binding assays, cellular toxicity assays, filter retardation assays of insoluble amyloid fibre (Wanker *et al.*, 1999) as well as traditional affibody and antibody binding screens.

Developing atomic level structural models has also been a major research goal to aid therapeutic development and is now coming to fruition with the development of SSNMR and complementary biophysics/biochemistry methods (Petkova *et al.*, 2002, Luhrs *et al.*, 2005). These methods include multi-dimensional solution-state nuclear magnetic resonance spectroscopy (NMR) to map intermolecular interactions, electron and atomic-force microscopy to visualise changes in oligomeric and fibrous morphology. Furthermore, fluorescence spectroscopy, circular dichroism spectroscopy

and size-exclusion chromatography quantify changes in protein/peptide structure and the underlying kinetics and thermodynamics, while mass spectrometry identifies effector molecules and the binding stoichiometries. The use of these techniques to determine the underlying modal activity of novel therapeutics will be essential to persuade reluctant pharmaceutical companies to perform clinical trials and develop pharmaceuticals for the spectrum of amyloid-related diseases.

Therapeutic molecule binding must be strong (μM, or ideally sub-μM) to allow for low physiological concentrations and have favourable pharmacokinetics. Good pharmacokinetics is essential and remains the largest stumbling block for pharmaceutical development. The acronym ADMET, describes the five essential pharmacokinetics criteria; absorption, distribution, metabolism, excretion and toxicity.

Amyloid-related disease research has been concentrated on widespread diseases such as Alzheimer's and Parkinson's disease, with the aim of treating as many people as possible with the least time and work required. For these diseases it is easier to acquire funding and a breakthrough treatment may lead to development of treatment for other related diseases. Corollary to this, therapeutic development of rare 'orphan' diseases e.g. familial amyloidotic polyneuropathy (FAP) is stringent but often requires a lower degree of clinical testing due to lower numbers of affected individuals, thus is favoured by some pharmaceutical companies, partly due to additional government support and extended exclusivity rights. This simpler mode of development could be an ideal route for treatments using unconventional mechanisms, with success leading to novel therapeutics for more prevalent diseases. The ideal strategy would be to stabilise the native state of a protein, thus enabling the proteins normal function and avoid any possible toxic state. This may not be possible in some cases, thus using a variety of approaches is essential.

1.8.4 Therapeutic Molecule Classes

Small Molecules

Small molecule inhibitors of protein aggregation are very different from traditional drug design, which has typically targeted enzyme function, cell signalling and membrane transport proteins. Many of the currently investigated small molecule inhibitors act as general anti-amyloid agents, with effects on a range of amyloidogenic proteins (Ehrnhoefer *et al.*, 2008, Bulic *et al.*, 2009, McKoy *et al.*, 2012). This non-specific behaviour is typically an undesirable property for drug candidates but does offer the potential for broad-spectrum therapeutics, so long as the side-effects are manageable and do not adversely affect functional human amyloid fibres such as those in the endocrine system (Maji *et al.*, 2009).

Natural products are an excellent source of novel small molecule therapeutics as they provide a greater diversity of chemical space than combinatorial chemistry. Natural products and existing drug treatments also contain a higher number of chiral centers, elevated structure rigidity and a smaller number of aromatic moieties than molecules in combinatorial chemistry libraries (Feher and Schmidt, 2003). The increased variety and structural specificity of natural products therefore leads to increased chance of discovering a pharmacophore, with combinatorial chemistry utilised in later development stages for optimisation.

Small molecule classes identified include N-phylamines, anthraquinones, thioxothiazolidinones, inositol derivatives, porphyrins and polyphenols have been studied. Removal or substitutions of hydroxyl groups has been shown to reduce efficacy, thus indicating hydrogen bonding is essential in the interaction (Bulic *et al.*, 2009).

The polyphenols are a broad class of small molecules which have received significant research interest and show inhibitory activity on a variety of amyloids such as α-synuclein, IAPP, Aβ, PrP^{sc} and tau (Porat *et al.*, 2006). Polyphenols are often observed in plants such as ginko biloba, tea bush, grape or turmeric and thus are regarded as natural products (Bulic *et al.*, 2009). The polyphenols include flavonoids such as baicalein (Zhu *et al.*, 2004) and epigallocatechin-gallate (EGCG) (Ehrnhoefer *et al.*,

2006), tannic acids (Bulic *et al.*, 2009) and lignins such as curcumin, resveratrol and rosmarinic acid (Ono *et al.*, 2004). Some of these small molecule therapeutics require a relatively low inhibitor concentration for a marked effect, which mirrors the polymer chemist's use of small-molecule nucleating agents in plastics polymerisation (Kodali and Wetzel, 2007). However, polyphenols commonly display very high metabolic turnover, which frequently manifests itself in reduced oral bioavailability and reduced blood-brain barrier penetration, thus minimising potential as lead therapeutic structure, without modification (Bulic *et al.*, 2009)

The exemplary plant derived polyphenol (-)-epigallocatechin gallate (EPCG) is found in large amounts in green tea, efficiently inhibits amyloid fibre formation of both Aβ, α-synuclein, huntingtin and produces oligomeric structures (Ehrnhoefer et al., 2008, Ehrnhoefer et al., 2006). EPCG also destabilises Aβ and α-synuclein amyloid fibre to produce protein aggregates that are non-toxic to mammalian cells (Bieschke *et al.*, 2010). The mechanism by which it performs this action is poorly understood and controversial.

Affibodies

Affibodies are small protein domains, capable of specific binding to target proteins in a similar manner to antibodies. Most affibodies are based on the α-helical bacterial receptor domain Z, derived from staphylococcal protein A and using combinatorial approaches to rationally design or randomise 13 solvent-accessible surfaces, specific interaction can be achieved and screened for with phage display (Nord *et al.*, 1997). Affibodies therefore can potentially be used for binding amyloid-forming native protein state, oligomer or fibre-like structures with lower immunogenicity, see section 1.8.5 and figure 1.8.4 for the ZAβ3 example.

Antibodies

Antibody mediated therapy has been cited as one of the most promising with two in late stage clinical trials (Palmer, 2011, Pul *et al.*, 2011). The mechanism of action of immunotherapy is controversial, with some antibodies showing reduced protein concentrations in neuronal cells, cerebral spinal fluid and blood plasma. Antibodies have been suggested to cross the blood-brain barrier and clear amyloid by Fc-mediated phagocytosis by microglia (Bard *et al.*, 2000) or act as a 'peripheral sink' in blood plasma and cerebral spinal fluid, thus shifting the equilibrium concentrations of amyloid in the central nervous system and thereby reducing concentrations of toxic amyloid species (DeMattos *et al.*, 2001).

Bapineuzumab, a lead monoclonal antibody which binds the Aβ peptide N-terminal, has in 2012 failed clinical testing with 6% of the mild-moderate Alzheimer's disease subjects developing aseptic meningoencephalitis and rare brain microhemorrhages, which indicates undesirable immunogenicity. Solanezumab, a second lead monoclonal antibody which binds the middle region of Aβ peptide has been proven safe but the efficacy is under question with only 30% of the phase III trial subjects showing improvements in symptoms, when compared with the control group (Panza *et al.*, 2012). Although the first trials with these antibodies are discouraging, the methodology remains feasible and a range of antibodies which target specific protein states (Kayed *et al.*, 2003) such as amyloid fibre and specific oligomers remains of interest for research and treatment purposes.

1.8.5 Therapeutic Discovery Examples

Native-State Stabilisation

Therapeutic molecule binding to the native state thermodynamically increases the population relative to non-native state (Miroy *et al.*, 1996). An ideal example of this is the transthyretin (TTR) stabilising therapeutic tafamidis meglumine (Vyndaqel, Pfizer), which has been approved for use by the European Medicine Agency in November 2011 to treat familial amyloidotic polyneuropathy (FAP). This small molecule works by binding the native state thyroxine binding sites (see figure 1.8.3), thus bridging the dimeric interface via hydrophobic and ionic interactions which stabilise the tetrameric native-state thus avoiding the aggregation-prone monomeric state (Connelly *et al.*, 2010).

In a similar manner to FAP treatment, novel therapeutics for amyloid lateral sclerosis (ALS) caused by superoxide dismutase-1 (SOD-1) amyloidosis, have been designed. *In silico* screening has been performed (Ray *et al.*, 2005) followed by experimentally testing small molecule activity to bind and stabilise the dimer, thus avoiding the amyloidogenic monomer state (Nowak *et al.*, 2010). In much the same way, a single-domain fragment of a camelid antibody raised against WT lysozyme stabilises the native-state D67H lysozyme mutant and inhibits amyloid formation, which otherwise causes hereditary systemic amyloidosis (Dumoulin *et al.*, 2003).

Native-state stabilisation of proteins and peptides which are thought to be natively unfolded is also possible, although less straightforward due to challenging specificity. The therapeutics described above use relatively well understood techniques to design and develop the pharmaceutical in a manner analogous to enzyme binding and inhibition, whereas natively unfolded protein binding is much less explored and understood by the pharmaceutical industry. Some solace may be had by the finding that α-synuclein, which was previously thought to be natively unfolded, forms a folded tetrameric α-helical structure (Bartels *et al.*, 2011) when expressed in mammalian cells. This discovery yields hope for a TTR-like treatment method for Parkinson's disease and points to possibly incorrect findings and assumptions of other proteins and peptides commonly described as natively unfolded, an example being the Aβ peptide.

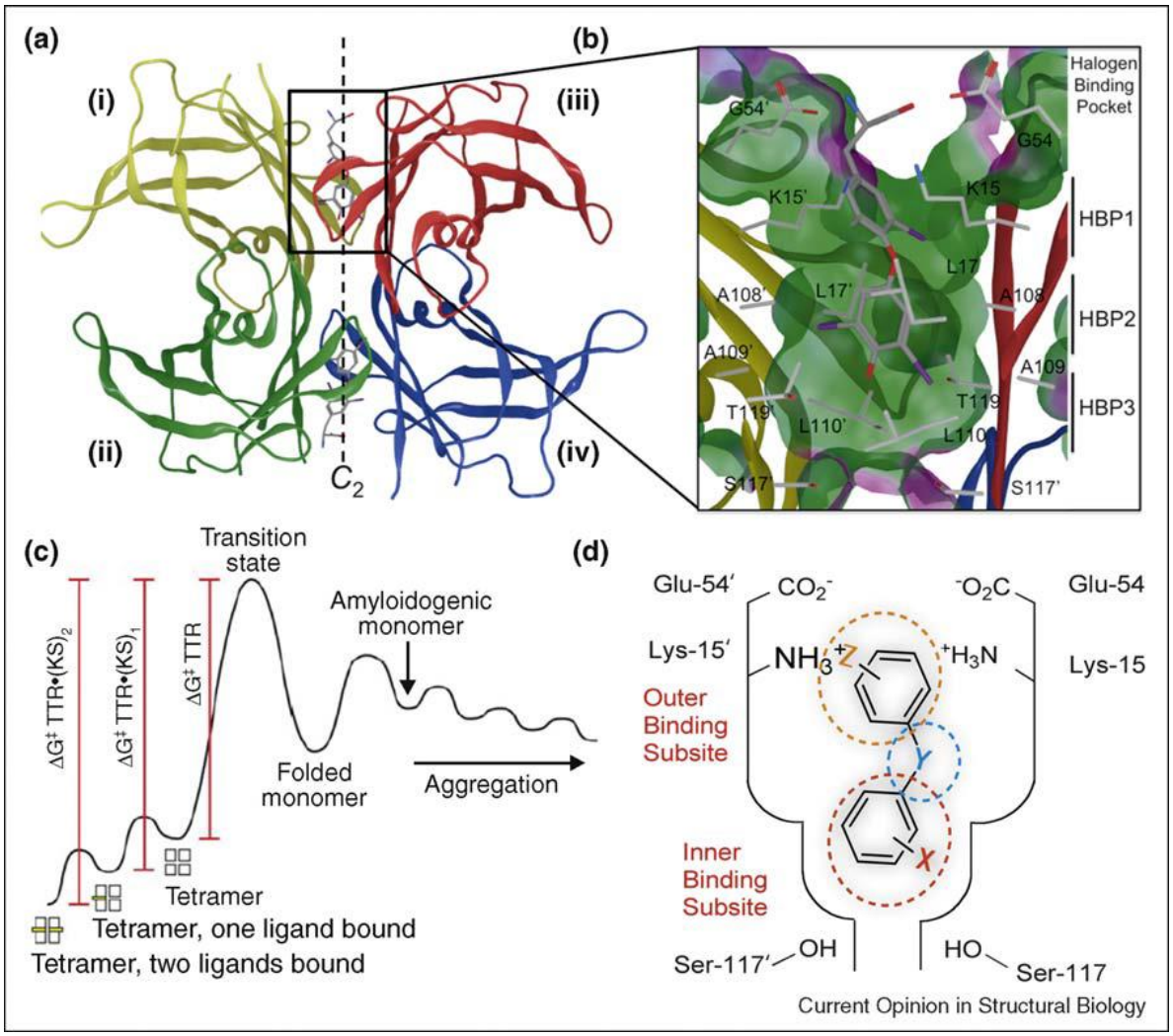


Figure 1.8.3 Structure of the Transthyretin: Thyroxine (T4) Complex and the Basis for Kinetic Stabilisation.

(a) Crystal structure of TTR with T4 bound to each T4 binding site (PDB code: 2ROX). **(b)** Close-up view of one T4 binding site with T4 bound. **(c)** Kinetic stabilization of TTR by small molecule binding to the natively folded tetramer, which stabilises the ground state and increases the tetramer dissociation free-energy barrier, thus preventing amyloidogenesis and pathology. **(d)** Schematic of one of the two T4 binding pockets within TTR occupied by a small molecule TTR kinetic stabilizer. Y represents a linker of variable chemical structure (e.g. NH, O, CH CH, C(O)NH, etc.) joining the two aryl rings, which typically bear a combination of alkyl, carboxyl, halide, trifluoromethyl, or hydroxyl substituents (X and Z). Figure reproduced from (Connelly *et al.*, 2010) with permission from Elsevier.

Aggregation Inhibition

Strong binding of therapeutic agents to aggregation prone regions of amyloidogenic peptides will prohibit self-assembly and avoid aggregation. An example of this is the affibody ZAβ3 which forms a non-polar, large tunnel-like binding site in which Aβ binds as a β-strand arc (β-hairpin) (see figure 1.8.4). The affibody was generated by biotinylation of $A\beta_{1-40}$ and immobilisation on streptavidin coated beads followed by selection of affibody binders from a phage display library. ZAβ3 monomer is comprised of a three helix bundle which forms a disulfide bonded dimer and generates the central binding site. Aβ residues 17-36 interact with two affibody β-strands to form a four stranded intermolecular β-sheet (Hoyer *et al.*, 2008). The equilibrium dissociation constant (K_d) of ZAβ3 for $A\beta_{1-40}$ and $A\beta_{1-42}$ is 17 nM indicating ideal tight binding, however the affibody:Aβ complex involves a high activation free-energy barrier, thus may require optimisation (Hoyer and Hard, 2008).

ZAβ3 inhibits amyloid fibre formation in a 1:1 stoichiometric ratio and acts to dissolve pre-formed aggregates over several hours. ZAβ3 also acts to dissolve mature amyloid fibre by sequestering monomeric Aβ, thus shifting the equilibrium, however, the fibre dissolution kinetics remained very slow in physiological buffers, thus indicating high kinetic barriers. In addition, the ZAβ3 affibody when expressed in WT or E22G $A\beta_{1-42}$ expressing transgenic *Drosophila melanogaster, restores the* life-span to that of WT flies with WT $A\beta_{1-42}$ and to ¾ of the lifespan in E22G flies. In these affibody flies, aggregation is inhibited and Aβ is cleared from the fly by an unknown mechanism (Luheshi *et al.*, 2010).

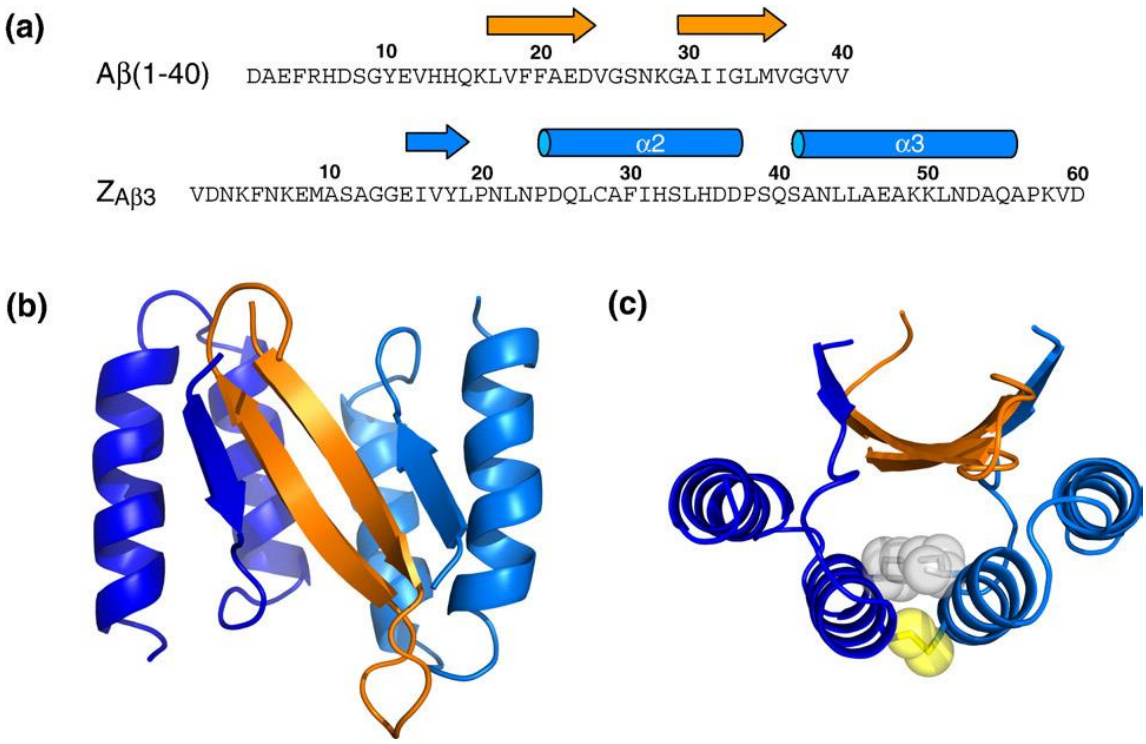


Figure 1.8.4 Structure and Amino Acid Sequences of the ZAβ3:$A\beta_{1-40}$ Complex

(a) Amino acid sequences of $A\beta_{1–40}$ and ZAβ3. Helical and β-sheet secondary structure in the ZAβ3: $A\beta_{1–40}$ complex is indicated by cylinders or arrows respectively. ZAβ3 helices are labelled according to their position in the Z domain affibody scaffold. **(b and c)** Cartoon topology of the ZAβ3: $A\beta_{1–40}$ complex from equatorial (b) and meridian (c) perspectives. $A\beta_{1–40}$ is depicted in orange, ZAβ3 subunits in blue. The subunit-linking disulfide bond is shown in yellow and the essential I31 side chains of both ZAβ3 subunits are shown in grey (c). Disordered N-termini are not displayed in (b) and (c). NMR structure reproduced from (Hoyer and Hard, 2008) with permission from Elsevier.

Inhibition of Amyloid Elongation

β-sheet breakers work by binding to the terminal exposed central hydrophobic regions of amyloid fibres and inhibit elongation (Tjernberg *et al.*, 1996). Protease-resistant D-amino acid peptides, proline residues and charged terminal amino acids have been inserted into these therapeutic peptides to avoid rapid proteolysis and disrupt intrinsic peptide β-sheet propensity. N-methylation is used in many β-sheet breaker peptides to inhibit formation of amyloid fibres from the blocking peptide themselves, improving ADMET properties and is also reported to improve membrane penetration though rapid phospholipid bilayer diffusion (Gordon *et al.*, 2002). Alternating N-methylated peptide presents a hydrogen-bonding 'complementary' face to the amyloid target, whilst the other 'blocking' face has N-methyl groups in place of backbone NH groups, thus preventing further aggregation (Hughes *et al.*, 2000). Much of this work has been

directed at binding the Aβ peptide, often targeting residues 16-20 (KLVFF) peptide region (Tjernberg *et al.*, 1996, Soto *et al.*, 1998, Kokkoni *et al.*, 2006) and other β-sheet forming regions such as residues 25-35 (Hughes *et al.*, 2000). These studies yielded positive results in biochemical/biophysics/*in vitro* assays often showing inhibition of amyloid formation, disaggregation of mature fibres and reduced toxicity.

A Novel Therapeutic with Unknown Activity:

The phage particle, NPT001, developed by NeuroPhage Pharmaceuticals has been shown to potently clear pathologic protein plaques. This M13 based filamentous bacteriophage was serendipitously discovered to mediate disruption of amyloid assemblies (Dimant *et al.*, 2009, Dimant and Solomon, 2010). NPT001 has been shown to reduce Aβ and tau protein plaques in Alzheimer's disease (AD) preclinical studies, whilst also disrupting alpha-synuclein amyloid fibres, which play a critical role in Parkinson's disease (PD). A study by Neurophage Pharmaceuticals demonstrated that a single intracranial NPT001 treatment produced significant reductions in neuropathology along with improved motor performance in a PD mouse model. Specifically, NPT001 significantly reduced alpha-synuclein protein plaques in studied mouse brains. In addition, NPT001 was well-tolerated and produced no observable adverse effects. The effect of this therapeutic agent is non-specific and therefore provides additional scope for treatment of rare medical conditions such as frontal temporal dementia, progressive supranuclear palsy, and prion diseases. Further experimentation is being performed on this therapeutic including NMR binding studies and H/D exchange NMR (Peter Davis, Andrea Hounslow and Jon Waltho, unpublished).

1.8.5 Conclusions

A range of potential therapeutics are being developed using a wide range of approaches including small molecules based on natural products and larger, more sophisticated biomolecules. Small molecules have potential advantages in increasing the delivery through the blood-brain barrier and longer in vivo lifetimes while careful development of larger biomolecules are yielding increased specificity on novel targets.

1.9 Cystatins as a Model Amyloid

1.9.1 Cystatin Superfamily

The cystatin superfamily are reversible, high affinity, broad specificity cysteine protease inhibitors (Hou et al., 2004) and have roles implicated in protein turnover (Turk and Bode, 1991, Turk et al., 2001), neurogenesis (Taupin et al., 2000), the innate immune response (Kopitar-Jerala, 2006) and sperm maturation and fertilisation (Cornwall et al., 2011). Cystatins bind protease active sites through hydrophobic interactions of the N-terminal residues and the loops between β-strand 2-3, containing the highy conserved QVVAG motif and the loop between β-strands 3-4 (see figure 1.9.1) (Gottlieb et al., 1997). The highly constrained loop between β-strand 2-3 provides the required resistance to protease hydrolysis (Abrahamson, 1993). The structurally homologous cystatin superfamily is divided into three families; type I cystatins (also known as stefins) which are typically intracellular, type II cystatins which are distinguished from Type I by intramolecular disulphide bonding and are typically extracellular and kininogens (type III) which are extracellular and composed of three cystatin domains (Barrett *et al.,* 1986, Abrahamson, 1993).

1.9.2 Human Cystatin B

Human cystatin B (alternative nomenclature, stefin B) is a 98 amino acid, 11 124 Da, mixed α-helix/β-sheet globular cysteine protease inhibitor containing no disulphide bonds. The native topology of cystatin B is that of a five-stranded β-sheet surrounding a central perpendicular α-helix ($\beta_1\alpha_1\beta_2\beta_3\beta_4\beta_5$, "hot dog fold" [PDB:1STF]) as shown in figure 1.9.1 and 1.9.2 (Stubbs *et al.,* 1990). Cystatin B is ubiquitously expressed and located in the cytoplasm, nucleus and associates with a specific subset of lysosomes, possibly counteracting inappropriate proteolysis caused by cathepsins leaking from lysosomes (Alakurtti *et al.,* 2005). *In vitro,* Cystatin B binds tightly to cathepsins B, H, L and S. The main function of cathepsins is non-selective degradation of intracellular proteins and antigen processing in apoptosis. It has therefore been suggested that Cystatin B is part of a multi-protein complex involved in apoptosis (Joensuu *et al.,* 2008). Cystatin B also appears to have an endogenous neuroprotective role as deficient mice exhibit cerebral granular cell apoptosis which results in loss of motor control, ataxia (loss of voluntary coordination of muscle movements) and myoclonus (twitching

of muscle) (Pennacchio *et al.*, 1998). Additional functions are suggested by double knockout mice deficient in cystatin B and cathepsin B which display ataxia and monoclonus (uncontrolled twitching) but show markedly reduced apoptosis (Houseweart *et al.*, 2003).

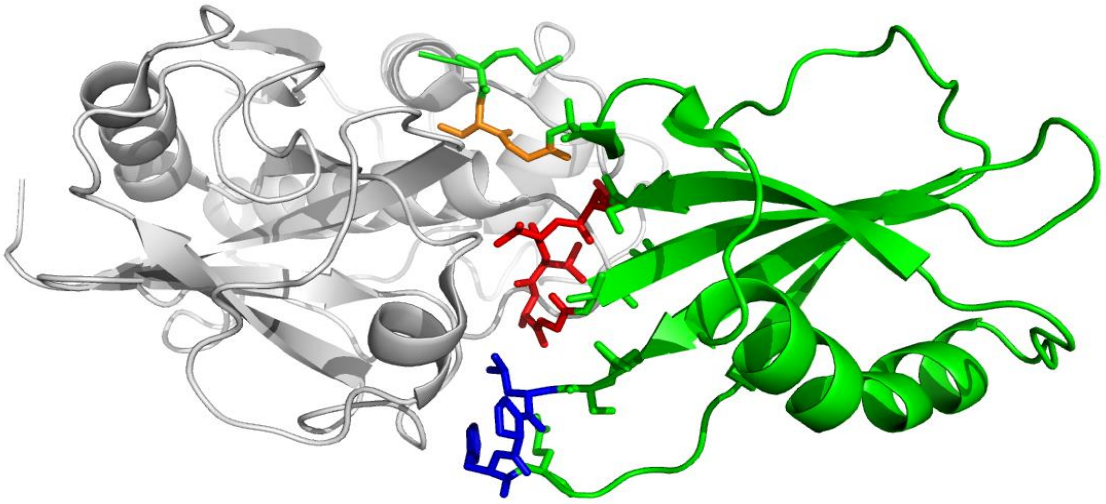


Figure 1.9.1 Cystatin B Inhibition Binding Site

(Green) Cystatin B, **(gray)** papain, **(red)** primary binding site loop QVVAG which is highly conserved in the cystatin family **(blue)** secondary binding loop residues LPH and **(orange)** N-terminal glycine and serine which are weakly involved in binding. X-ray crystal structure of cystatin B complexed with papain [PDB:1STF] graphically rendered in Pymol (DeLano Scientific, CA.

The neuroprotective role of cystatin B is further highlighted with mutations resulting in myoclonus epilepsy of type I (EPM1), a rare frequency (1 in 20 000, in Finland), monogenic autosomal recessive progressive and degenerative epilepsy also known as Unverricht-Lundborg disease. EPM1 has an onset between 6 and 13 years with variable progression to cerebellar ataxia and mental deterioration (Norio and Koskiniemi, 1979). Four protein coding region mutations in EPM1 have been reported and are proven to be responsible for disease onset; nonsense (g.2388C>T, p.Arg68X), frame-shift (g.2400_2402delTC, pLys73fsX2) and two missense (g.426G>C, p.Gly4Arg and g.2398A>C, p.Glu71Pro) (Rabzelj *et al.*, 2005). The G4R mutant cystatin B displays similar folding stability and structure to the WT, however in fibrillisation conditions remains in the prefibrillar aggregate state for four times longer than the WT. The R68X mutant on the other hand is reported to be unstructured in standard conditions (pH 7.0, 25°C) and very rapidly forms amyloid fibres (Rabzelj *et al.*, 2005).

However, protein coding region mutations only account for 14% of 58 EPM1 alleles studied. Instead, the main incidence of EPM1 is caused by >3 dodecamer (12-mer, CCCCGCCCCGCG) nucleotide repeats located ~70 nucleotides upstream of the transcription start site nearest the 5' end of the cystatin B gene. The pathophysiology of these gene instabilities is unclear, although is most likely responsible for absence or reduction of cystatin B in specific tissues of affected individuals thereby reducing the inhibition of proteases involved in apoptosis and leading to uncontrolled cell death (Lalioti et al., 1997).

1.9.3 Human Cystatin B Amyloid Fibre

Cystatin B fibre preparations with 30 μM protein produced at 37 °C, pH 4.7 and 10 % 2,2,2, trifluoroethanol (TFE) consistently produce long unbranched smooth fibres, over a period >5 days. The addition of TFE is essential for fibre formation *in vitro* for experimentally useful timescales and has been shown to induce NMR chemical shift changes on the dimeric domain swapped β-strands 2 and 3 (Paramore et al., 2012). These fibres extensively laterally associate and share identical morphology in EM micrographs, see figure 1.9.2. The fibres produced are 3.4 (±0.3) nm in height exhibiting a twist with a periodicity of 27 nm as determined by AFM, (Zerovnik *et al.*, 2002a) with most mature fibres tens of micrometers long. These fibres are 1% sodium dodecyl sulphate (SDS) and 6 M guanidine hydrochloride (GuHCl) soluble.

Cystatin B Amyloid Formation Pathway

The cystatin B amyloidogenesis pathway described in figure 1.9.2 is complicated by the *in vitro* dynamic equilibrium of the dominating dimer species (~90%), followed respectively by monomer (~10%) and tetramer (~1%). However, it is suggested that the human cystatins A, B, C and chicken cystatin simply form amyloid fibres via a 3D domain swapped dimer intermediate (Staniforth et al., 2001, Janowski et al., 2001). A range of oligomeric states is not uncommon, with several low molecular weight oligomers (dimer-dodecamer) observed in soluble Aβ preparations (Stine *et al.*, 2011).

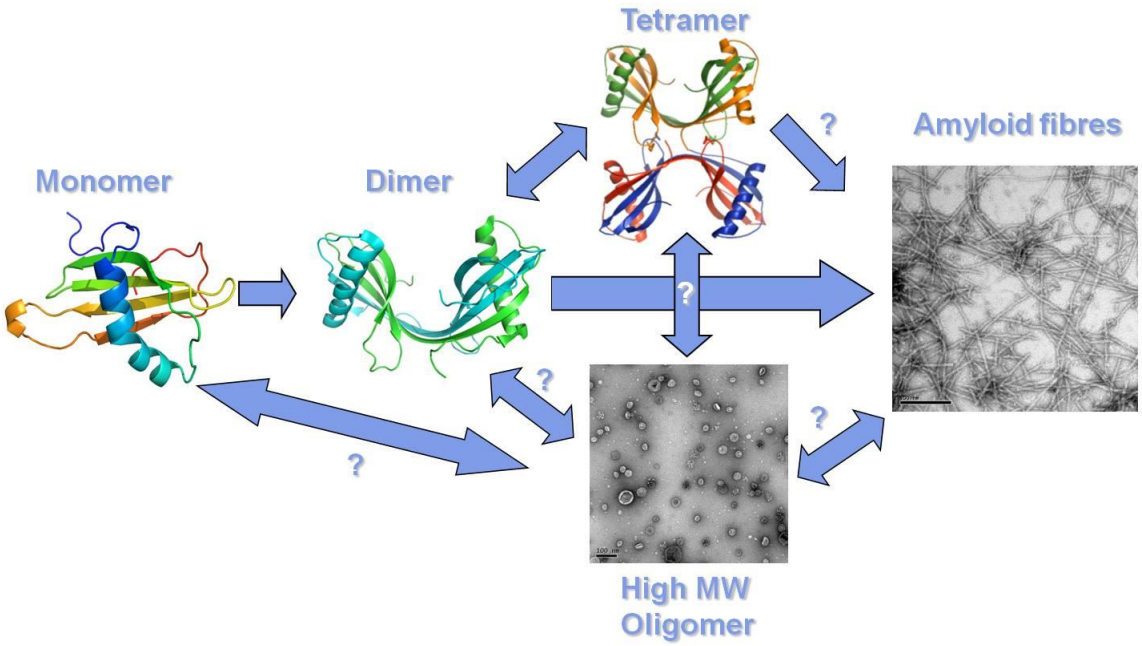


Figure 1.9.2 Cystatin B Amyloid and Oligomer Formation Pathway

Cystatin B *in vitro* equilibrates to form predominantly domain swapped dimer along with some monomer and tetramer. In specific destabilising conditions WT cystatin B forms amyloid fibre and rare instances of high molecular weight oligomer. PDB: monomer 1STF, tetramer/dimer 2OCT, graphically rendered in Pymol (DeLano Scientific, CA). The tetramer structure is from (Kokalj *et al.*, 2007) reprinted with permission from Elsevier.

1.9.4 Human Cystatin C

Human cystatin C (HCC) is structurally homologous to cystatin B and in 120 amino acids shares 26% sequence homology and 38% sequence similarity to the shorter human cystatin B (Y31 polymorph) protein (see figures 1.9.3 and 1.9.4). Therefore human cystatin B is an amenable and suitable homologue for HCC amyloid fibre formation. HCC is an extracellular type II cysteine protease inhibitor and is cellularly the most widely distributed human cystatin. HCC contains two disulphide bonds, one in the extended loop between β-strands 3 and 4, whilst the other stabilises the C-terminal by linking β-strands 4 and 5. Similar fold stability is achieved in the intracellular cystatin B by a reduced three residue loop linking β-strands 3 and 4, as well as an extended c-terminal which forms a hydrogen bond back to β-strand 4 (Japelj et al., 2004, Martin et al., 1995, Waltho et al., 1995).

HCC is directly associated with human cystatin C amyloid angiopathy (HCCAA), a systemic progressive amyloidosis, with symptoms including recurrent stroke, paralysis and dementia followed by pathological cerebral haemorrhage typically from the age of 17 to 40 years of age. The L68Q polymorphism is the only mutation responsible for HCCAA pathology (Jensson et al., 1987, Ghiso et al., 1986, Palsdottir et al., 1988). The replacement of a hydrophobic leucine for a hydrophilic glutamate in the largely hydrophobic interface between the β-sheet and α-helix, reduces the folding stability of HCC which when partially unfolded can form amyloid fibres.

HCCAA exhibits amyloid deposition in the tunica intima and/or outer wall of the small arteries and arterioles of the leptomeningea, cerebral cortex, basal ganglia, brainstem and cerebellum in all autopsied cases (Levy et al., 1989, Revesz et al., 2003, Gudmundsson et al., 1972). Further to this HCC has been found to co-localise with Aβ fibre deposits in the brains of patients with AD, hereditary cerebral haemorrhage with amyloidosis - Dutch type (HCHWA-D) and sporadic cerebral amyloid angiopathy (CAA) (Levy et al., 2001, Deng et al., 2001, Vinters et al., 1990, Haan and Roos, 1992).

```
Query:      Human cystatin B, 2OCT chain: A, Length: 98
Subject:    Human cystatin C, 1G96 chain: A, Length: 120

2OCT.A   1 MMSGAPS--ATQPATAETQHIADQVRSQLE---EKYNK------KFPVFKAVSFKSQVVA  49
             | |       |  |  .   . ||  |.    .|||           . |  . |.||
1G96.A   1 SSPGKPPRLVGGPMDASVEE--EGVRRALDFAVGEYNKASNDMYHSRALQVVRARKQIVA  58

2OCT.A  50 GTNYFIKVHVG-----------D----EDFVHLR-----VFQ--SLPHENKSLTLSNYQT  87
           | |||. | .|           |     |  ||.      ||  ..| .  ..|||
1G96.A  59 GVNYFLDVELGRTTCTKTQPNLDNCPFHDQPHLKRKAFCSFQIYAVPWQG-TMTLSKSTC 117

2OCT.A  88 NKAKHDELTYF  98
             |
1G96.A 118 QDA-------- 120
```

Figure 1.9.3 Human Cystatin B and C Global Sequence Alignment

(Green) identical residues, **(pink)** similar residues, **(blue)** sequence mismatch, **(brown)** insertion/deletion. Needleman-Wunsch global sequence alignment (Needleman and Wunsch, 1970) was performed on PDB sequences of human cystatin B polymorphism Y31 (2OCT) and human cystatin C (1G96). Alignment tool available at http://www.rcsb.org/pdb/ (Prlic et al., 2010).

```
Cystatin B:       ------MMSGAPSATQPATAE-TQHIADQVRSQLEEKYNK--
Cystatin C:       SSPGKPPRLVGGPMDASVEEEGVRRALDFAVGEYNKASNDMY

Cystatin B:       KFPVFKAVSFKSQVVAGTNYFIKVHVG----------------
Cystatin C:       HSRALQVVRARKQIVAGVNYFLDVELGRTTCTKTQPNLTTCTK

Cystatin B:       ------------------------------------DEDFV
Cystatin C:       TQPNLTTCTKTQPNLTTCTKTQPNLDNCPFHDQPHLKRKAFCS

Cystatin B:       HLRVFQSLPHENKSLTLSNYQTNKAKHDELTYF
Cystatin C:       FQIYAVPW-----QGTMTLSKSTCQDA
```

Figure 1.9.4 Human Cystatin B and C Structural Alignment

Structural alignment performed with with X-ray crystallography PDB coordinates of human cystatin B **(Blue)** and human cystatin C domain swapped dimer**(Green)**, (1OCT) and (1G96) respectively. The secondary structure elements of both proteins align very closely with only two cystatin B loop regions ($\beta_2\beta_3$ and C-terminal) being distinct and larger than those in cystatin C **(Dark grey)**, whilst one large loop in cystatin C ($\beta_3\beta_4$) is much larger than that of cystatin B **(Light grey)**. Dimers were compared due to the availability of comparable coordinates and clear indication of shared secondary structure elements. The sequence segments determined to reliably align are labelled in yellow. The comparison was performed with the jFATCAT-flexible algorithm (Ye and Godzik, 2003) available at http://www.rcsb.org/pdb/ (Prlic et al., 2010) and rendered in Pymol (DeLano Scientific, CA).

1.10 Introduction to Thesis

The initial aim of this project was to develop faster, more generally applicable methods for defining the structure of cystatin amyloids. H/D exchange methods and lengthy site-directed disulphide bridge mapping of the structure had yielded a model where fibres were composed of stacked dimer-like structures where the topology of the native β-sheet remained relatively intact despite the exclusion of a large part of the N-terminal region of the molecule. A main task was to establish whether this architecture had implications for cystatin B amyloids formed in different solution conditions and whether oligomeric non-fibrillar forms contained structure reminiscent of the fibre. The amyloid fibre structural work presented in chapter 3 of this thesis uses four strategies to investigate the structure of cystatin B amyloid fibre. The first method is limited proteolysis of mature cystatin B amyloid fibre to probe for the fibre core amino acid residues and indicate structural features. Scanning transmission electron microscopy (STEM) mass-per-unit-length (MPL) and transmission electron microscope (TEM) width measurements were performed on cystatin B fibres to determine the protein molecule frequency and maximum spatial dimensions, which restrains the possible structures.

In order to apply this methodology to other forms of cystatin B, the formation of oligomeric structures, described in chapter 4, was performed with preliminary EM and biochemical characterisation with the intention of future structural study. Compared with its extracellular analogue, cystatin C, these cystatin B oligomeric structures are poorly populated and a challenge to isolate (Paramore, 2011), however this was partially overcome. Further work is also described in the appendices.

Chapter 2: Materials and Methods

2.1 Introduction

This chapter details common experimental procedures used throughout the work presented. Further details of material and methods particular to specific experiments are presented in the results chapters.

2.2 Buffers and Reagents

Unless stated otherwise, all reagents were purchased from, BDH Laboratory Supplies, Fisher, Melford or Sigma-Aldrich. Deionised water (18.2 Ω) from an Elga Purelab 611 Classic UVF was used throughout all experiments. Buffers were prepared according to the protocol described in (Sambrook *et al.,* 1989) and filtered through a 0.2μm filter.
1 mM sodium azide (NaN_3) was added as standard to all buffers, except those used for bacterial growth and unless otherwise stated.

2.3 Growth Media and Solutions

2.3.1 Luria-Bertani Media

Taken from (Sambrook *et al.,* 1989). For each litre, the following was added to deionised water:

10 g bacto-tryptone (Oxoid ltd, UK)
5 g yeast extract (Oxoid)
10 g NaCl (Melford)

Media was sterilised by autoclaving, and antibiotics added after the media had cooled. LB-agar was produced by the addition of 15 g/l bacto-agar (Oxoid) to the media prior to autoclaving.

2.3.2 Antibiotic Stock Solutions

Ampicillin (Melford): 1000x stock solution was prepared by dissolving ampicillin sodium salt in water at a concentration of 100 mg/ml and subsequently 0.2 μm filter sterilised. 1 ml and 5 ml aliquots were stored at -20°C until needed, when they were gently thawed and added to growth media to a final concentration of 100 μg/ml.

Chloramphenicol (Melford): Stock solution was prepared by dissolving chloramphenicol in 20% ethanol/water at a concentration of 100 mg/ml, and 0.2 μm filter-sterilised (Sambrook et al., 1989). A 10 ml stock was stored at 20°C until needed and added to growth media to a final concentration of 50 μg/ml or 25 μg/ml for recently weakened transformed cells. This antibiotic was only required for pLysS containing cells to maintain the T7 repressor encoding plasmid with an incorporated chloramphenicol resistant gene.

2.3.3 Isopropyl-β-D-galactosidase (IPTG)

1 M stock solution was prepared by dissolving 120 mg/ml in water and subsequent 0.2 μm filter sterilisation. Fresh solution (or aliquots stored at -20°C) was added to growth media to a final concentration of 0.5 mM to induce protein overexpression.

2.4 DNA Manipulation

2.4.1 Gene Origin and Expression Vector

To date, the Y31 variant of cystatin B, the first to be isolated and cloned (Jerala *et al.*, 1988) is the most characterised form. However, the E31 variant is the most widely-occurring form in the human population, and is the sequence present in Genbank (www.ncbi.nlm.nih.gov, accession number 20357564). Therefore work in this study has been performed on the E31 variant with a C3S mutation remaining from crystallographic studies (Rabzelj *et al.*, 2005). The WT-E31 cystatin B used is encoded on a pET-11a vector which was a kind gift from Eva Zerovnik of the Josef Stefan Institute, Ljubljana, Slovenia. The pET-11a vector places expression of cystatin B under control of the T7 promoter, and additionally carries the β-lactamase gene, which provides ampicillin resistance to transformed cells.

2.4.2 Competent Cells

Escherichia coli strain XL10-Gold was used for stock plasmid production and *E. coli* XL1-Blue was used in mutagenesis. BL21 (DE3)-Gold and BL21 (DE3)-Gold pLysS *E. coli* cells were used for protein expression. Glycerol stocks of untransformed strains in 15% glycerol 25% LB were stored at -80°C.

E.coli cells of the desired strain were streaked from glycerol stocks onto non-selective LB-agar plates and incubated at 37°C overnight (pLysS cells were grown on 25 μg/ml chloramphenicol plates to avoid plasmid loss). A single colony was picked and used to inoculate 5 ml of LB media, subsequently incubated at 37°C overnight with shaking at 230-250 rpm. 200 μl of the overnight culture was used to inoculate an identical 10 ml of LB, incubated at 37°C, 230-250 rpm until the optical density at 600 nm (OD_{600}) reached 0.6.

Cells were pelleted by centrifugation, supernatant discarded and the cells gently resuspended in 3.3 ml ice-cold RF1 buffer, pH 5.8:

30 mM KCH_3CO_2
100 mM RbCl
10 mM $CaCl_2$
50 mM $MnCl_4$
15 % glycerol

After 30 minutes of incubation on ice, the cells were again pelleted by centrifugation, the supernatant discarded and the cells gently resuspended in 1 ml ice-cold RF2 buffer, pH 6.5:

10 mM MOPS
10 mM RbCl
75 mM $CaCl_2$
15% glycerol

After a further 30 minute incubation on ice, the cells were divided into 200 μl aliquots and either immediately transformed or stored at -80 °C.

2.4.3 Transformation

1 µl of plasmid DNA was added to ~200 µl competent cells and incubated on ice for 30 minutes. Cells were transferred into a cold 14 ml Falcon tube, heat-shocked at 42 °C for 90 seconds and returned to ice for a further 2 minutes. 800 µl of non-selective LB media was added, and the cells incubated at 37 °C with shaking at 230-250 rpm for 90 minutes. 100 µl of undiluted, ten-fold and one hundred-fold diluted cells were plated on selective LB-agar and incubated overnight at 37 °C.

2.4.4 Plasmid Extraction

Plasmid DNA was extracted and purified from 5 ml overnight cultures of *E. Coli* XL1-Blue or XL10-Gold using QIAprep Spin Miniprep kits (Qiagen) according to the manufacturer protocol. The plasmid was eluted in 50 µl of autoclave sterilised deionised water, quantified and stored at -20°C.

2.4.5 Quantification of DNA Concentration

The concentrations of plasmid and oligonucleotide solutions were calculated using the optical density at 260 nm, where an absorbance reading of 1.0 is equivalent to a nucleotide concentration of 33 µg/ml. Plasmid preparations were considered free of protein contamination if the ratio $A_{260\ nm}/A_{280\ nm}$ was greater than 1.7. All absorbance readings were taken with a Varian Cary 50-Bio UV-Vis spectrophotometer.

2.4.6 Site-Directed Mutagenesis

Site-directed mutagenesis (SDM) was used for single amino acid substitution mutants or stop codon addition for truncate mutants of cystatin B. The QuikChange® II mutagenesis kit (Stratagene) was used in all the mutants produced in this study and the manual can be referred to for more details.

125 ng of each forward and reverse primer containing the desired mutation were added to a PCR reaction mix containing:

30 ng plasmid template,

1 µl dNTP mix,

5 µl 10x reaction buffer (100 mM KCl, 100 mM $(NH_4)_2SO_4$, 200 mM Tris-HCl pH 8.8, 20 mM $MgSO_4$, 1 % Triton® X-100, 1 mg/ml nuclease-free BSA).

The reaction mixes were then made up to a final volume of 50 µl with sterilised deionised water.

1µl Pfu Turbo DNA polymerase was added to the reaction. PCR was then performed in a Techne Progene thermocycler, using 18 cycles of:

1 minute at 95°C (melting)

1 minute at 55°C (annealing)

7 minutes at 68°C (extension).

Reactions were subsequently digested for 1 hour with Dpn1 restriction endonuclease to selectively digest methylated template DNA. Plasmid DNA was transformed into competent XL1-Blue *E.coli* cells and plated onto selective LB-agar, containing the relevant antibiotic. Colonies which grew on the selective agar were grown overnight and the plasmid extracted as described in section 2.4.4.

The primers were synthesised, a-la-carte by MWG and were optimised for site-directed mutagenesis using the Stratagene QuikChange® Primer Design Program.

2.4.7 DNA Sequencing

Plasmids produced by site-directed mutagenesis were sequenced to confirm the mutation by the Core Genomic Facility, University of Sheffield. Sequences were aligned and analysed using Finch TV Version 1.4.0 (Geospiza, Inc).

2.5 Protein Expression and Purification

2.5.1 Predicted Protein Physical-Chemical Characteristics

Predicted extinction coefficients, molecular weights and pKa were calculated from proteins sequences using the "ProtParam" tool on the ExPASy Proteomics Server at http://www.expasy.ch/tools/protparam.html (see table 2.5.1). This data was utilised in buffer choice to maximise protein solubility and ion-exchange suitability. The predictions additionally aided the choice of correct ultrafiltration and dialysis membranes, identification in SDS-PAGE and calculation of protein concentration.

Protein	pI	Predicted ε_{280nm}	Predicted Molecular Weight (Da)
WT-E31	6.97	4470	11 124
G4R-E31	8.04	4470	11 222
α-3C-β-E31	6.97	4470	11 927
D61X-E31	9.22	1490	6 565
R68X-E31	6.25	1490	7 421

Table 2.5.1 Predicted Protein Physical-Chemical Characteristics

Characteristics determined from the amino acid sequence using the ProtParam program.

2.5.2 Protein Overexpression

A single colony of freshly transformed *E. coli* BL21 (DE3) Gold cells was used to inoculate a 50 ml overnight starter culture of LB media containing 100 μg/ml ampicillin (additional 50 μg/ml chloramphenicol was added for pLysS cells). The cells were subsequently incubated overnight at 37°C, 230-250 rpm. The starter culture was diluted 1:100 commonly into 1 litre of selective LB media and incubated at 37 °C with shaking at 230 rpm until OD_{600} reached ≈ 0.6. Cultures were then induced with IPTG to a final concentration of 0.5 mM and mostly incubated for a further 4 hours at 37 °C, 230 rpm.

2.5.3 Cell Harvesting

Induced cells were harvested by centrifugation at 10 000 rpm, 4 °C for 10 minutes. The growth media discarded, and cell pellets resuspended in ~50 ml 10 mM sodium phosphate buffer, pH 6.0. Cell resuspensions were either sonicated immediately or stored at -80 °C.

2.5.4 Cell Lysis and Protein Solubilisation

Immediately before sonication ~0.1 mg/ml DNAse 1 (Sigma), one Complete protease inhibitor tablet (Roche Diagnostics) and $MgCl_2$ to a final concentration of 20 mM was added accordingly to 50 ml cell resuspensions. Cells were kept in an ice/water bath and sonicated at 15 μm intensity for 4 x 1 minute bursts with one minute cooling intervals between each sonication cycle. Cell debris was removed by centrifugation at 15 000 rpm, 4°C for 15 minutes. For WT-E31 cystatin B the soluble protein supernatant was directly used in further purification steps.

2.5.5 Purification

Protein cell extract was loaded onto a 100 ml SP-Sepharose (GE Healthcare) cation-exchange column at 2 ml/min, equilibrated in the case of WT with 10 mM sodium phosphate buffer, pH 6.0 (Kenig *et al.*, 2004). The column was washed until the OD_{280} of the eluent stabilised. Bound protein was then eluted with a 160 ml total volume buffer with a 0-500 mM NaCl gradient. Eluent was collected in 4 ml fractions and presence of cystatin B verified by SDS-PAGE (see figure 2.1). Cystatin B typically elutes at ~ 200 mM NaCl.

Fractions containing cystatin B were pooled and concentrated to ~10 ml volume (13 ml maximum) using an Amicon ultra-filtration stirred-cell device in conjunction with a Millipore regenerated cellulose membrane with a molecular weight cut-off (MWCO) of 10 000 or 3 000 Da. The concentrated protein was 0.2 μm filtered and loaded onto a 400 ml preparative Superdex 75 (GE Healthcare) size-exclusion column, equilibrated with filtered and degassed 10 mM appropriate buffer/100 mM NaCl (Kokalj *et al.*, 2007). Buffer was run through the column at a rate of 3 ml/min and 6 ml fractions were collected. The OD_{280} of the eluent was monitored and fractions identified by SDS-PAGE as containing pure cystatin B were pooled, 0.2 μm filtered and stored at 4 °C. A typical purification scheme for WT-E31 cystatin B is illustrated in figure 2.1.

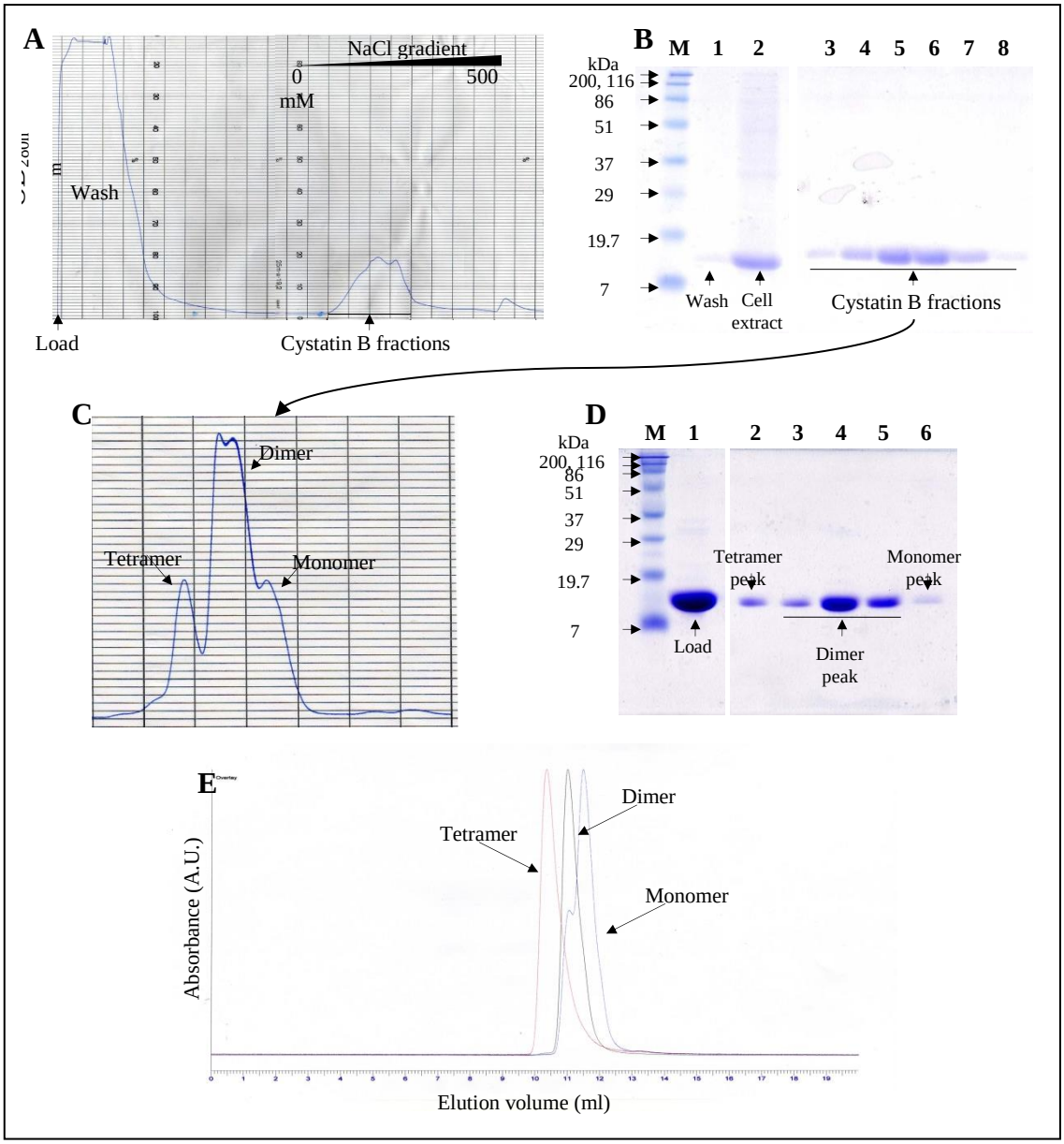


Figure 2.5.1 Purification of WT-E31 Cystatin B

A: Elution profile of WT-E31 from SP-sepharose ion exchange chromatography

B: SDS-PAGE of ion exchange fractions. Lanes are as follows: M: broad range pre-stained markers (BioRad), lane 1: wash, lane 2: cell extract, lanes 3-8: eluent with increasing salt concentration.

C: Elution profile of WT-E31 cystatin B from SEC. **D:** SDS-PAGE of SEC fractions. Lanes are as follows: M: broad range pre-stained markers, lane 1: load, lane 2: tetramer peak, lane 3-5: dimer peak, lane 6: monomer peak. **E:** SEC-HPLC trace showing purity and oligomeric state of purified cystatin B. Peaks are as follows: monomer is observed at 11.6 ml, dimer at 11.2 ml and tetramer at 10.6 ml. Diagram from Sharma (2009).

2.6 Protein Procedures

2.6.1 Electrophoresis

All SDS-PAGE was carried out using a Bio-Rad Mini Protean II apparatus. The standard method described below was used for all experiments except where described otherwise.

SDS-PAGE Buffers

4 x upper buffer	0.5 M Tris/HCl, 0.4 % (w/v) SDS, pH 6.8
4 x lower buffer	1.5 M Tris/HCl, 0.4 % (w/v) SDS, pH 8.8
Running buffer	25 mM Tris/HCl, 190 mM glycine, 0.1 % (w/v) SDS, pH 8.3
2 x loading buffer	100 mM Tris/HCl pH 6.8, 200 mM DTT, 4 % (w/v) SDS, 0.2 % (w/v) bromophenol blue, 30 % (v/v) glycerol. 200 µl aliquots were frozen at -20 °C and defrosted when required.

Gel Preparation

4 % stacking gels were cast above 16 % resolving gels as described below:

16 % resolving gel

- 2.5 ml 4 x lower buffer
- 4 ml 40 % Bis-acrylamide (acrylamide: bis-acrylamide ratio 37.5:1; Bio-Rad)
- 3.5 ml deionised water
- 100 µl 10 % (w/v) ammonium persulphate (APS)
- 10 µl N, N, N', N'-tetramethylethylenediamine (TEMED) (Bio-Rad)

4 % stacking gel

- 2.5 ml 4 x upper buffer
- 1.1 ml 40 % Bis acrylamide
- 6.4 ml deionised water
- 100 µl 10 % (w/v) ammonium persulphate (APS)
- 10 µl N, N, N', N'-tetramethylethylenediamine (TEMED) (Bio-Rad)

Samples were typically prepared with a 1:1 ratio of 2x loading buffer to protein solution, and 5 – 20 μl were loaded depending on the expected sample concentration. Samples were not heated prior to loading, as this has been observed to promote oligomerisation of cystatins (Rosie Staniforth, personal communication); instead samples were incubated at room temperature for approximately 10 minutes before loading.

Molecular Weight Markers

Bio-Rad pre-stained broad range molecular weight markers were used with the typical mass values stated below:

Myosin	200 000 Da
B-galactosidase	116 250 Da
Bovine serum albumin	86 000 Da
Ovalbumin	51 000 Da
Carbonic anhydrase	37 000 Da
Soybean trypsin inhibitor	29 000 Da
Lysozyme	19 700 Da
Aprotinin	7 000 Da

Electrophoresis

SDS-PAGE gels were electrophoresed for ~60 minutes at constant voltage of 180V. Cystatin B WT E31 runs at an apparent molecular weight of 15 kDa. Gels were stained with (0.25 % (w/v) Coomassie Brilliant Blue, 10 % (v/v) acetic acid, 45 % (v/v) methanol) for a minimum of 1 hour with gentle agitation at room temperature, then destained at room temperature with agitation (10 % (v/v) acetic acid, 45 % (v/v) methanol) until bands were clearly visible. Destained gels were scanned between two transparencies with an Epson Stylus DX3800 combination printer.

2.6.2 Protein Solution Concentration and Buffer Exchange

For large volumes (greater than 15 ml), an Amicon ultra-filtration stirred-cell was used in conjunction with an appropriate Millipore 3 or 10 kDa molecular weight cut-off (MWCO) filter. For volumes less than 15 ml, solutions were concentrated using Vivaspin centrifugal concentrators (Vivascience) with a MWCO of 3 or 10 kDa. For volumes less than 0.5 ml, a Microcon centrifugal concentrator (Millipore) with a 3 or 10

kDa MWCO was used. Buffer exchange was achieved either by repeated rounds of concentration and dilution with the required buffer, or by dialysis using Spectra/Por dialysis tubing with an appropriate 6 – 8 or 3 kDa MWCO.

2.6.3 Determination of Protein Concentration

UV absorption spectra were recorded on a Varian Cary 50-Bio UV-Vis spectrophotometer using a 500 µl Hellma quartz black-walled cuvette with a pathlength of 1 cm. The concentration of protein solutions were calculated using the Beer-Lambert Law:

$$A = \varepsilon \times c \times t$$

A is absorbance, ε is molar extinction coefficient (M^{-1} cm^{-1}), c is concentration (M), and t is pathlength (cm).

2.6.4 Analytical Size Exclusion Chromatography

The purity and oligomeric states of protein solutions was typically analysed by size exclusion high pressure liquid chromatography (SEC-HPLC) using a Shodex KW803 column and KW-G guard column (Shodex, Japan) with a Perkin Elmer Series 200 HPLC system equipped with a UV-visible absorbance detector. Usually, a 20 µl sample was injected, and the optical density (OD) of the eluent monitored at $OD_{280\ nm}$.

2.6.5 Dot Blot Immunoassay

Reagents

Primary antibody:	Human cystatin B Polyclonal goat IgG, raised against 2-98 (full length) *E. coli* derived recombinant cystatin B (R&D Systems, cat. no. AF1408).
Secondary antibody:	Rabbit anti-goat IgG (H and L chain) horse radish peroxidase conjugate (Calbiochem, cat. no. 401515-2ml).
Chemiluminescence Reagent:	Super Signal West Pico Stable Peroxidase solution and Luminol Enhancer (horse radish peroxidase) (ThermoFisher cat. no. 2124 and 1862123)

Buffers

Tricine Buffer Saline (TBS): 20 mM Tris-HCl, pH 7.5, 150 mM NaCl

TBS-Tween (TBS-T): As above with 0.05% v/v Tween 20

TBS-T-milk (TBS-TM): As above with 3% v/v low fat (1%) milk powder

Blotting

Serially-diluted samples (2µl) were spotted onto a marked grid of nitrocellulose blotting membrane (Amersham Hybond ECL, RPN3030) and dried at RT. Blocking of non-specific sites was carried out at RT for 30-60 minutes with TBS-TM. The membrane was then washed 3 times in d.H_2O and incubated with the primary antibody (2-0.2 µg/ml) for 30-90 minutes at RT. The membrane was then washed a further 3 times in d.H_2O and three times with TBS-T, the latter for 5 minutes each. Incubation of the secondary antibody (0.1 µg/ml) was performed for 30 minutes at RT, followed by a wash 3 times in d.H_2O and three times with TBS-T, the latter for 15 minutes (x1) and 5 minutes (x2) and once with TBS for 5 minutes. The samples on the membrane were then incubated for 2 minutes with the chemiluminescence reagents mixed in a 1:1 ratio followed by exposure of X-ray film in a dark room for 30 seconds, 5 minutes and 18 hours for reliable luminescence detection.

2.6.6 Circular Dichroism

Circular dichroism was performed with a Jasco J-810 spectropolarimeter using 0.1 cm pathlength cells over a 190-300 nm range. Spectra were acquired at RT or 25°C with a data pitch of 1 nm at a rate of 20-50 nm/min a suitable response time such as 8 or 2 seconds respectively and typically averaged over 5 accumulations. Mean residue elipticity, $[\Theta]_{MRW}$ (deg·cm^2 dmol^{-1}) was calculated to allow reliable comparison of different proteins using the equation:

$$[\Theta]_{MRW} = \frac{(\Theta \times 100 \times M_r)}{(c \times d \times N_A)}$$

Θ is the measured elipticity in degrees, M_r is the protein molecular weight in kDa, c is the protein concentration in mg/ml, d is the pathlength in centimetres, N_A is the number of amino acids per protein. The factor of 100 in the equation originates from the conversion of the molar concentration to dmol/cm^3 (Creighton, 1989).

2.7 Amyloid Fibre

2.7.1 Cystatin B Amyloid Formation

Cystatin B amyloid fibres were typically generated by quiescent incubation of 30 µM soluble protein at 37 °C in fibrilisation buffer (15 mM sodium acetate buffer, pH 4.7, 150 mM NaCl and 10 % 2,2,2, trifluorethanol (TFE)) (Zerovnik *et al.*, 2002b). The protein was first buffer exchanged by ultrafiltration in TFE-free fibrilisation buffer, for a maximum of a few days before triggering fibrilisation. Addition of TFE initiates the polymerisation process, and was thus added immediately before kinetics were measured. For work with mature amyloid fibres, the morphology was examined by transmission electron microscopy (TEM) following a minimum three week incubation. This incubation period ensured 100% soluble to insoluble protein conversion and almost exclusive fibre formation as determined by TEM and subsequent UV spectroscopy of soluble fractions following centrifugation.

2.7.2 Amyloid Fibre Formation Kinetics

Amyloid fibre formation was monitored by thioflavin-T (ThT) binding. When bound to amyloid fibres, the dye fluoresces maximally at ~480 nm (Naiki *et al.*, 1989, Naiki and Gejyo, 1999). To monitor fibrilisation kinetics, an aliquot of freshly prepared 10 mM ThT stock solution was added, where possible, to the batch of protein to be used in the fibrilisation, prior to triggering with the addition of 2,2,2, trifluoroethanol (TFE). In these conditions ThT has no observable effects on reaction kinetics (Giannini, 2004).

A ThT dye control was monitored for degradation and photobleaching during the incubation period alongside the protein preparations which were incubated in the dark. Spectra were recorded on Shimadzu RF-5301PC or Cary Eclipse spectrofluorimeters, using 1.4 ml fluorimetry quartz cuvettes (Hellma), with an excitation and emission pathlength respectively of 10 and 4 nm. Samples were excited at 442 nm and emission spectra were recorded from 450 – 600 nm. Emission at 482 nm was plotted as a function of time to follow kinetics of the fibrilisation reaction. Data analysis and fitting was performed typically in a manner described in (Hortschansky et al., 2005) using Microsoft Excel and Graphpad Prism 6.0 for Windows (GraphPad Software, La Jolla California USA). A single or double exponential function was used with the premise

that in the absence of a reliably determined multi-state kinetic model the basic assumption of concentration dependence is the simplest means of kinetic determination and line fitting for means of clarity.

2.7.3 Transmission Electron Microscopy

TEM negatively stained sample grids were prepared by addition of 2 µl sample solutions, for 1 minute, onto the glow-discharged surface of a carbon-coated copper/palladium grid (400-square mesh, Agar scientific). Samples were then blotted and washed twice in deionised water, then washed and blotted in 1% (w/v) uranyl formate stain and finally a 20 second soak in stain before blotting and drying. 200-250Å carbon-coated grids were prepared in-house to avoid contamination of polymer which is sometimes observed in commercially prepared grids, which increases the background noise. Uranyl formate stain is used to enhance contrast compared to the standard uranyl acetate. TEM was carried out on a Phillips CM100 instrument operating at 100 kV and magnification of 5 200 x, 11 500 x, 21 000 x typically and when warranted up to 52 000 x. Images were recorded on a 1024 x 1024 pixel Gatan CCD camera using Gatan Digital Micrograph 3 software.

Chapter 3: Structural Determination of Cystatin B Amyloid Fibre

3.1 Introduction

3.1.1 Background

The current cystatin B fibre structure model is primarily based on hydrogen/deuterium (H/D) exchange NMR data (Morgan, 2006) and is presented in the introduction (section 1.5.5). A major challenge facing structural biologists in the amyloid field is the observation of a variety of amyloid structures produced from the same protein or peptide (Petkova et al., 2006), in different conditions and even within the same preparation. The relevance of the currently posited model to different experimental conditions and indeed to other members of the cystatin superfamily, including the amyloidosis-linked cystatin C needed to be tested further. H/D exchange studies involve lengthy analyses and may not be applicable to all fibre preparations so different, more rapid methods were investigated for the characterisation of the cystatin B model system.

3.1.2 Limited Proteolysis

Limited proteolysis has been exploited successfully to analyse amyloid fibre structure in many systems including Aβ peptide (Kheterpal *et al.*, 2001), α-synuclein (Miake *et al.*, 2002), HET-s (Balguerie *et al.*, 2003a), Ure2p (Baxa *et al.*, 2003), PI3-SH3 (de Laureto *et al.*, 2003), bovine α-lactalbumin (de Laureto *et al.*, 2005), β_2-microglobulin (Myers *et al.*, 2006) and lysozyme (Frare *et al.*, 2006). Limited proteolysis occurs primarily at sites with high chain mobility and rarely within regular secondary structure. Therefore, conformational parameters such as accessibility, mobility and protrusion correlate with limited proteolysis data (Hubbard, 1998).

The use of serine proteases, which monomeric cystatin B inhibits (see figure 1.9.1) (Barrett *et al.*, 1986), allows for clear determination of fibre structure without the complication of soluble protein digestion products. The cystatin B protease binding motif is composed primarily of a constrained loop linking β-strand 2 and 3 (Abrahamson, 1993), which is inactivated by the straightening and extension of the β-

strands in the 3D domain swapped dimer and tetramer (Staniforth *et al.*, 2001, Kokalj *et al.*, 2007) and is also observed within the fibre (Morgan et al., 2008). Although the cystatin B monomer *in vitro* typically constitutes 10% of the protein population it remains in excess in the 1:1000 and 1:50 ratio protease: cystatin digests,thus leading to effective protease inhibition in soluble protein preparations. The inhibitory action upon elastase and proteinase K makes them highly useful proteases because any soluble cystatin B available through dynamic fibre exchange processes ('molecular recycling') is unlikely to be cleaved and thereby clarifying the results obtained in this ideal system.

3.1.3 Mass Mapping

Mass mapping of cystain B amyloid fibre produces a measurement of the mass per unit length (MPL), e.g. χ kDa/nm which can be used to determine the spacing of repeating units (Wall *et al.*, 2008). This measurement has been performed on several amyloid fibre forming proteins and forms a major part in amyloid fibre structural modelling (see section 1.2) (Petkova *et al.*, 2002). Knowledge of the mass of the fibre as well as its dimensions provides even greater restrictions on the models proposed, in the absence of a fuller 3-dimensional reconstruction of the structure from cryoEM data. STEM MPL can be used on particle masses from 30 kDa to 1 000 MDa with increasing accuracy the larger the particle, e.g. ±10% for 100 kDa particles to ±4% with 1 MDa particles.

These measurements were performed with a scanning transmission electron microscope (STEM) equipped with a digital dark annular field detector on unstained samples bound to a carbon coated copper grid. The electron elastic scattering from the sample atoms is directly related to the particle mass and was detected by the dark annular field detector. The intensity of the elastically scattered electrons was compared to the reference sample (tobacco mosaic virus) and the MPL of the sample then determined (Wall *et al.*, 2008). To measure the elastic scattering intensity each measurement requires the designation of a rectangular axial area around a specific fibre section and a suitable amount of surrounding background. The background intensity is then subtracted and the remaining intensity from the fibre integrated to provide intensity values. These intensity values were then calibrated against the tobacco mosaic virus (TMV) standard (131 kDa/nm) and the final intensities corrected for electron-irradiation mass loss. The intensity over a given span (1 nm) produces the final mass per unit length measurement (x kDa/nm).

3.1.4 Sequence Specific Amyloid Fibre Prediction

A prevalent theory posits that short regions of protein sequence modulate protein aggregation and amyloid fibre formation (Chiti *et al.*, 2003), which have been termed 'amyloid hotspots' (Conchillo-Sole *et al.*, 2007). These sequence based prediction programs are primarily based on a set of values in (Chiti *et al.*, 2003) whereby the rates of aggregation of unfolded polypeptide chain were correlated with physicochemical properties such as hydrophobicity, secondary structure propensity and charge. Therefore an appraisal of the currently available amyloidosis sequence prediction programmes was performed and matched to the experimental data available.

3.1.5 Combining Structural Data

In this chapter, data is presented on the limited proteolysis of cystatin B to produce a map of protease accessible regions external to the fibre core. Electron Microscopy and AFM has been previously used by the Zerovnik group to determine cystatin B fibre dimensions. However, it was felt that in-house determination was essential due to potential variation in condition and thus morphological differences. Mass per unit length measurements were then carried out at the University of Manchester in collaboration with Dr David Holmes. By combining the results with existing data from hydrogen-deuterium exchange NMR spectroscopy (Morgan *et al.*, 2008), EM and atomic force microscopy (AFM) (Zerovnik *et al.*, 2002a), the results significantly challenge the current model.

3.2 Materials and Methods

3.2.1. Amyloid Fibre Formation

The human cystatin B used in this study is the 98 amino acid E31 polymorph, C3S mutant (Rabzelj *et al.*, 2005) expressed and purified by the method described in section 2.5. Amyloid fibres were formed by 37°C incubation of 30 µM cystatin B in pH 4.7, 15 mM sodium acetate buffer with100 mM NaCl, 1 mM sodium azide and 10% (v/v) 2,2,2, trifluoroethanol (Zerovnik *et al.*, 2002b). All pH dependent fibrilisations were carried out using the method described in section 2.7. The pH dependent fibrilisations were performed at 37°C, in pH 3.3 (no TFE), pH 4.0, 4.7, 6.0 and 7.4 (see table 3.2.1) with 10% TFE with WT cystatin B and were trialled along with WT G4R monomer/dimer mixtures and G4R high molecular weight oligomer (HO) at pH 4.7, 10% TFE.

<table>
<tr><th>pH</th><th>Component 1</th><th>Component 2</th><th>Component 3</th><th>Component 4</th></tr>
<tr><td>3.3</td><td>15 mM glycine/HCl</td><td>260 mM sodium sulfate</td><td rowspan="5">1 mM sodium azide</td><td>N/A</td></tr>
<tr><td>4.0</td><td rowspan="2">15 mM sodium acetate/acetic acid</td><td rowspan="4">150 mM NaCl</td><td rowspan="4">10% v/v 2,2,2 trifluoroethanol</td></tr>
<tr><td>4.7</td></tr>
<tr><td>6.0</td><td rowspan="2">10 mM Na_2HPO_4/NaH_2PO_4</td></tr>
<tr><td>7.4</td></tr>
</table>

Table 3.2.1 pH Dependant Fibrilisation Buffer Table

Soluble cystatin B preparations and amyloid fibres are entirely composed of full length 1-98 residue protein, determined by MS. Fibre morphology was examined by TEM following a minimum three week incubation which leads to 100% soluble to insoluble protein conversion and almost exclusive fibre formation as determined by TEM and subsequent UV spectroscopy of soluble fractions following centrifugation. For limited proteolysis the fibre preparation was concentrated to 90 µM monomer equivalent and exchanged into buffers suitable for each protease (table 3.2.2) by centrifugation at 13 000 g $RCF_{(max)}$ for 15 minutes, the supernatant removed, replaced with the proteolysis buffer and centrifuged for a further 15 minutes before repeating the process.

Protease	Protease: Cystatin B Mass Ratio	Protease Final Conc.	Cystatin Conc.	Digest Buffer	Inhibitor	Inhibitor Final Conc.
First WT Digest						
Endoproteinase Lys-C	1:1000	36 nM, 1 µg/ml	90 µM, 1.0 mg/ml	25 mM Tris-HCl, pH 8.0	TLCK	10 µM
Elastase	1:1000	38 nM, 1 µg/ml		10 mM Tris-HCl, pH 8.0	PMSF	100 µM
Pepsin	1:500	57 nM, 1 µg/ml		15 mM Na Acetate, pH 4.0	Pepstatin A	7 µg/ml
Proteinase K	1:1000	35 nM, 1 µg/ml		10 mM Tris-HCl, pH 8.0, 1 mM CaCl	PMSF	5 mM
Second WT Digest						
Elastase	1:50	380 nM, 20µg/ml	90 µM, 1.0 mg/ml	25 mM Tris-HCl, pH 8.0	PMSF	5 mM
Pepsin	1:50	570 nM, 20µg/ml		15 mM Na Acetate, pH 4.7	Pepstatin A	100 µg/ml
Proteinase K	1:50	350 nM, 20µg/ml		10 mM Tris-HCl, pH 8.0, 1 mM CaCl	PMSF	5 mM

Table 3.2.2 WT Cystatin B Limited Proteolysis Experimental Parameters

3.2.2 Limited Proteolysis Reaction

Proteinase K from *Tritirachium album limber*, elastase from porcine gut mucosa and pepsin were used in this study due to their wide coverage of predicted cut sites within the soluble cystatin B protein whilst endoproteinase Lys-C was used for greater specificity. Proteolysis reactions were carried out initially at 1:1000 protease: cystatin B ratio (by mass) at 25 °C with and then in a 1:50 ratio at 37 °C except for pepsin which was initially carried out at a 1:500 ratio.

Proteinase K digestion was carried out in 10 mM Tris-HCl pH 8.0 buffer with 1 mM $CaCl_2$ and quenched with a final concentration of 5 mM phenylmethanesulfonyl fluoride (PMSF). Elastase digestion was carried out in 10 mM Tris-HCl pH 8.0 buffer and quenched with a final concentration of 100 µM and 5 mM phenylmethanesulfonyl fluoride (PMSF) respectively in the 1:1000 and 1:50 protease: protein digests. Pepsin digestion was carried out initially in 15 mM sodium acetate buffer, pH 4.0 and then in the same buffer at pH 4.7 and inhibited with 0.7µg/ml and 0.1 mg/ml of pepstatin A respectively in the 1:500 and 1:50 protease: protein digests. Endoproteinase Lys-C digestion was performed in 25 mM Tris-HCl, pH 8.0, 1 mM ethylenediaminetetraacetic acid (EDTA) buffer and the reaction was inhibited by a final concentration of 10 µM N_α-Tosyl-L-lysine chloromethyl ketone hydrochloride (TLCK). All proteases and inhibitors were purchased from Sigma Aldrich, with the same supply used in all

experiments, and stored frozen at -20°C. Fresh aliquots were used for each digestion and stored according to the manufacturer instructions to maintain activity.

The proteolysis method was devised to determine the hydrolysis of cystatin B fibre with separation of the still intact fibre and fragments in solution and is described in figure 3.2.1 and in detail below. The reaction solution containing fibre in proteolysis buffer was equilibrated at the desired reaction temperature for ≥ 5 minutes prior to proteolysis. Protease was added to the reaction solution and a T = 0 hour sample was immediately removed and the reaction quenched with an inhibitor. An SDS-PAGE aliquot was taken and an equal volume of loading buffer added. The remaining aliquot was centrifuged at 13 000 g $RCF_{(max)}$ for 15 minutes. The supernatant was retained for analysis, the pellet washed by repeating centrifugation in fresh reaction buffer and the final pellet resuspended in an equal volume of 6 M GuHCl. The aliquots were initially stored on ice and frozen at – 20 °C shortly after processing. Each time point was treated in an identical manner. Controls included; a soluble protein digest for each protease and condition, fibre in proteolysis buffer to monitor molecular recycling and dissociation in the buffers used, protease alone and protease with inhibitor. Controls were all treated in an identical manner to the digests to allow observation of autolysis and subsequent erroneous fragment detection.

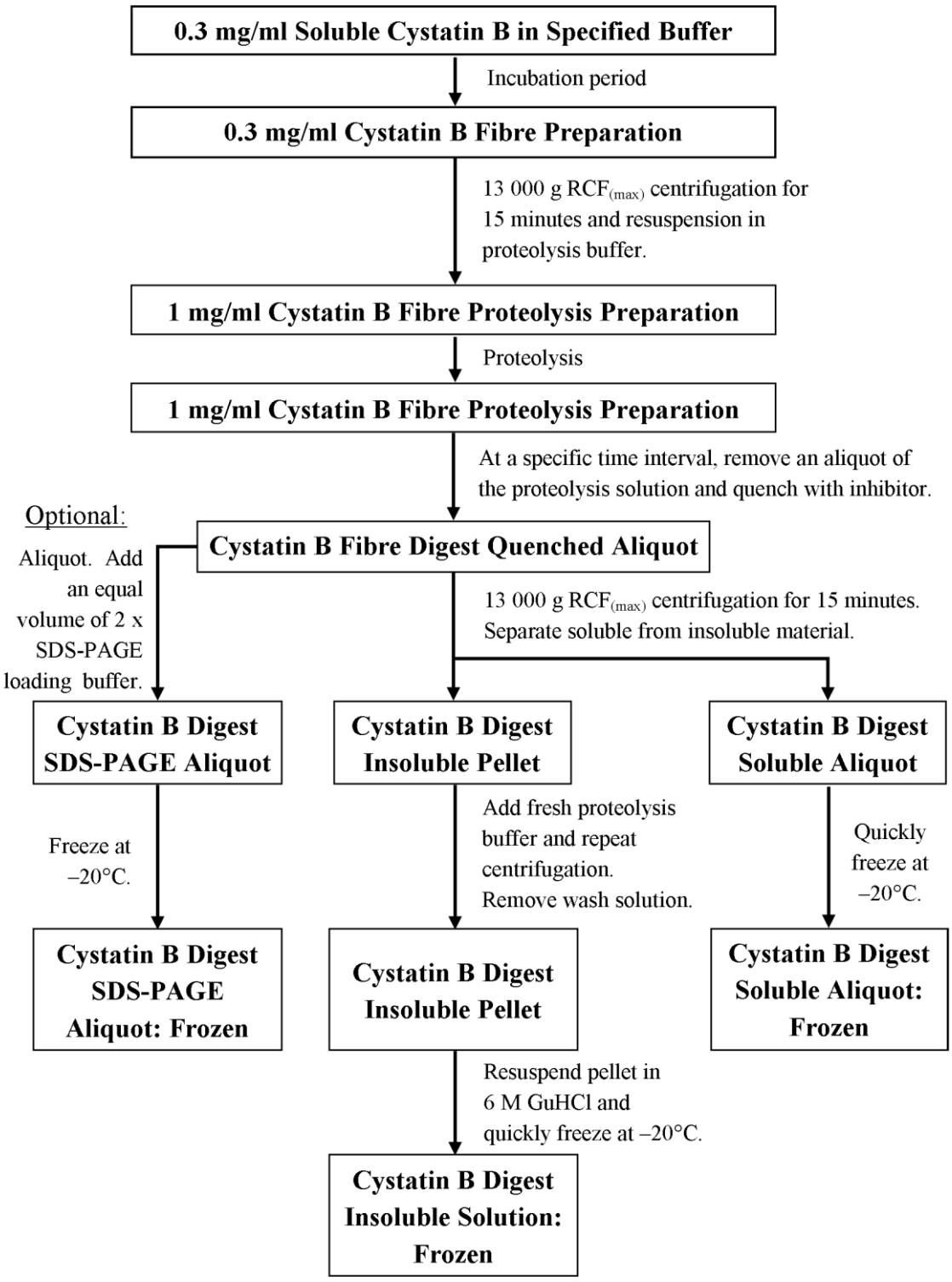


Figure 3.2.1 Limited Proteolysis Sample Preparation Method

The method described can be used for amyloid fibres from most proteins with some specific alterations. The SDS-PAGE sample is suggested to be optional as its use was limited in fragment analysis, although is useful in inferring that hydrolysis has been effective before use of MS equipment.

3.2.3 Protease Selection and Experimental Design Implications

This study used proteinase K, elastase and pepsin due to their wide coverage of predicted cut sites within the soluble cystatin B protein, whilst endoproteinase Lys-C was used for greater specificity in specific regions. Proteinase K (from *Tritirachium album limber*) is one of the most active serine endopeptidases known (three times more active than bovine trypsin) with little cleavage specificity and can readily degrade natively folded protein to individual amino acids. Carboxyl groups of aliphatic, aromatic and hydrophobic amino acids show some preferential cleavage (Ebeling *et al.*, 1974). Proteinase K is predicted to cleave cystatin B at 41 sites spread across the entire protein sequence (cleavage typically at residues F,Y,W,G,A,V,L,I and T). Predicted cut sites were determined from amino acid sequence with PeptideCutter software, which is available online via the Expasy server. See figure 3.2.2 for the predicted cut sites map, produced with the assumption of complete accessibility.

Elastase (from porcine pancreas) is a serine protease which typically hydrolyses amino acids with straight side chains; alanine >> valine > leucine > glycine and serine (Atlas *et al.*, 1970, Thompson and Blout, 1973, Gold and Shalitin, 1975), unless proceeded immediately by proline or preceded by proline-X-Y- (e.g. P,A,A,A) (Thompson and Blout, 1973), thus 31 specific cleavage sites are predicted (see figure 3.2.2). Endoproteinase Lys-C (from *Lysobacter enzymogenes*) is a serine protease and hydrolyses lysine residues with nine cleavage sites predicted in cystatin B using PeptideCutter. Some minor non-specific cleavage with this protease has also been reported.

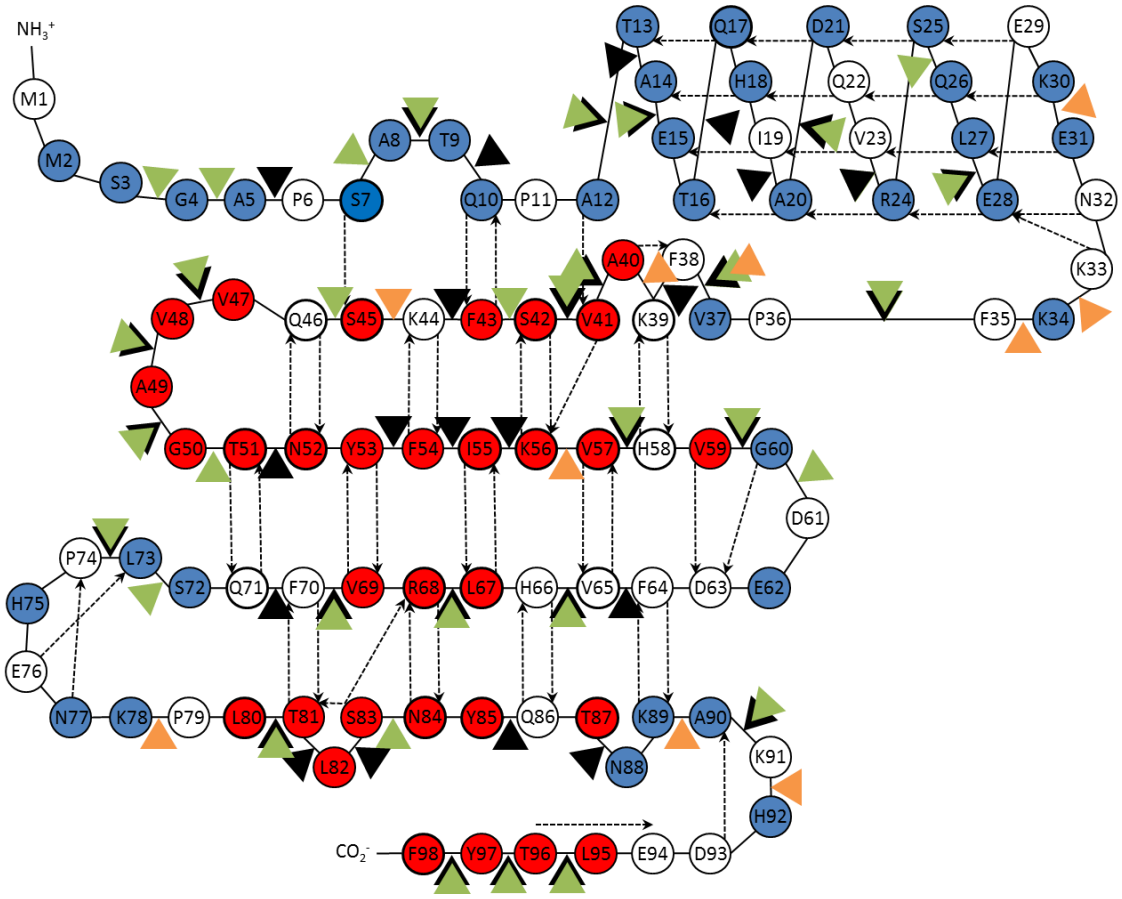


Figure 3.2.2 WT Cystatin B Topology Map of Predicted Hydrolysis Positions Based on Sequence Alone

The predicted cleavage positions shown are structural model free and based purely on the sequence. **(Black triangles)** proteinase K proteolysis, **(green triangles)** elastase and **(orange triangles)** endoproteinase Lys-C digestion. The topology map is modified from (Morgan *et al.*, 2008), where hydrogen bonding present in the dimer is shown using dashed arrows. The regions of the molecule involved in the formation of secondary structure in the fibre are shown as H/D exchange protected residues (red) whereas unstructured regions of the amyloid fibre which readily exchange hydrogen for deuterium are highlighted in blue. Residues for which no H/D exchange data is available are coloured white.

3.2.4 SDS-PAGE

Tricine-SDS-PAGE (Schagger, 2006) was used to resolve the low molecular weight fragments generated and stained with coomassie R250 blue in 45:45:10 v/v water: methanol: acetic acid staining solution.

3.2.5 Reverse Phase HPLC

A Waters C4 reverse-phase (RP) column (300Å pore, 5.0 µm beads, 4.6 mm x 250 mm column) was used on a Perkin Elmer Series 200 HPLC machine over a linear gradient of 5-20% aqueous-organic solvent over 20 minutes and 20-80% over the 20 mins at 0.8 ml/min. Aqueous phase: milli-Q H_2O (0.1% trifluoroacetic acid, TFA), organic phase acetonitrile (0.1% TFA). Effluent was monitored by recording absorbance at 224 nm. This methodology is described as C4-RP-LC.

3.2.6 Reverse Phase HPLC-ESI-MS and Mass Fragment Identification

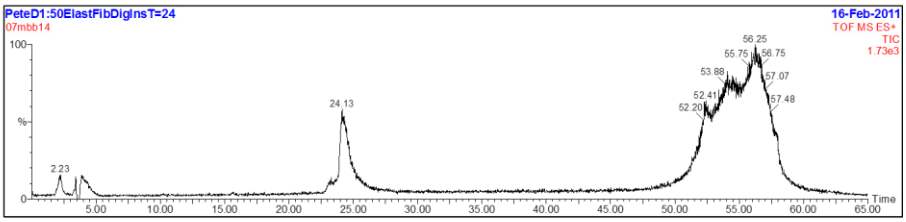
A Grace Vydac Everest C18 reverse-phase column (300Å pore, 5.0 µm beads, 2.1 mm x 250 mm column) was run on a Waters Alliance 2695 HPLC machine over a linear gradient of 5-50% aqueous-organic solvent over 40 minutes and 50-95% over 5 mins at 0.2 ml/min. Aqueous phase: milli-Q H_2O (0.1% formic acid), organic phase acetonitrile (0.1% formic acid). Masses were detected by a Micromass LCT electrospray-ionisation time-of-flight instrument (Waters Corporation, Manchester) in positive ion mode. This method is described as C18 RP-HPLC-ESI-TOF-MS. Masslynx 3.5 software (Waters Corporation, Manchester) was used to determine masses present and FindPept (ExPASy) (Artimo *et al.*, 2012) was used to elucidate protein fragment identities by matching predicted and actual masses. See figure 3.2.3 for a flow diagram of the analysis methodology.

Cystatin B Digest Samples

Samples of interest transferred to vials

RP-HPLC-MS

Performance of RP-HPLC-ESI-TOF-MS. Example chromatogram:



Mass Spectra Interpretation

Manual determination of masses primarily using the maximum entropy function to determine the mass of large multi-charged species. Example below.

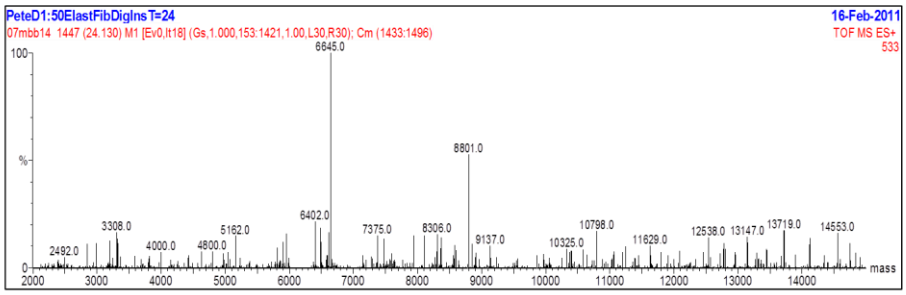


Table of Fragment Masses

Send the fragment masses to the FindPept server which matches the experimental mass values to the mass of predicted fragments using the protein sequence.

FindPept tool

The entered sequence is:

```
        10         20         30         40         50         60
MMSGAPSATQ PATAETQHIA DQVRSQLEEK ENKKFPVFKA VSFKSQVVAG TNYFIKVHVG

        70         80         90
DEDFVHLRVF QSLPHENKPL TLSNYQTNKA KNDELTYF
```

98 Amino Acids.

Entered peptide masses: 6645.000 8801.000

Tolerance: ±1.5 daltons

Using **average** masses of the occurring amino acid residues and interpreting your peptide masses as **[M]**.

Cysteine in reduced form.

New FindPept Search | PeptideMass | FindMod | GlycoMod

FindPept documentation
Mass values and considered PTMs

Matching peptides for unspecific cleavage:

User mass	DB mass	Δmass (daltons)	peptide	position
6645.000	6643.613	-1.386	(V)RSQLEEKENKKFPVFKAVSF KSQVVAGTNYFIKVHVGDED FVHLRVFQSLPHENKPL(T)	24-80
6645.000	6646.438	1.438	(S)ATQPATAETQHIADQVRSQL EEKENKKFPVFKAVSFKSQV VAGTNYFIKVHVGDEDFVH(L)	8-66
8801.000	8800.938	-0.062	(H)IADQVRSQLEEKENKKFPVF KAVSFKSQVVAGTNYFIKVH VGDEDFVHLRVFQSLPHENK PLTLSNYQTNKAKHDE(L)	19-94
8801.000	8800.938	-0.062	(I)ADQVRSQLEEKENKKFPVFK AVSFKSQVVAGTNYFIKVHV GDEDFVHLRVFQSLPHENKP LTLSNYQTNKAKHDEL(T)	20-95

Peptides are displayed with the adjoining residues before cleavage greyed out and in parentheses.

SDS-PAGE and N-Terminal Sequencing

Digest time points with ambiguously determined fragments were selected, run on SDS-PAGE, blotted onto a nitrocellulose membrane and individual bands were N-terminal sequenced to aid identification.

Fragment Identification

Figure 3.2.3 Limited Proteolysis Mass Spectroscopy and Fragment Mass Identification Method

3.2.7 Cystatin B Fibre Diameter

To measure the fibre width, 16 electron microscopy preparations produced at pH 4.7, 37 °C with 30µM cystatin B (15 mM sodium acetate/acetic acid buffer, 150 mM NaCl and 10 % TFE) and collected over a period of 7 years by 3 PhD students, were each analysed independently, and contributed a maximum 30 values each to ensure no single preparation produced errant results skewing the average. Most of the images used were collected by the author, due to the requirement of high resolution images and selection of regions with clearly defined single fibres, parameters not particularly required in previous studies where only fibre identification was needed. Digital Micrograph 3 software (Gatan, UK) was used to generate electron intensity histogram profiling perpendicular to the fibre axis to accurately determine and measure the fibre width. Images were collected on a Phillips CM100 (Philips/FEI Corporation, Holland) electron microscope with a 1024 x 1024 pixel Gatan CCD (Gatan, UK) camera using 21 000 x magnification negatively stained electron micrographs, thus each pixel is equivalent to 1 nm and defines the maximum level of accuracy as ± 1 nm.

3.2.8 Mass Mapping

MPL measurements were made from a standard preparation of cystatin B amyloid fibre (see section 2.7) exchanged into ultrapure water by centrifugation at 14 000 g, supernatant removal, water addition and the process repeated twice. These fibres were adsorbed for two minutes onto a 600 mesh copper grid with 3 nm thick carbon film, washed with ultrapure water and left to air dry for 15 minutes and then dried further in the EM vacuum. 50 STEM annular dark field (ADF) 48 000 x electron micrographs of unstained cystatin B amyloid fibres were obtained on a Tecnai 12 TEM/STEM instrument (Electron Optics, Eindhoven, The Netherlands), equipped with a high angle, dark-field detector and digital scan generator. This instrument is capable of a spacial resolution of ~2 nm (Holmes *et al.*, 2001). Image analysis was performed using SEMPER (Synoptics, Cambridge, UK), whereby for each measurement a rectangular axial area of fibre and surrounding background was selected, the background intensity subtracted and the remaining intensity from the fibre integrated to provide intensity values. These intensity values were then calibrated against tobacco mosaic virus (TMV) standard (131 kDa/nm) and the final intensities corrected for electron-irradiation mass loss.

3.2.9 Sequence Specific Amyloid Prediction

TANGO

TANGO, available at http://tango.crg.es/ is designed to predict β-sheet aggregation and not amyloid formation, however the two show a good correlation. TANGO is a statistical thermodynamics algorithm, which predicts the percentage occupancy of the β-aggregation conformation, hydrophobicity and estimates desolvation penalties. Peptides are considered to have some aggregation tendency when they possess segments of at least 5 consecutive residues populating the β-aggregated conformation. TANGO is unable to accurately predict low levels of aggregation propensity (<5% total sequence) (Fernandez-Escamilla *et al.*, 2004) and inherently may fail to predict all hotspots in larger polypeptides due to the system being constrained to a maximum detection of 2 hotspots (Pande, 2004). TANGO does not account for position specific differences (Maurer-Stroh *et al.*, 2009) and cannot in principle quantitatively compare polypeptides differing widely in sequence. (Mutants can be compared in identical conditions.) (Fernandez-Escamilla *et al.*, 2004). Overall the authors of TANGO determine the prediction program is 87% accurate (Fernandez-Escamilla *et al.*, 2004).

WALTZ

WALTZ is available at: http://switpc7.vub.ac.be/cgi-bin/submit.cgi, and combines a position-specific sequence profile with multiple physical property parameters (helical, β-sheet and hydophobicity propensities), complemented with structural modelling data from the all-atom force field FoldX to identify motifs that can nucleate amyloid fibre formation in proteins. A position-specific scoring matrix for amyloid propensity was generated from a modified and expanded non-redundant hexapeptide database (AmylHex) of experimentally verified amyloid fibre forming hexamers. A 6 residue polypeptide is the minimal amyloid nucleating core, which can as an insertion induce amyloid fibre formation of an entire protein domain, thus is the justification for using hexameric sequences in the prediction program (Maurer-Stroh *et al.*, 2009).

AGGRESCAN

AGGRESCAN available at: http://bioinf.uab.es/aggrescan/ is based on aggregation propensity values for individual amino acids derived from *in vivo* experiments and utilises an experimentally determined hotspot threshold (HST). The program calculates the average aggregation potential over a sliding window of given length and assigns the average value to the central residue in the window. The window size used is: 5 residues for < 75 amino acids, 7 for <175, 9 for <300 and 11 for >300. The reasoning for this window size change is that for longer sequences, larger 'hotspots' are required to significantly increase the aggregation propensity (due to the more extensive native folding), whilst short-stretches suffice for smaller peptides. The hotspot threshold (HST) is defined as the average of the aggregation potential of the 20 amino acids weighted by their frequencies in the SwissProt database. A region in the polypeptide sequence is then considered an aggregation hotspot if there are 5 or more sequentially continuous residues with an average aggregation potential larger than the HST and none of them are an aggregation breaking proline (Conchillo-Sole *et al.*, 2007).

ZYGGREGATOR

ZYGGREGATOR Combines the aggregation propensity and folding propensity for each individual amino acid which is then averaged in a seven residue window, or smaller at the N and C-terminal (Tartaglia *et al.*, 2008). This method also accounts for specific patterns of alternating hydrophobic and hydrophilic residues. Prediction of folding propensity is calculated with the CamP program which predicts flexibility and solvent accessibility (Fernandez-Escamilla *et al.*, 2004). Zygregator is available on http://www-vendruscolo.ch.cam.ac.uk/zyggregator.r.php, which considers the pH and http://www-vendruscolo.ch.cam.ac.uk/zyggregator.plus.php, which considers many environmental conditions.

AmylPred

Amylpred, available at http://biophysics.biol.uoa.gr/AMYLPRED/input.html uses a consensus method from five different prediction systems: average packing density (Galzitskaya *et al.*, 2006), conformational switches (ability to convert from α helix to β sheet) (Hamodrakas *et al.*, 2007), sequence patterns (hexapeptide) that match amyloid formation (Lopez de la Paz and Serrano, 2004) TANGO (β-sheet propensity) and conformational energy values (Zhang *et al.*, 2007). The consensus method works by each method individually predicting amyloid propensity, any residue that is individually

predicted twice to be amyloid forming is deemed as likely to be amyloidogenic. The arbitrary consensus methodology has an inherent conservatism, thus this system is likely to produce a larger number of false-negatives, although minimise false positives produced from any one prediction programme.

Conformational Energy

Conformational energy (3D profiling) is based on the threading of six-residue peptides through the known crystal structure of the cross-β spine formed by the hexapeptide NNQQNY from yeast Sup35 PrP and also an ensemble of 'near-native' templates (n=2 511). The conformational energy fit is then evaluated using ROSETTADESIGN. A permissive and conservative energy threshold is then used to determine the amyloidogenicity of the hexapeptide. Conformational energy is claimed to reach an accuracy of ≈80% when a conservative energy threshold is used to separate between amyloid peptides and amorphous β peptides. Cysteine or proline residues have to be removed as they cause errors in the energy calculations thus producing ambiguous results (Thompson *et al.*, 2006). The conformational energy program by (Zhang *et al.*, 2007) is based on this method and is available through the AmylPred server.

Other Programs

Average Packing Density, is based upon the observation that regions with strong packing density are likely to form amyloid. This prediction program measures the observed number of residues within the given distance from the considered residue and if this value is greater than the threshold it is determined to be amyloidogenic. The authors determine the predictions are consistent in 8 out of the 12 tested disease-related amyloid-forming proteins (Galzitskaya *et al.*, 2006). Conformational Switches, utilises the observation that sequence regions with high amyloid propensity show α-helix – β-sheet ambivalence (Hamodrakas *et al.*, 2007). Pattern, searches for peptide sequences that are known to individually form amyloid fibre from experimentally derived peptide databases (Lopez de la Paz and Serrano, 2004).

3.3 Results

3.3.1 Limited Proteolysis

The strategy was to carry out limited proteolysis of WT cystatin B fibres using four carefully chosen proteases with the reaction quenched at specific time points. Fragment analysis was then carried out with reverse-phase HPLC coupled to an electro-spray ionisation time of flight mass spectrometer (RP-HPLC-ESI-MS). Mass spectrometry peaks were then analysed manually and the corresponding fragment mass identified using the FindPept tool available on the Expasy proteomics site.

Formation and Characterisation of Amyloid Fibre

Cystatin B fibre preparations produced at 37 °C, pH 4.7 and 10 % 2,2,2, trifluoroethanol (TFE) consistently produce long unbranched smooth fibres which extensively laterally associate and share identical morphology in TEM micrographs, shown in panels A and B of figure 3.3.5. These amyloid fibre preparations are free of Cystatin B soluble species following centrifugation and measurement by UV spectroscopy, sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) or size-exclusion (SEC-HPLC). In addition, no significant quantities of non-fibrillar aggregate or high molecular weight oligomer was detected by TEM.

Fibre Stability

H/D exchange protected residues within the cystatin B amyloid fibre exchange at a rate of around 10^{-4} s^{-1}, independent of pH and residue type (Morgan *et al.*, 2008). This indicates that each individual amino acid, backbone amine group becomes available for exchange on average every 2.8 hours. Although this exchange may be concerted for individual molecules of cystatin within the fibre, this lifetime still supports the use of proteolysis in analysing the structure of cystatin B fibres without a significant contribution from soluble molecules cycling on and off the fibre assembly ("molecular recycling").

Given the above considerations, a significant population of soluble full length protein was nevertheless clearly detected by RP-HPLC in a pepsin digestion of cystatin B amyloid fibre at pH 4.0. This free protein could potentially have introduced erroneous

amyloid fibre digestion results, thus the digest results were discarded. We propose that Cystatin B amyloid fibres generated at pH 4.7 are perhaps unstable at pH 4.0 or are more readily perturbed by pepsin addition in these conditions. A further pepsin fibre digestion was performed at pH 4.7, with a 1:50, protease: cystatin B ratio in which the fibres remained intact but undigested following 24 hour incubation at 37 °C, thus indicating insufficient pepsin activity in these conditions.

At pH 8.0, in the proteinase K digestion buffer, a peak corresponding to 5% of the total protein was observed in the soluble fraction of the fibre control detected by C4 RP-HPLC. This could be attributed to a number of factors; (i) fibre being removed from the pellet and subsequently dissolving in the water/acetonitrile solvent (ii) low levels of soluble cystatin in fibre preparations, previously undetected (iii) the fibre is mildly destabilised at pH 8.0 contrary to previous work indicating stability over a period of months at pH 9.0 (Morgan *et al.*, 2008). All these factors could contribute, although the latter two are less likely to be responsible for the observed protein, following previous experimental work. This population is thus most likely the result of experimental limitations (i) and is very unlikely to alter the results obtained.

To conclude, soluble protein cleavage can be reduced by minimising complete digestion time to a matter of a few hours by manipulating the protease concentration and temperature. In addition, the primary data used has been collected from fibres separated by centrifugation and dissolution in denaturant, thus further ensuring the digestion fragments observed are only from protein remaining incorporated in fibrous material. Very small populations of soluble cystatin B were occasionally observed by MS in the fibre digests, approaching the limits of detection. This is most likely the result of carry-over from the centrifugation step, and is not entirely removed in the wash step. Overall confidence in the results presented is high due to the various methods and features employed in the experiment to minimise soluble protein presence and digestion in the fibre digests.

Soluble Cystatin B Digestion

Cystatin B is a strong inhibitor of serine proteases as well as cysteine proteases. This property was therefore employed to ensure that only inactived oligomeric forms of the protein would be digested and form part of our analysis. Soluble cystatin B control digests were performed identically to fibre digests and analysed with SDS-PAGE and C4-RP-LC and C18 RP-HPLC-ESI-TOF-MS to ensure suitable protease inhibition in the conditions used. In 1:1000 protease: soluble cystatin B digests, proteinase K, elastase and endoproteinase Lys-C showed complete inhibition with no proteolysis intermediates detected by all the methods used (see figure 3.3.1 for SDS-PAGE gels of the digest). Only in a 1:50 proteinase K: soluble cystatin B digest were small N-terminal cleavage events detected. For instance at T = 0.5 hours the full size 98 residue protein is detected alongside , 4-98, 6-98 and 11-98 truncate species. Following 24 hours of proteolysis, only 4-98 residue species are detected. No soluble digest controls contained extensively digested medium sized protein fragments similar to those observed in fibre digests. Once cleaved the short peptide generated is readily cleaved further into several small species (< 15 residues). Therefore the results of the amyloid fibre digests are clearly and unambiguously from hydrolysis of amyloid fibre and not soluble protein.

Cystatin B Amyloid Fibre Digestion

The procedure employed identifies cystatin B protein fragments that remain integrated into the fibre core, see tables 3.3.1 and 3.3.2. To map the cystatin B amyloid fibre core components, fibre was digested for 0.5, 1, 4 and 24 hours, after which the insoluble fibre digest material was separated by centrifugation, the pellet washed, centrifugation repeated and the pellet, dissolved in 6 M guanidine hydrochloride followed by C18 RP-HPLC ESI-TOF-MS analysis. Figures 3.3.2 and 3.3.3, show that the deepest N-terminal (NT) cleavage site frequently observed in digested but intact amyloid fibre is up to residue 27 and produced by proteinase-K whilst both proteinase-K and elastase digests up to the C-terminal (CT) residue 80. The NT 'native α-helical region' is highly susceptible to proteolysis (see figure 3.3.4), thus confirming the unstructured nature of this region in the cystatin B amyloid fibre, and corroborates the H/D exchange data (Morgan *et al.*, 2008). However, in simple terms, the observation of rapid proteolysis in the CT contradicts the high levels of H/D exchange protection observed.

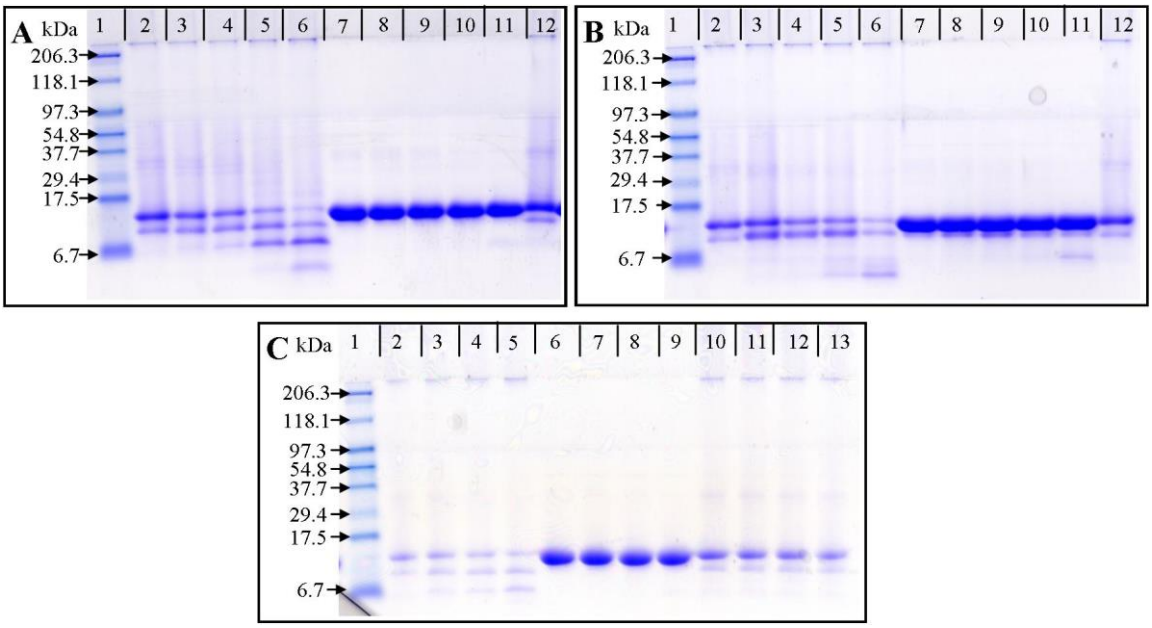


Figure 3.3.1 1:1000 Protease: WT Cystatin B Limited Proteolysis SDS-PAGE

In all gels Lane 1: Bio Rad Broad Range Pre-Stained Marker. **(A)** Elastase Digest. **(B)** Endoproteinase Lys-C Digest. **(A and B)** Lane 2-6 T = 0, 0.5, 1, 4, 24 hr amyloid fibre digest insoluble products, Lane 7-11 T = 0-24 hr soluble cystatin B digest indicating protease inhibition and lane 12 which contains the control fibre and shows two fragments, which are not seen in the soluble preparation of the protein and could be the result of an SDS-resistant compact post-fibre species. **(C)** Proteinase K Digest: Lane 2-5 T = 0, 0.5, 1, 4 hr amyloid fibre digest insoluble products, Lane 6-9 T = 0-4 hr soluble cystatin B digest indicating protease inhibition, lane 10-13 contains the T = 0-4 hr control fibre. All digests over time show at least 2-3 bands from amyloid fibre proteolysis. When compared to the multiple peaks observed in RP-HPLC and MS it is clear SDS-PAGE does not resolve sufficiently, although is a useful tool for simply detecting protease activity.

The observation that compared to the NT, the CT exhibits lower susceptibility to proteolysis, is shown in figure 3.3.4 which shows the average percentage value of N and C-terminal cleavage per unit time. The lower susceptibility of the CT to hydrolysis is indicated by fewer CT cleavage fragments detected in earlier digestion time points. For instance, proteinase-K and elastase at T = 0.5 hours in the 1:1000 digests indicate extensive N-terminal cleavage whilst the C-terminal remains entirely intact, up the limit of detection.

Digest Time (hrs)	Fragment Mass (Da)	Fragment Identity	Fragment Mass (kDa)	Fragment Identity
	Proteinase K 1:1000		**Proteinase K 1:50**	
0.5	8427	27-98 (0.6)	6274	1-57(1.8),27-80(1.7),5-61(0.1),15-68(1.1)
	9462	18-98 (0.7)	6488	25-80(0.6),27-82(1.5)
	9995	13-98 (1.8)	8430	18-90(1.4)
	10161	11-98 (0.4)		
	10860	3-98 (1.1)		
	10991	2-98 (1.3)		
1	8220	19-89 (0.3), 11-82 (0.3)	8428	27-98(0.7)
	8427	27-98 (0.6)	6273	1-57 (0.8),27-80(0.8),5-61(1.1)
	9463	18-98 (0.3)		
	10161	11-98 (0.4)		
	10646	6-98 (0.1)		
	10861	3-98 (0.1)		
4	6488	25-80(0.2), 27-82 (1.1)	6273	1-57 (0.8),27-80(0.8),5-61(1.1)
	7523	26-90(1.5), 27-91(1.5), 18-82 (1.5)		Complete Digestion
	7609	18-83(0.4), 25-90(0.5)		
	8429	27-98(1.4)		
	9464	18-98(1.4)		
	10863	3-98(1.6)		
	10994	2-98(1.2)		
	11125	1-98 (0.0)		
24	N/A	N/A	N/A	Complete Digestion
	Elastase 1:1000		**Elastase 1:50**	
0.5	8798	24-98 (1.0)	6859	24-82(1.9)
	9820	15-98 (1.0)	8801	19-94(0.1), 20-95(0.1)
	9994	13-98 (0.8)	6645	24-80(1.4), 8-66(1.4)
	10161	11-98 (0.4)	8054	7-77(1.0),13-82(1.9)
	10860	3-98 (1.1)	8491	6-80(0.4)
	10991	2-98 (1.3)		
1	8220	19-89 (0.4), 11-82 (0.4)	6644	24-80 (0.4)
	8798	24-98 (1.0)	8799	24-98(0.0),19-94(1.9),20-95(1.9)
	9820	15-98 (1.0)	8488	23-95(1.4),12-85(0.4),21-93(0.4),24-96(0.6)
	9992	6-93 (0.2), 13-98 (1.2)		
	10161	11-98 (0.4)		
	10860	3-98 (1.1)		
	10990	2-98 (2.3)		
4	6857	24-82 (0.9)	6644	24-80 (0.4)
	6945	6-67 (1.2), 9-69 (1.2), 20-79 (1.1), 24-83 (0.0)	8800	24-98(1.0),19-94(0.9),20-95(0.9)
	8052	13-82	8489	12-85(1.4),21-93(1.4),24-96(0.4),6-80(1.6)
	8139	13-83		
	8220	19-89 (0.4), 11-82 (0.4)		
	8306	20-91 (0.4), 11-83 (1.4)		
	8798	24-98 (1.0)		
	9153	18-96 (0.7)		
	9067	17-94 (0.8)		
	9992	6-93 (0.2), 13-98 (1.2)		
	10160	11-98 (1.4)		
24	6858	24-82 (0.1)	6644	24-80 (0.4)
	6945	24-83 (0.0)		
	8800	24-98 (0.5)		

Table 3.3.1 WT Cystatin B Amyloid Fibre Limited Proteolysis Digestion Fragments from Proteinase K and Elastase

The bracketed figure in the fragment identity column is the number of mass unit variance from the experimental and predicted fragment mass.

Digest Time (hrs)	Fragment Mass (Da)	Fragment Identity
	Endoproteinase Lys-C 1:1000	
0.5	9056	10-88 (0.2), 11-89 (0.2)
	9755	3-89 (0.9), 12-96 (1.0)
	9886	2-89 (1.1)
	10161	11-98 (0.4)
	10860	3-98 (1.1)
	10991	2-98 (1.3)
	11123	1-98 (1.5)
1	9056	10-88 (0.2), 11-89 (0.2)
	9414	3-86 (1.5), 6-88 (1.5)
	9755	3-89 (0.9), 12-96 (1.0)
	9886	2-89 (1.1)
	10161	11-98 (0.4)
	10860	3-98 (1.1)
	10991	2-98 (1.3)
	11124	1-98 (0.5)
4	5172	45-89 (0.2), 44-88 (0.2)
	9056	10-88 (0.2), 11-89 (0.2)
	9414	3-86 (1.5), 6-88 (1.5)
	9755	3-89 (0.9), 12-96 (1.0)
	9886	2-89 (1.1)
	10161	11-98 (0.4)
	10860	3-98 (1.1)
	10991	2-98 (1.3)
24	9052	2-82 (0.7), 15-92 (0.2), 18-95 (0.8)
	9252	2-84 (0.6)
	9417	2-85 (0.9)
	9756	3-89 (0.4)
	9888	2-89 (0.4)
	10018	1-89 (0.5)
	10862	3-98 (0.9)
	10992	2-98 (0.7)
	11125	1-98 (1.0)

Table 3.3.2 WT Cystatin B Amyloid Fibre Limited Proteolysis Digestion Fragments from Endoproteinase Lys-C The bracketed figure in the fragment identity column is the number of mass unit variance from the experimental and predicted fragment mass.

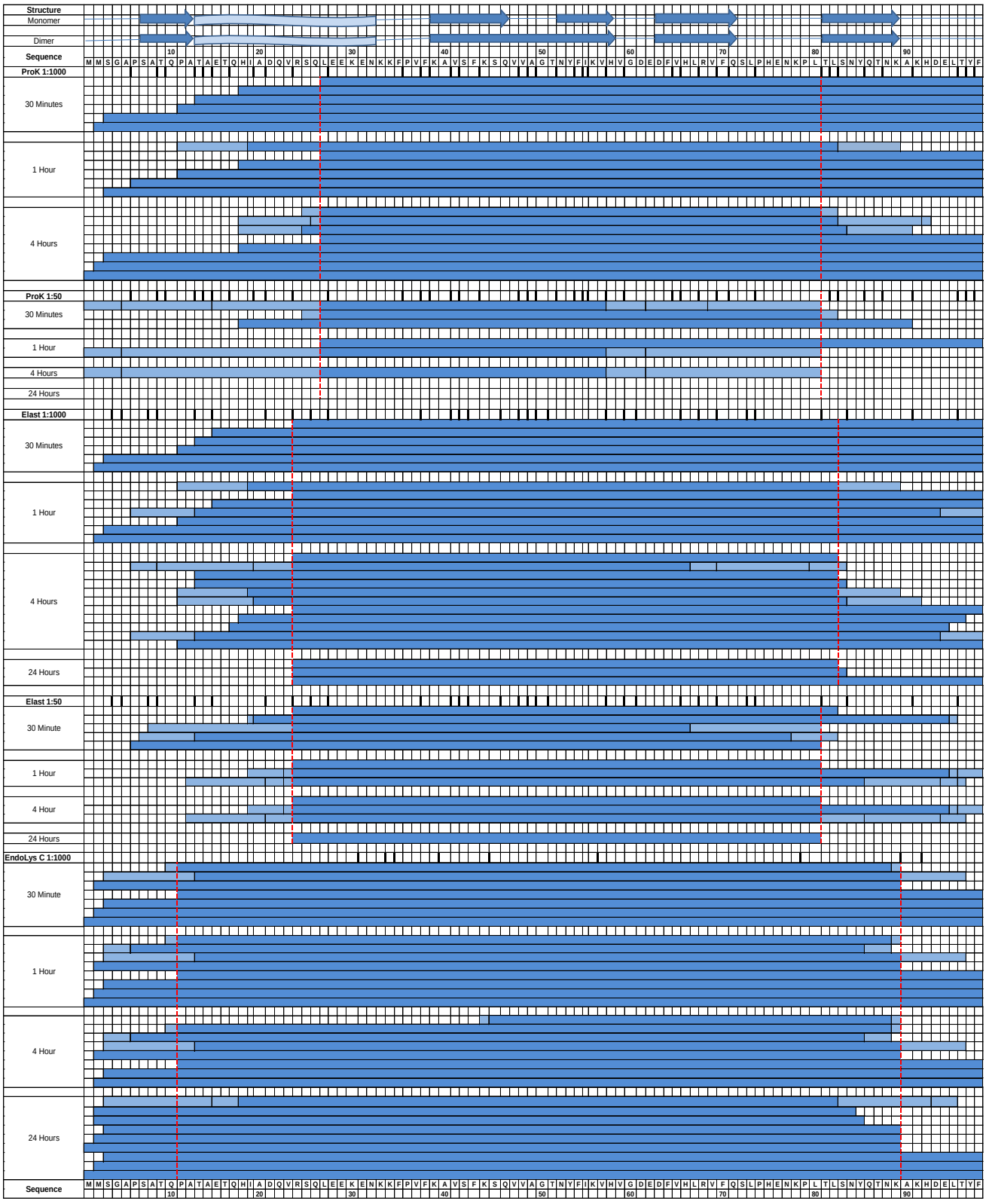


Figure 3.3.2 Limited Proteolysis Digestion Map of WT Cystatin B Amyloid Fibre 1:1000 and 1:50 ratio protease digestions are shown, with the dark blue bars showing clearly defined protein digest fragments or regions, whilst the light blue bars indicate ambiguous fragment regions. The far left column indicates protease and ratio used, plus the digestion time. ProK represents proteinase K, Elast is elastase and EndoLysC is endoproteinase Lys-C. At the top of the diagram, the cystatin B native secondary structure elements of monomer and domain swapped dimer are shown along with the amino acid sequence. The thick black lines indicate predicted specific protease cleavage sites. Red dotted lines define the reliable cleavage maxima.

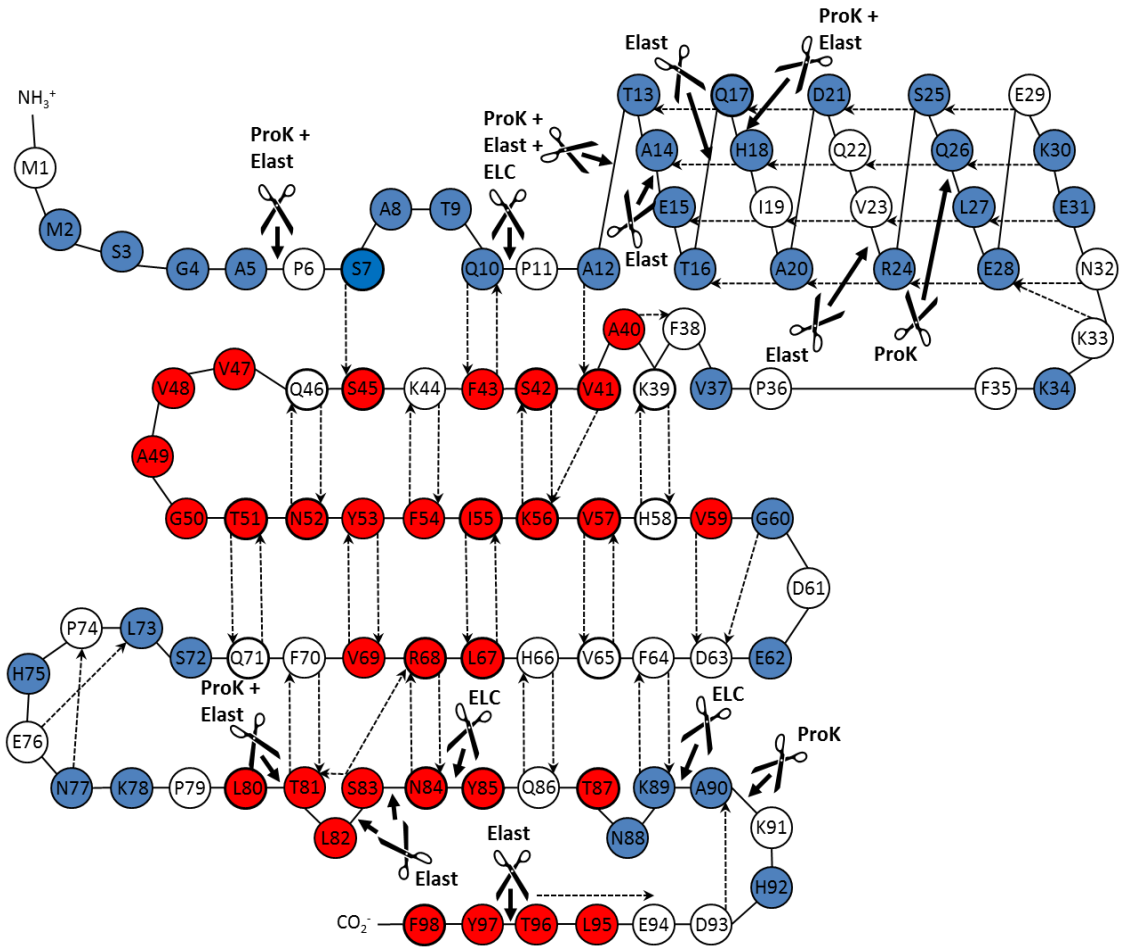


Figure 3.3.3 WT Cystatin B Topology Identifying Positions of Hydrolysis

Cleavage positions are identified by 'scissors' and the protease responsible is labelled: **(ProK),** proteinase K; **(Elast)** elastase, **(ELC)** endoproteinase Lys-C. Topology map modified from (Morgan *et al.*, 2008), where hydrogen bonding present in the monomer is shown using dashed arrows. The regions of the molecule involved in the formation of secondary structure in the fibre are shown as H/D exchange protected residues (red) whereas unstructured regions of the amyloid fibre which readily exchange hydrogen for deuterium are highlighted in blue. Residues for which no H/D exchange data is available are coloured white.

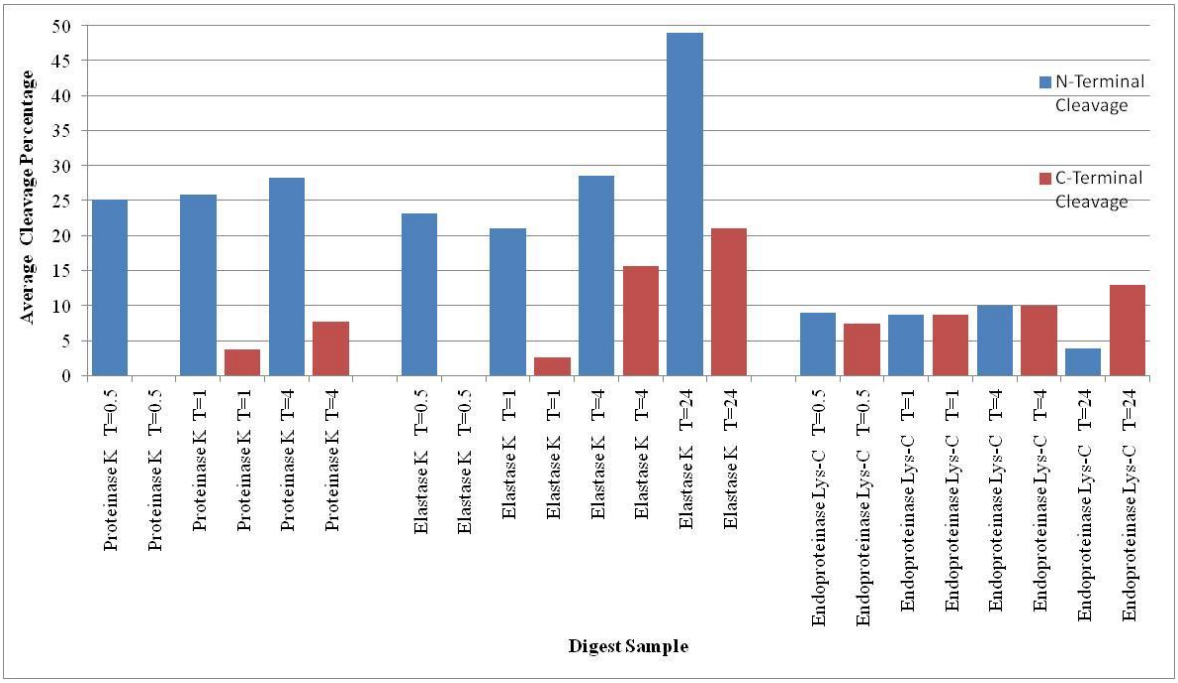


Figure 3.3.4 Terminal Region Accessibility in the 1:1000 Limited Proteolysis Digest

Average percentage of N and C-terminal cleavage per unit time. The N and C termini is described as the first and last 49 residues in the 98 amino acid protein respectively. The percentage cleaved value is calculated for an individual proteolysis fragment and averaged for each digest time-point. This graph shows that the CT is less susceptible to proteolysis than the readily cleaved NT region.

The proteases used in this study rapidly cleave soluble protein fragments released from fibre digestion as indicated by the multitude of <1000 Da polypeptides detected in mass spectrometry of the soluble fraction of fibre digests. The deepest NT and CT fibre digestion products, if produced in a single co-ordinated reaction would have a mass of 2714 Da (NT) and 2173 Da (CT). Fragments of this size however were not detected, thus the protease either digests frequently, perhaps in a step-wise exogenous manner, commencing at the terminal ends along the unstructured region and/or once cleaved from the fibre the larger soluble peptides are rapidly hydrolysed into smaller fragments. Identification of the soluble protein fragments by the method employed was not possible due to the mass differences being insufficient to determine distinct identities. The rapid fragment proteolysis indicates that protease inhibition is not significantly limiting in amyloid fibre digests.

The proteases used in this study rapidly cleave soluble protein fragments released from fibre digestion as indicated by the multitude of <1000 Da polypeptides detected in mass spectrometry of the soluble fraction of fibre digests. The deepest NT and CT fibre digestion products, if produced in a single co-ordinated reaction would have a mass of 2714 Da (NT) and 2173 Da (CT). Fragments of this size however were not detected, thus the protease either digests frequently, perhaps in a step-wise exogenous manner, commencing at the terminal ends along the unstructured region and/or once cleaved from the fibre the larger soluble peptides are rapidly hydrolysed into smaller fragments. Identification of the soluble protein fragments by the method employed was not possible due to the mass differences being insufficient to determine distinct identities. The rapid fragment proteolysis indicates that protease inhibition is not significantly limiting in amyloid fibre digests.

Proteinase-K is shown in the 1:50 digests at 24 hours to have promoted complete amyloid fibre destruction most likely by extensive protein-wide cleavage of fibre sub-units leading to subsequent dissociation of the products. The described process most likely occurs at the fibre termini which may exhibit less protected, partially associated species as suggested in molecular dynamics simulations of polyethylene-like polymers (Yamamoto, 2010). An additional process to consider is the possibility that unfolded region cleavage reduces fibre stability. Additional stability inferred from the unfolded regions may occur through a mechanism in which the unfolded regions transiently and loosely interact and thus protect the fibre from sub-unit melting. Furthermore, unfolded regions may sterically hinder protease access to the fibre core or act as a preferred 'suicide' substrate.

Notably, cleavage of β-strands 2, 3 and 4 does not readily occur within the insoluble material of a fibre digest. This indicates that this β-sheet segment remains integral to the fibre core, whilst the α-helix and β-strands 1 and 5 are readily cleaved from the fibre without clearly causing fibre dissociation. The persistence of this fibre core region strongly suggests that the structure is maintained independently of the regions which have been removed, see Figure 3.3.6. Without major structural rearrangement, which seems unlikely, it is difficult to reconcile the current H/D exchange based cystatin B fibre model with these results and thereby suggests the model is incorrect.

Morphology of Partially Digested Material

Cystatin B amyloid fibre following protease digestion remains fibrous and shows no clear morphological change in electron micrographs (see figure 3.3.5), thus suggesting the β-sheet core remains intact and stable throughout the digestion of extraneous unfolded regions of the protein. Fibre width remains the same following 1:50 ratio protease: cystatin proteolysis. The average fibre width after a 24 hour elastase digest was 9.7 ±1.5 nm (SD, n=21, max 13.3, min 7.7) and a proteinase K digest for 4 hours produced fibres with an average width of 9.4 ±2.3 nm (SD, n=15, max 14.4, min 6.7). The deviation of these values from the average undigested fibre (8.6 nm) is most likely due to experimental error including a small sample size and the error remains within a single standard deviation. The maintenance of fibre width indicates that no major structural event or morphology change is likely to have occurred.

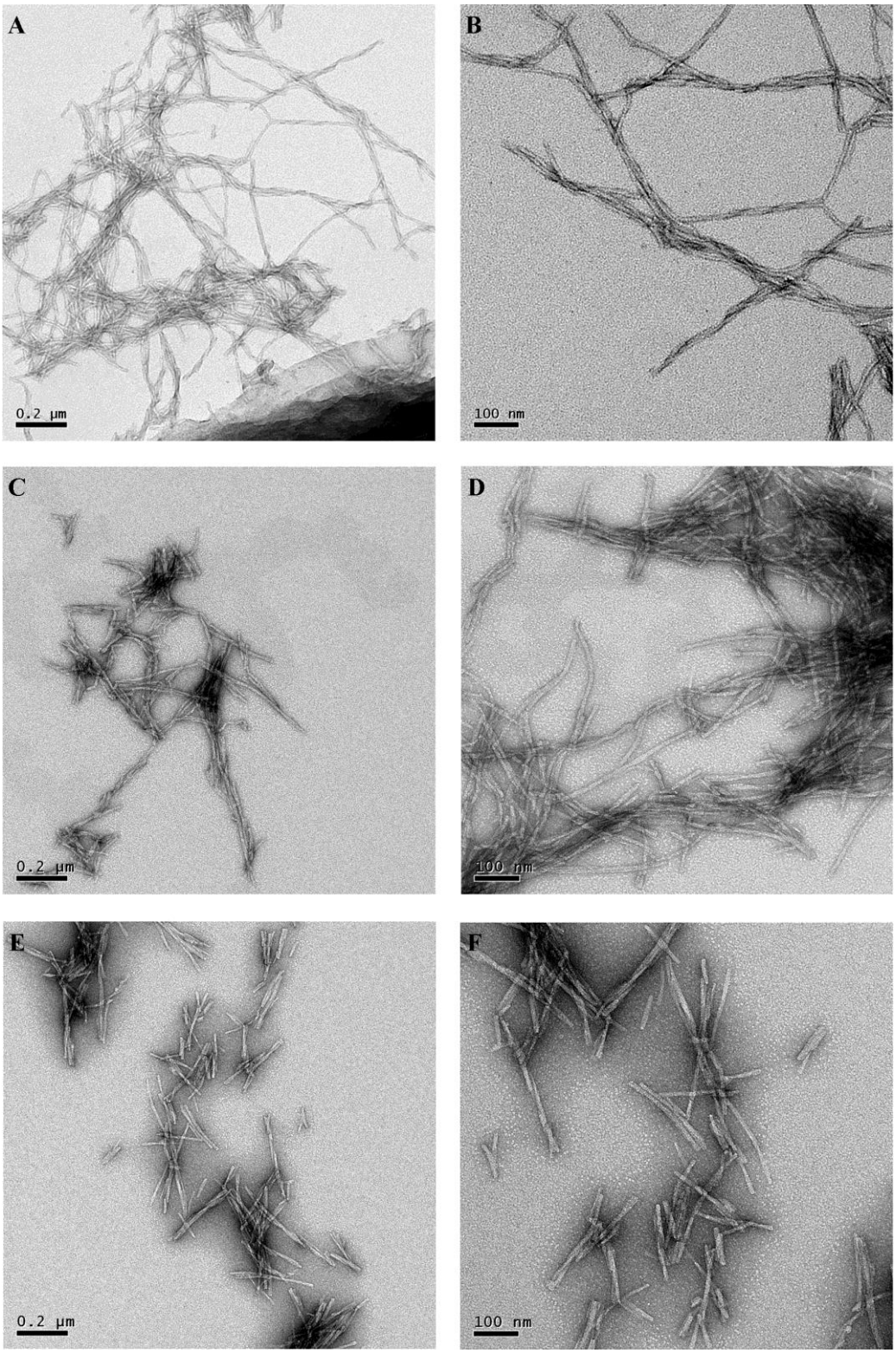


Figure 3.3.5 Electron Microscopy of Hydrolysed WT Cystatin B Amyloid Fibres

EM images following limited proteolysis of WT cystatin B fibres in a 1:50, protease: cystatin B ratio with 11 500x magnification to the left and 21 000x magnification on the right. **(A and B)** Fibres prior to digestion. **(C and D)** Proteinase K digested fibres after 4 hours. **(E and F)** Elastase digested fibres after 24 hours. Shortened fibres were sometimes observed, which most likely occurs through the repeated pipetting required for analysis and not by hydrolysis, although the latter cannot be entirely ruled out.

Fibre Polymorphism

Limited proteolysis of fibres produced at varying pH was used to explore cystatin B amyloid fibre polymorphism as observed in other amyloid forming proteins and peptides such as Aβ (Petkova *et al.*, 2005), the null hypothesis being that fibres produced in different conditions have the same morphology. pH dependent fibrilisation thioflavin T screening assays were initially performed with the help of undergraduate project students (Jennifer Thompson, Lauren Wareham, and Oliver Raimond) in an attempt to produce cystatin B amyloid fibres in conditions closer to those experienced physiologically. The pH 3.3, 4.7 and 6.0 conditions formed fibres at approximately the same elongation rate, (see figure 3.3.6) with increasing lag times as the pH increased. This indicates that the formation of the amyloidogenic nucleus is pH dependent whilst the nucleus/fibre templating is pH independent. pH 4.0 was an exception with increased polymerisation rate and no observable lag phase, however this does not correlate with the trend or previous results. The pH 7.4 conditions failed to yield detectable fibres in the timescale measured. EM of the fibres formed in the varied pH (figure 3.3.7) confirmed the presence of fibrillar species in all incubations.

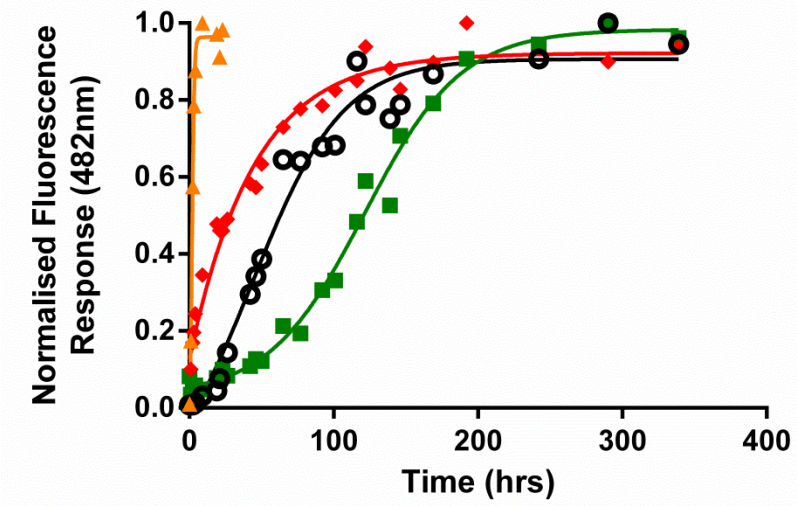


Figure 3.3.6 Thioflavin T Assay of WT Cystatin B Fibrilisation at Varying pH

(Red diamonds) pH 3.3 conditions (15 mM glycine, 260 mM Na_2SO_4 buffer), **(yellow triangles)** pH 4.0 (15 mM sodium acetate buffer, 150 mM NaCl, 10% v/v TFE), **(black open circles)** pH 4.7 (15 mM sodium acetate buffer, 150 mM NaCl, 10% v/v TFE) and **(green squares)** pH 6.0 (10 mM sodium phosphate buffer, 150 mM NaCl, 10% v/v TFE). Fluorescence measurements were normalised for comparison because direct fluorescence yields are difficult to interpret considering the non-linear nature of thioflavin T pH-dependence. Fit performed using a variable slope model for reasons of clarity and thus can only be used qualitatively.

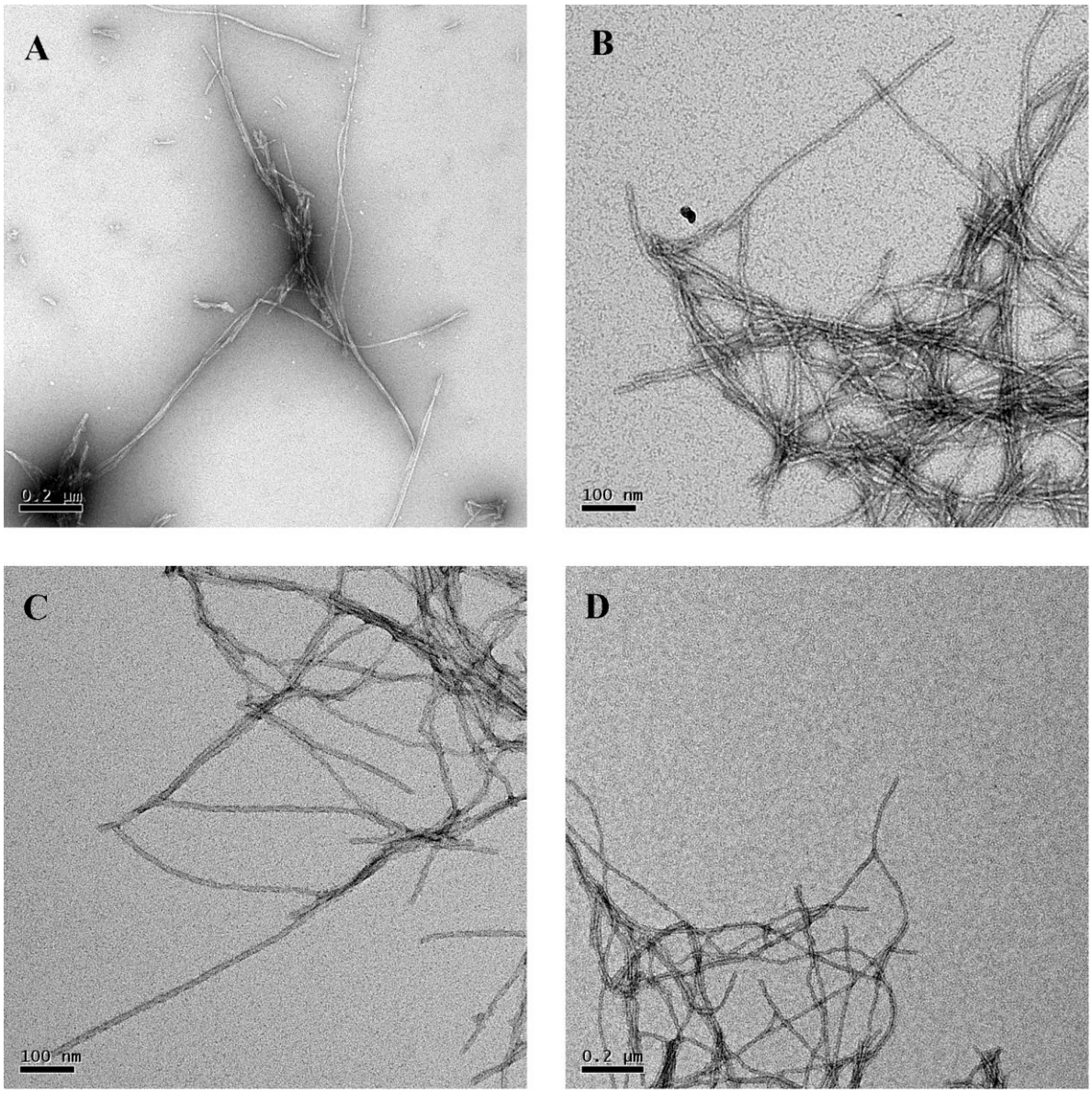
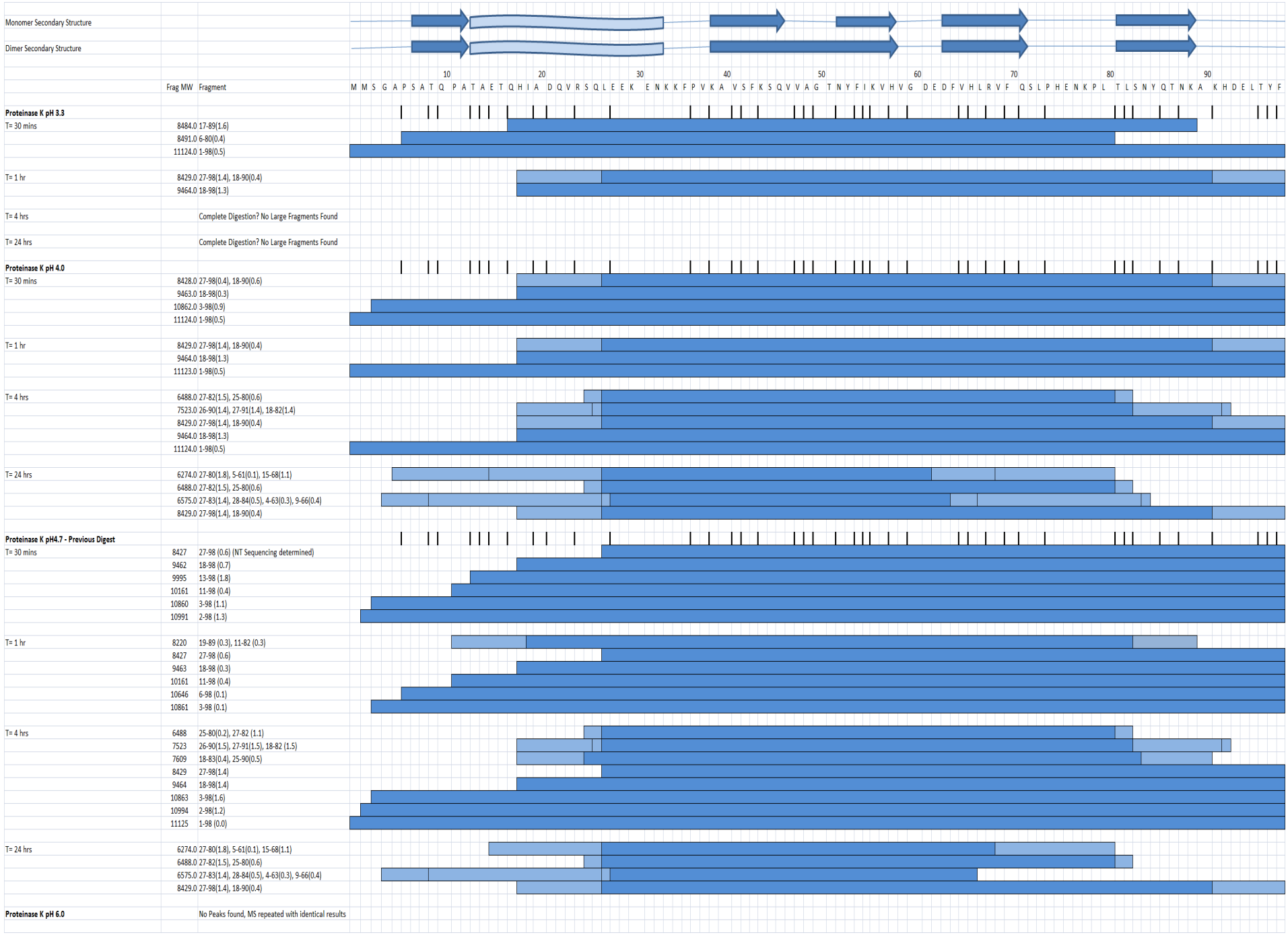


Figure 3.3.7 EM of Cystatin B Amyloid Fibre Produced at Varying pH

The fibres produced at varying pH all show similar morphology. **(A)** pH 3.3 generated fibre, 11 500 x magnification. **(B)** pH 4.0 formed fibre, 21 000 x magnification. **(C)** standard pH 4.7 fibre, 21 000 x magnification. **(D)** pH 6.0 fibre, 11 500 x magnification.

To probe for fibre polymorphism, limited proteolysis was performed on the fibres produced at varying pH. The digestion profiles from cystatin B fibres produced at pH 3.3 and 4.0 were similar to that of fibres generated at pH 4.7, thus indicating that the results are reproducible (see figure 3.3.8). These results support the hypothesis of polymorphism absence in these fibre preparations. However, although the fragments generated from fibres grown at pH 3.3, 4.0 and 4.7 are the same, the rate of proteolysis appears to be faster in the pH 3.3 fibres compared to the others tested. This could be due to fibre polymorphism, which has not been observed in EM. However, subsequent repeats of this proteolysis (by other members of the lab) indicates that the rate of hydrolysis can vary significantly and is most likely due to differences in the activity of the protease preparations and not the morphology of fibre samples. Reduced mass spectrometer sensitivity was ruled out by repeating the fragment analysis with a high concentration protein control being performed, nevertheless this produced similar results.



	Frag MW	Fragment
Proteinase K pH 3.3		
T= 30 mins	8484.0	17-89(1.6)
	8491.0	6-80(0.4)
	11124.0	1-98(0.5)
T= 1 hr	8429.0	27-98(1.4), 18-90(0.4)
	9464.0	18-98(1.3)
T= 4 hrs		Complete Digestion? No Large Fragments Found
T= 24 hrs		Complete Digestion? No Large Fragments Found
Proteinase K pH 4.0		
T= 30 mins	8428.0	27-98(0.4), 18-90(0.6)
	9463.0	18-98(0.3)
	10862.0	3-98(0.9)
	11124.0	1-98(0.5)
T= 1 hr	8429.0	27-98(1.4), 18-90(0.4)
	9464.0	18-98(1.3)
	11123.0	1-98(0.5)
T= 4 hrs	6488.0	27-82(1.5), 25-80(0.6)
	7523.0	26-90(1.4), 27-91(1.4), 18-82(1.4)
	8429.0	27-98(1.4), 18-90(0.4)
	9464.0	18-98(1.3)
	11124.0	1-98(0.5)
T= 24 hrs	6274.0	27-80(1.8), 5-61(0.1), 15-68(1.1)
	6488.0	27-82(1.5), 25-80(0.6)
	6575.0	27-83(1.4), 28-84(0.5), 4-63(0.3), 9-66(0.4)
	8429.0	27-98(1.4), 18-90(0.4)
Proteinase K pH4.7 - Previous Digest		
T= 30 mins	8427	27-98 (0.6) (NT Sequencing determined)
	9462	18-98 (0.7)
	9995	13-98 (1.8)
	10161	11-98 (0.4)
	10860	3-98 (1.1)
	10991	2-98 (1.3)
T= 1 hr	8220	19-89 (0.3), 11-82 (0.3)
	8427	27-98 (0.6)
	9463	18-98 (0.3)
	10161	11-98 (0.4)
	10646	6-98 (0.1)
	10861	3-98 (0.1)
T= 4 hrs	6488	25-80(0.2), 27-82 (1.1)
	7523	26-90(1.5), 27-91(1.5), 18-82 (1.5)
	7609	18-83(0.4), 25-90(0.5)
	8429	27-98(1.4)
	9464	18-98(1.4)
	10863	3-98(1.6)
	10994	2-98(1.2)
	11125	1-98 (0.0)
T= 24 hrs	6274.0	27-80(1.8), 5-61(0.1), 15-68(1.1)
	6488.0	27-82(1.5), 25-80(0.6)
	6575.0	27-83(1.4), 28-84(0.5), 4-63(0.3), 9-66(0.4)
	8429.0	27-98(1.4), 18-90(0.4)
Proteinase K pH 6.0		No Peaks found, MS repeated with identical results

Figure 3.3.8 Digestion Map of Varied pH Generated Cystatin B Amyloid Fibre Proteinase K at 1:1000 ratio protease: WT cystatin B fibre digestions were used. Dark blue bars show defined protein digest fragments or regions, whilst the light blue bars indicate ambiguous fragment regions. The far left column indicates protease and ratio used, plus the digestion time. The column second on the left has the mass value obtained from MS of the detected fragment. The third column from the left is the fragment/s identified from the mass value with the bracketed values indicating the mass deviation between observed and predicted fragment/s. At the top of the diagram, the cystatin B native secondary structure elements of monomer and domain swapped dimer are shown along with the amino acid sequence. The thick black lines indicate predicted specific protease cleavage sites. Digestion profiles from fibres grown at pH 3.3, 4.0 and 4.7 are the same, although the rate of proteolysis appears to be faster in the pH 3.3 fibres compared to the others tested.

3.3.2 Cystatin B Fibre Diameter

In order to produce models of cystatin B amyloid fibre consistent with the new information from the limited proteolysis analysis, clear fibre dimensions were required. Cystatin B fibres are extensively laterally associated in the *in vitro* conditions tested (see figure 3.3.9 (panel A)), with the majority of fibrous material present in dense, 'tangles'. The fibres measured in this study were the lowest width single fibrous structure that could be unambiguously defined. Fibres which appeared to be laterally associated were omitted for several reasons; (i) in most cases, accurate determination of the total number of fibres present in one area was not possible, thus producing poor fibre width estimates, (ii) individual fibres are randomly associated and separated along the fibre length and can associate with several different fibres in loose bundles/tangles, (iii) association is not an essential maturity step in cystatin B fibre formation, as shown by very mature (> 3 week − >1 year, 37°C preparations) producing substantial populations of associated and non-associated fibres. Therefore fibre association is unlikely to be structurally significant in terms of protofibril/mature fibre assembly.

The result was that fibre width was measured at an average of 8.6 nm (± 1.4 nm SD, n = 237) with a minimum and maximum of 5.0 and 13.4 nm (see figure 3.3.9). No single preparation clearly deviated in width or morphology. Previously reported cystatin B fibre widths of fibre grown in similar conditions from 'many different preparations' is 7 ± 1.5 nm and 14 nm by TEM (Zerovnik *et al.*, 2002a), whilst AFM with fibres generated in similar conditions (45 µM cystatin B, TFE free, produced over 3 months) are 13 ± 3 nm in width and 3.4 ± 0.3 nm in height (n = many fibres) with a suggested β-helical twist periodicity of 27 ± 2 nm (Zerovnik *et al.*, 2002a). The results generated in

this study have clarified the fibre width by accurate measurements in well-defined conditions and determination of the minimum observable fibre size, without the addition of laterally associated fibres described as different structural entities. I propose the fibres measured at 13 and 14 nm are actually two fibres laterally associated which may not be apparent in lower resolution EM micrographs and AFM images.

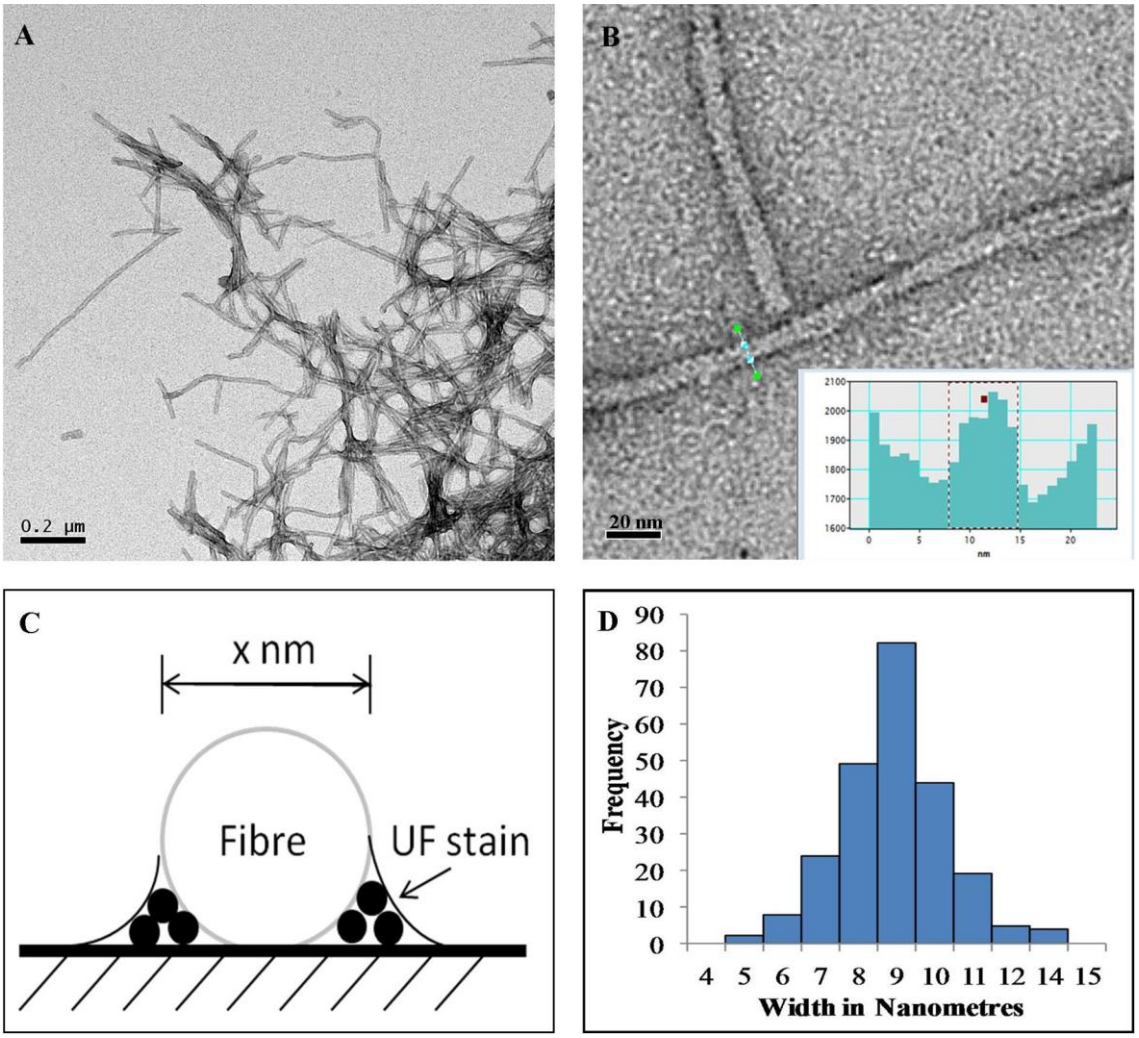


Figure 3.3.9 WT Cystatin B Fibre Width Determination

(A) Typical cystatin B fibre image showing the edge of a 'tangle', laterally associated and individual fibres. **(B)** Measurement of cystatin B fibre width using the electron intensity histogram profiling function of Digital Micrograph 3 software. **(C)** Measurement of fibre width uses framing of the fibre with the heavier uranyl formate stained regions limiting the boundaries. (D) Histogram of cystatin B fibre width measurements with the peak distribution around 9 nm.

3.3.3 Mass Mapping of Cystatin B Amyloid Fibre

To determine the spacing of repeating units in the amyloid fibre mass mapping was performed to generate χ kDa/nm mass per unit length values (MPL). From the 590 measurements, four major fibre classes were observed (see figure 3.3.10 and table 3.3.3) and Gaussians were then fitted to determine the MPL. The MPL Gaussian peak values were 26.2 ±5.0 kDa/nm, 38.7 ±3.3 kDa/nm, 49.8 ±16 kDa/nm and 97.5 ±22 kDa/nm. Fitting Gaussians also determines peak amplitudes, which can then be used as a crude measure of species occupancy and suggests that the first 3 classes are approximately equally populated. The third class is actually more widely distributed suggesting a higher population; however, it is thought that an additional fibre class could be incorporated and account for this variability. The extra class however, could not be fitted due to the quality of the data. The standard deviation is also calculated in the fit and provides a measure of data accuracy.

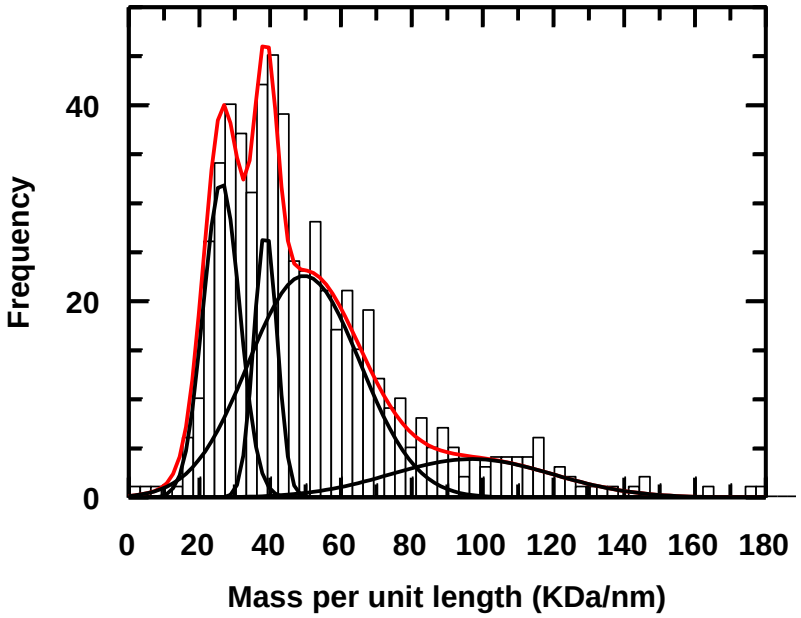


Figure 3.3.10 Mass Per Unit Length Measurements of WT Cystatin B Fibre

Histogram (3 kDa/nm bins) of the STEM MPL measurements, with the data fitted to four Gaussians, shown in black and the combined Gaussian fit is plotted in red. 4-Gaussian fit to: $f = a_1 * e^{\left(-0.5\left(\frac{x-b_1}{c_1}\right)^2\right)} + a_2 * e^{\left(-0.5\left(\frac{x-b_2}{c_2}\right)^2\right)} + a_3 * e^{\left(-0.5\left(\frac{x-b_3}{c_3}\right)^2\right)} + a_4 * e^{\left(-0.5\left(\frac{x-b_4}{c_4}\right)^2\right)}$ where f is the frequency of a mass per unit length measurement (χ), a = peak amplitude, b = peak position, c = standard deviation.

When considering the MPL data, other experiments must be taken into account to explain the results beyond a simple measurement. This includes EM fibre width measurements, (section 3.3.2) average of 8.6 ± 1.4 nm, and an average fibre thickness from AFM of 3.4 ± 0.3 nm (Zerovnik *et al.*, 2002a). These measurements define the two height and width dimensions whilst the physical size of the protein monomers making up the fibre determines the length of the 'box' that x number of protein molecules must fit into. MPL measurements then determine the x number of molecules value.

The H/D exchange data collected on cystatin B fibres reliably predicts the existence of secondary structure present with these protected regions highlighted red in figure 3.3.3. The fibre secondary structure includes the extended β-strand present in the dimer, which is composed of strand 2 and 3 of the monomeric structure. The native β-strand 4, β-strand 5 and a 5 residue element at the C-terminus of the protein is also structured in the amyloid fibre. Considering the double length of the extended β-strand (20 residues), it is useful to think of the monomer occupying the space of 4 strands (in effect, the original strands 2-5) with the potential of a short, fifth strand (the C-terminus). Building a model requires arranging these secondary structural elements, which we believe to be β-strands (based on FTIR measurements, Morgan 2006) in a way consistent with the EM dimensions and available X-ray fibre diffraction data which suggests a cross-β arrangement (meridional and equatorial reflections of 4.7Å, and 10 Å, which are indicative of intra-β-sheet and inter-β-sheet interactions respectively (Jenko et al., 2004)).

The collected data therefore suggests either 4 or 5 β-strands per cystatin B molecule are incorporated into the fibre (native β-sheet 2-5 or 2-5 and the C-terminal). For simplicity, we start with the model proposed in Morgan et al. (2008) where 4 β-strands were arranged down the same β-sheet such that each dimer gave a rise of 4.7 Å multiplied by 4 which equals 18.8 Å (1.88 nm) (see figure 3.3.11). It is also useful to consider whether five β-strands for each molecule is accommodated, giving a rise of 23.5 Å (5 x 0.47 = 2.35 nm). Table 3.3.3 shows the breakdown of the fibre classes measured in the MPL experiment and the number of cystatin B molecules per 4 and 5 β-strand rises. This analysis shows that a 4-β-strand rise has approximately 4, 6, 8 and 16 cystatin B molecules whilst the 5-β-strand rise has approximately 6, 8, 10 and 20 cystatin B molecules. Both β-strand rise values repeat approximately every 2 molecules, except for class IV which is most likely a doublet of class III.

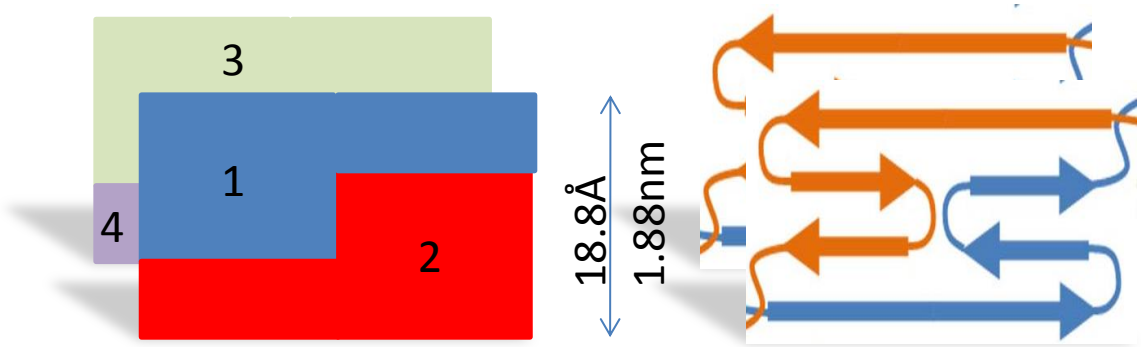


Figure 3.3.11 Cystatin B Fibre 'Sub-Unit'.

β-sheet length based on meridional reflections of 4.7Å from X-ray fibre diffraction data (Jenko et al., 2004).

Fibre Class	MPL (KDa/nm)	Amplitude (% total)	Cystatin B number per 4 β-strand rise	Cystatin number B per 5 β-strand rise
I	26.2 ± 5	37 %	4.4 ± 0.85	5.5 ± 1
II	38.7 ± 3.3	32 %	6.5 ± 0.55	8.2 ± 0.7
III	49.8 ± 16	26 %	8.5 ± 2.7	10.5 ± 6.3
IV	97.5 ± 22	4 %	16.5 ± 11	20.5 ± 4.6

Table 3.3.3 Values of WT Cystatin B Fibre Mass Per Unit Length

MPL ; Mass per unit length expressed in KDa/nm. The amplitude shown in percentage can be used as a crude measure to suggest that fibre classes I-III are approximately equally populated. The cystatin B number per 4 β-strand rise is calculated by determining the mass (KDa) for each 4 β-stranded rise in the fibre, where each β-strand occupies 0.47 nm and 4 strands thus give a rise of 1.88nm, e.g. 1.88 x 26.2. Following this, the number of cystatin B molecules in this mass are calculated by dividing the mass value by the mass of monomeric cystatin B (11 124 Da). The same calculation is repeated for a 5 β-stranded (2.35 nm) rise. The calculations assume a model where the core of the amyloid fibres assumes a continuous β-sheet rather than a discontinuous model with breaks in the core β-structure, which is currently understood to be the most prevalent structure (Petkova *et al.*, 2005, Wasmer *et al.*, 2008).

3.3.4 Sequence Dependent Amyloid Prediction

A comparison of the 'amyloid hotspot' prediction programs was performed on cystatin B to elucidate the expected core regions of the fibre and the results are shown in figure 3.3.12. The programs performed well with typically conservative results similar to the H/D exchange data. In several cases the prediction programmes favourably matched the results from limited proteolysis in that they did not include β-strand 5 and the N-terminal region. This poor matching could indicate that β-strand 5 is a relatively weak amyloid β-sheet forming section of the protein and thus is more susceptible to proteolysis and a higher dynamic association/dissociation with the other fibre interfaces.

Description		10	20	30	40	50
H/D Ex NMR	MMSGAPSATQPATAETQHIADQVRSQLEEKENKKFPVFKAVSFKSQVVAG					
Sec. Structure		ββββββββββ α α α α α α αα α α α α α αα α α α			ββββββββββββββββββββ	
Chem Nature	MMSGAPSATQPATAETQHIADQVRSQLEEKENKKFPVFKAVSFKSQVVAG					

AGG	MMSGAPSATQPATAETQHIADQVRSQLEEKENKKFPVFKAVSFKSQVVAG
WALTZ	MMSGAPSATQPATAETQHIADQVRSQLEEKENKKFPVFKAVSFKSQVVAG
TANGO	MMSGAPSATQPATAETQHIADQVRSQLEEKENKKFPVFKAVSFKSQVVAG
AmylPred	MMSGAPSATQPATAETQHIADQVRSQLEEKENKKFPVFKAVSFKSQVVAG
Packing	MMSGAPSATQPATAETQHIADQVRSQLEEKENKKFPVFKAVSFKSQVVAG
Switches	MMSGAPSATQPATAETQHIADQVRSQLEEKENKKFPVFKAVSFKSQVVAG
Pattern	MMSGAPSATQPATAETQHIADQVRSQLEEKENKKFPVFKAVSFKSQVVAG
Con Energy	MMSGAPSATQPATAETQHIADQVRSQLEEKENKKFPVFKAVSFKSQVVAG
Zagg (pH7.0)	MMSGAPSATQPATAETQHIADQVRSQLEEKENKKFPVFKAVSFKSQVVAG
Zagg (pH4.7)	MMSGAPSATQPATAETQHIADQVRSQLEEKENKKFPVFKAVSFKSQVVAG

	60	70	80	90	98
H/D Ex NMR	TNYFIKVHVGDEDFVHLRVFQSLPHENKPLTLSNYQTNKAKHDELTYF				
Sec. Structure	ββββββββββββββ	ββββββββββββββ		ββββββββββββββ	
Chem Nature	TNYFIKVHVGDEDFVHLRVFQSLPHENKPLTLSNYQTNKAKHDELTYF				

AGG	TNYFIKVHVGDEDFVHLRVFQSLPHENKPLTLSNYQTNKAKHDELTYF
WALTZ	TNYFIKVHVGDEDFVHLRVFQSLPHENKPLTLSNYQTNKAKHDELTYF
TANGO	TNYFIKVHVGDEDFVHLRVFQSLPHENKPLTLSNYQTNKAKHDELTYF
AmylPred	TNYFIKVHVGDEDFVHLRVFQSLPHENKPLTLSNYQTNKAKHDELTYF
Packing	TNYFIKVHVGDEDFVHLRVFQSLPHENKPLTLSNYQTNKAKHDELTYF
Switches	TNYFIKVHVGDEDFVHLRVFQSLPHENKPLTLSNYQTNKAKHDELTYF
Pattern	TNYFIKVHVGDEDFVHLRVFQSLPHENKPLTLSNYQTNKAKHDELTYF
Con Energy	TNYFIKVHVGDEDFVHLRVFQSLPHENKPLTLSNYQTNKAKHDELTYF
Zagg (pH7.0)	TNYFIKVHVGDEDFVHLRVFQSLPHENKPLTLSNYQTNKAKHDELTYF
Zagg (pH4.7)	TNYFIKVHVGDEDFVHLRVFQSLPHENKPLTLSNYQTNKAKHDELTYF

Figure 3.3.12 WT E31 Cystatin B Amyloidosis Sequence Prediction

H/D Ex. NMR = Hydrogen/ deuterium exchange protected regions (red) in the amyloid fibre (Morgan *et al.*, 2008). Sec. Structure = secondary structure in the dimer, α = α-helix, β = β-sheet. Chem. nature = chemical nature of the amino acid: Blue = basic, Pink = acidic, Polar = green, Non-polar = yellow. AGG = AGGRESCAN (Conchillo-Sole *et al.*, 2007), WALTZ (Maurer-Stroh *et al.*, 2010), AmylPred (Hamodrakas *et al.*, 2007), Packing = Average Packing density (Galzitskaya *et al.*, 2006), Switches: = Conformational switches (Hamodrakas *et al.*, 2007), Pattern (Lopez de la Paz and Serrano, 2004), Con Energy: Conformational Energy (Zhang *et al.*, 2007), Zagg = Zyggregator (Tartaglia *et al.*, 2008).

3.4 Discussion

3.4.1 Limited Proteolysis Data Implication on Cystatin B Amyloid Formation

This study has generated additional evidence for the absence of a native-like α-helix in fibre associated protein, due to rapid hydrolysis throughout this region, and further supports the absence of H/D exchange protection previously observed in this region (Morgan *et al.*, 2008). This indicates that unfolding of the α-helix is essential in cystatin B fibre formation. The α-helix role is essential to cystatin B stability, indicated by dictation of the folding rate in chimeras with α-helices from different cystatins, despite the helix representing only 25 % of the protein chain (Jelinska *et al.*, 2011).

In identifying an amyloid fibre core, it must be considered that proteases typically require ≥10 amino acid residues with high chain mobility for binding upstream of the cleavage site, thus limiting the exact definition of cystatin B amyloid fibre core regions (Hubbard, 1998). Therefore, the protease resistant 27-80 amino acid fragment represents a maximal fibre core. Removing a further 10 residues from each termini of this fragment results in an NT, one amino acid from the beginning of the native β-strand 2 and a CT, one residue within β-strand 4. Therefore it is proposed the cystatin B fibre core is minimally composed of native-like β-strands, 2, 3 and 4, residues 38-70, and maximally no more than residues 27-80. Therefore, it is likely only one third of the cystatin B protein is incorporated into the fibre core. A more conservative estimate would be 54% incorporation, assuming the 27-80 residues fragment is a perfect delineation of the fibre core.

This result clearly indicates the currently posited simple structural model of native dimer-like stacked β-sheet (strand 2-3, 4 and 5) (Morgan *et al.*, 2008) is unlikely due to hydrolysis within native β-sheet 5, whilst the amyloid fibre remains intact (figure 3.3.5). If the protease cleaved β-sheet 5 whilst incorporated into the fibre core, then the core structure is likely to collapse and either induce fibre break down or structural rearrangement. It is improbable that digestion occurs within the fibre core structure with subsequent fibre structural remodelling, and there was no evidence of morphology changes or full-size soluble protein species which would be indicative of such events.

The implication is that strand 5 must form part of a separate sheet, which can be cleaved away independently of the main fibre structure. A number of models allow this including the widely proposed parallel in-register models of amyloid.

3.4.2 Mass Mapping Restricts the Model Further

The width of a single fibre 'tape' (8.6 ± 1. nm) is equivalent to the length of the dimer-like extended, straightened β-strand made from the original strands 2 and 3 which is suggested to be preserved in the fibres from the H/D exchange data. The β-sheets are most likely to be arranged perpendicular to the fibre axis as suggested from the persistence of the 4.7 Å reflection in X-ray diffraction data of a partially aligned sample of cystatin B (Jenko et al., 2004). The 3.4 ± 0.3 nm thickness of the fibre can thus only accommodate an absolute maximum of 3 β-sheets longitudinally, with each sheet contributing 1 nm (10 Å). Given the dimensions and the requirement to encompass the presence of large amounts (perhaps 50% by volume, a third by polypeptide length) of unfolded protein, it seems likely that the originally proposed 2 β-sheet thickness in Morgan et al. (2008) is consistent with the EM dimensions. This resulted in the presence of 4 cystatin B molecules per 4 β-strand rise (Class I), a doublet thereof explaining the presence of fibres with 8 and 16 molecules of cystatin B per 4 β-strand rise (Class III and IV). Class II fibres contain only 6 molecules per 4-strand rise, something which is more difficult to envisage within the reported dimensions but is reminiscent of proposed models for insulin, see figure 3.4.1 (Jimenez et al., 1999).

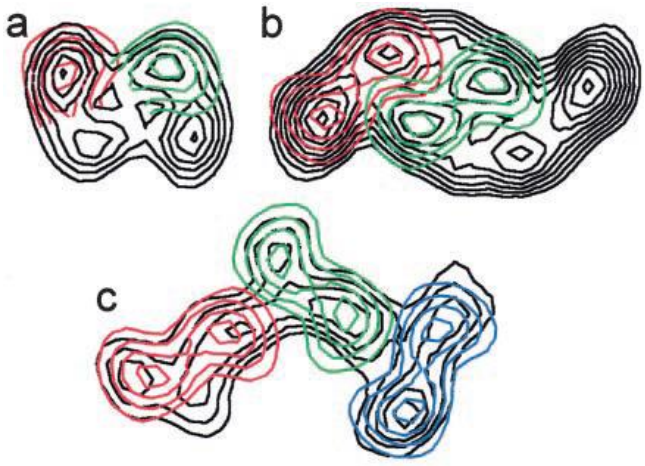


Figure 3.4.1 Cross Sections of Various Insulin Fibre Morphologies

Superposition of cross sections from different insulin fibre morphologies suggesting the presence of a common protofilament structure; shown in green, red and blue. Figure from (Jimenez et al., 1999).

The MPL measurements suggest that the 5-β-sheet rise is unlikely. A model of cystatin B fibres with unpaired numbers of molecules (5, 8 and 10) is more difficult to accommodate. Further to this, with β-strands 2 and 3 forming a single β-sheet, β-strands 4 and 5 are most likely required in the structure to form an opposing β-sheet. In this case, the H/D exchange protected C-terminal is most likely to either be included in this balancing β-sheet or form a small independent structure extraneous to the main fibre core.

Although the MPL measurements suggest that at least four fibre classes exist, this is not observed in the width measurements obtained, with the measurement distribution suggesting a single species. This discrepancy could be due to several issues, such as the negatively stained EM images failing to determine fibre thickness which could multiply, assuming the fibre lays flat and has no periodic twist. Equally the sampling method of the width measurements could be a reason for this, with sampling only of single fibres that could be unambiguously determined to be not laterally associating at the point of measurement. STEM MPL images are intrinsically low resolution and as such it is difficult to determine and observe the smallest fibres and determine whether the fibres are laterally associated. Therefore the minimum fibre structure could actually be composed of 2 cystatin B molecules per rise and were not measured due to the poor resolution which would explain the repeating unit of two cystatin B monomers per fibre class. This is unlikely however as the technique resolution is typically sufficient to select the smaller fibres and there is a near absence of MPL measurements at the expected values if this were the case. The most likely structural scenario is that the minimal fibre unit is composed of 2 protofibrils each with 2 cystatin B fibres per 4-β-strand rise similar to that observed in $A\beta_{1\text{-}40}$ (Petkova et al., 2002) and amyloid-like crystals (Serpell, 2000). Cystatin B can then also produce fibre composed of 3, 4 and 8 protofibrils and or/the class II fibres represent a different structural arrangement and class III and IV are laterally associated class I fibres.

Random errors from this experiment are due to noise from scattered electron signal from the fibre mass and the support carbon film. Error from electron scattering can be reduced using higher electron exposure, although this also increases the error from electron-beam-induced mass loss. However, with collagen fibre as an example 1-2 % accuracy can be achieved (Holmes *et al.*, 2001). Systematic error within the technique includes mass contamination, calibration errors and electron irradiation-induced mass loss. Mass contamination of the sample and/or the reference fibre with small molecules such as salts can lead to error with TMV mass contamination showing a typical 2-5% error (Muller *et al.*, 1992). Electron irradiation mass loss is reduced by using STEM, minimising the number of scans in a particular area and measuring the mass loss occurring in the TMV calibration microscopy over time. Thus the results produced are likely to be reliable with interpretation and data fitting being the most likely source of error.

3.4.3 Non-Native Structure Model of Cystatin B Amyloid Fibre

We propose a new simple working model (Figure 3.4.2) based on the limited proteolysis and H/D exchange data which posits that the native β-strand 2 and 3 is extended into a single strand stacking in parallel to form the fibre 'backbone' whilst β-strand 4 and 5 forms the second arm of a β-arc (hairpin). This model allows a number of things. Firstly, the conservation of the 20 amino acid continuous β-strand observed in the dimer (made from β-strands 2 and 3), secondly shielding of the hydrophobic regions of the protein is ensured, while still allowing the cleavage of β-strand 5, leaving a shortened second strand yet still maintaining a continuing β-sheet core of strand 2-3 and 4.

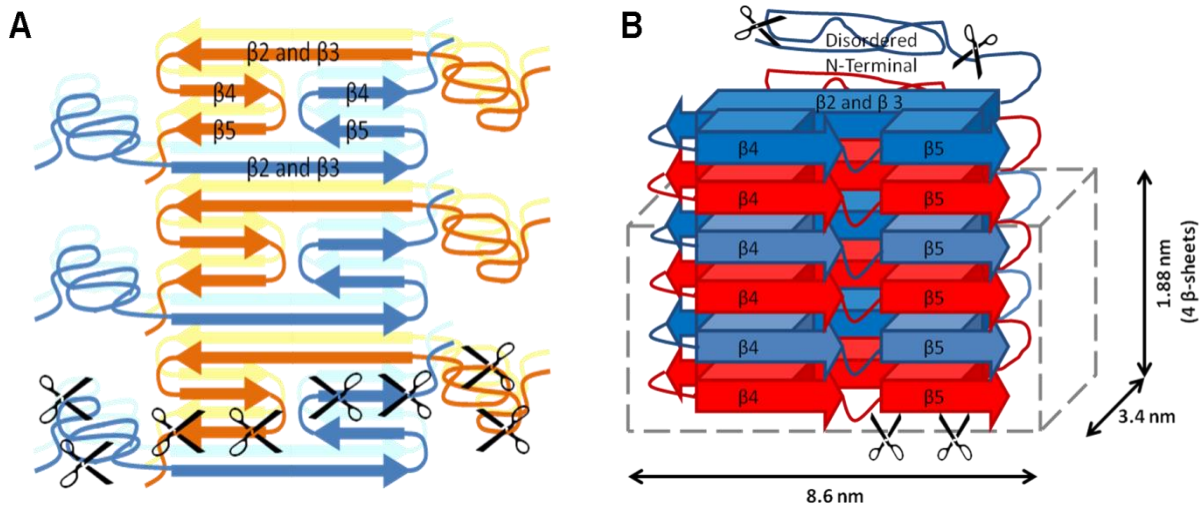
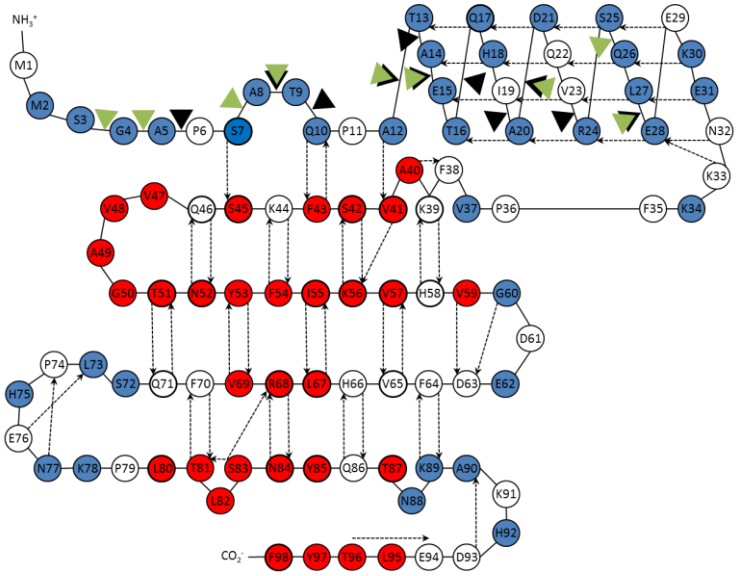


Figure 3.4.2 Modelling of Cystatin B Amyloid Fibre

A. The current model based on H/D exchange NMR maintains mostly native β-sheet structure with a disordered native α-helix (Morgan *et al.*, 2008). This model is not supported by the limited proteolysis data which requires β-strand 5 to be accessible to protease without fibre dissociation or large scale structural remodelling. **B.** Suggested working model, which allows access to β-strand 5 without major remodelling upon removal. Both models fit the EM and AFM size restrictions and account for MPL measurements (classes I, II and IV).

Figure 3.4.3 shows predicted patterns of hydrolysis for the two models proposed. In both models proteolysis in the NT region is accommodated as observed. However both models do not simply fit the limited proteolysis results when considering the cleavage within β-strand 5. The H/D exchange based model cannot be used to explain the proteolysis data because β-strand 5 is completely inaccessible and if it were cleaved the fibre would fall apart. However, the newly proposed β-strand arc model can accommodate this data if β-strand 5 is occasionally accessible by a melting and re-annealing process.

H/D Exchange Model



B-Arc (Hairpin) Model

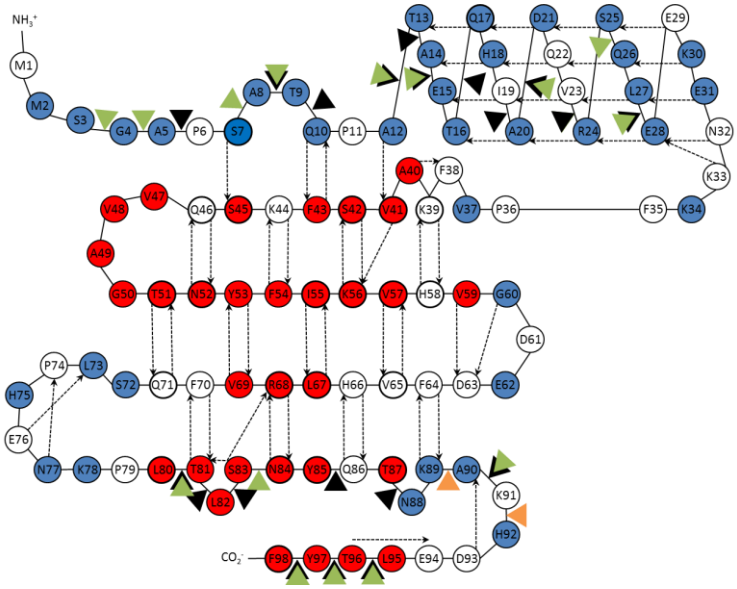


Figure 3.4.3 Predicted Limited Proteolysis with Fibre Structure Models

The predicted cleavage positions, shown by triangles, are based on the posited structural models and incorporates the observation that hydrolysis rarely occurs within 10 residues of structured regions. Proteinase K proteolysis is indicated by black triangles**,** elastase by green triangles and endoproteinase Lys-C digestion labelled with orange triangles. The H/D exchange model fits only the cleavage from the N-terminus. The βstrand arc (hairpin) model shown fits the proteolysis data robustly if the C-terminal region is occasionally accessible by a melting and re-annealing process. The topology map is modified from (Morgan *et al.*, 2008), where hydrogen bonding present in the dimer is shown using dashed arrows. The regions of the molecule involved in the formation of secondary structure in the fibre are shown as H/D exchange protected residues (red) whereas unstructured regions of the amyloid fibre which readily exchange hydrogen for deuterium are highlighted in blue. Residues for which no H/D exchange data is available are coloured white.

The accessibility of β-strand 5 could be mediated by dynamic fibre sub-unit melting and annealing leading to increased exposure and perhaps a lower frequency of secondary structure maintenance. This process may explain the difference in the rate of proteolysis in the N and C terminal region. In further support of this, it was noted that in the H/D exchange data at pH 9.0 the last 19 residues (including β-strand 5) are 10-fold less protected than those in the native β-strand 2, 3 and 4. This was the only condition examined in which differential stability of fibrous cystatin B could be observed (due to a significant contribution from EX2 as well as EX1 exchange). At pH 4.7 and 7.0 the protection amplitude is near identical for all protected residues (Morgan, 2006) and at pH 8.0 (the conditions for proteolysis) it is assumed that the exchange rates would appear somewhere between the pH 7.0 and 9.0 data. Alternatively the native loop (8 aa) between β-strand 4 and 5 may allow protease binding, leading to destabilisation of β-strand 5 and subsequent hydrolysis, perhaps in a processive manner.

The cystatin B amyloid fibre core consists maximally of 54% of the total monomeric residues, which is consistent with results from many other amyloid forming proteins. For instance; HET-s fibres produce a 218-289 (25%) proteinase K resistant fragment (Balguerie *et al.*, 2003a), α-synuclein has residues 31-109 (56%, 140 aa) protected from proteinase K proteolysis (Miake *et al.*, 2002), human lysozyme residues 32-108 (58%) are protected from pepsin digestion (Frare *et al.*, 2006), Ure2p filaments with addition of proteinase K were trimmed from 25 nm to 4 nm cores which composed of residues 1-70 (20%) (Baxa *et al.*, 2003). In comparison; β_2-microglobulin long straight fibres produced at low ionic strength were digested with pepsin to produce fragments of 9-99 aa (90%) (Myers *et al.*, 2006) and P13-SH3 fibres were completely resistant to proteinase-K proteolysis (de Laureto *et al.*, 2003). It is thus supportive that the data values obtained in this study is within the range observed in other systems.

3.4.4 Sequence Specific Amyloid Prediction

A comparison of the 'amyloid hotspot' prediction programs was performed on cystatin B to elucidate the expected core regions of the fibre. All the programs tested predict several 'hotspots' of aggregation within the β-sheet regions of cystatin B. Of particular note is the GTNYFIKV motif in β-strand 3 which nearly all the programs detect and is empirically thought to be the primary amyloidogenic region in the centre of the protein. It is difficult to decide which program is the best, although I personally prefer ZYGGREGATOR, closely followed by WALTZ and AGGRESCAN, the latter two

being the most user friendly. The programs overall performed well with typically conservative results similar to the H/D exchange data. In several cases the prediction programmes favourably matched the results from limited proteolysis in that they did not include β-strand 5 and the N-terminal region. This poor matching could indicate that β-strand 5 is a relatively weak amyloid β-sheet forming section of the protein and thus is more susceptible to proteolysis and a higher dynamic association/dissociation with the other fibre interfaces.

3.4.5 The Role of Domain Swapping in Cystatin B Amyloid Fibre

The data presented herein develops the structural model of cystatin B fibre with implications on the understanding of amyloid formation in a multitude of natively folded proteins. The primary relevance of simple 3D-domain swapping models to the mechanism of amyloid formation by cystatin B seems remote although the key element identified in the 3D domain-swapped dimer is preserved in the core of the fibres. The topology of the β-strands proposed here is more reminiscent of the parallel-in-register structures proposed to account for large numbers of amyloids prepared *in vitro* (Margittai and Langen, 2008). Further fibre structural information collected from a range of cysteine mutants analysed by X-ray absorption near-edge spectroscopy and solid-state NMR projects which are ongoing will help refine the model and early indications suggest that the native-like model is unlikely. It is of particular intrigue that amyloid fibres can accommodate large segments of exposed unstructured regions extraneous to the fibre core and accessible to proteases. Whilst many amyloid related diseases may have similar mechanisms of toxicity, yet to be conclusively determined, it is perhaps these unstructured regions which mediate variations in targeting and aetiology of disease. The elucidation of amyloid structure and intermediates will assist in drug development and treatment for many degenerative diseases.

Chapter 4: Oligomer Production

4.1 Introduction

The Zerovnik group report the production of cystatin B oligomers composed of 5,6,7,8 and up to 32 monomers (personal communication), (Ceru and Zerovnik, 2008). These high molecular weight oligomers (HOs), often referred to in this chapter as oligomers, are suggested to be toxic and therefore currently of interest in developing treatments for neurodegenerative diseases (Lashuel *et al.*, 2002). Oligomeric species are typically transient in nature, thus difficult to isolate with previous cystatin B attempts proving unsuccessful. Replication of the Zerovnik group oligomer generation has been attempted by both repeating and modifying the original protocol, but failed to yield HOs. Site-directed mutagenesis of WT cystatin B to produce the G4R EPM1 related mutant and a variation of the standard protocol described in section 2.5 produced HOs. The generated oligomer was then characterised by EM, solvent stability and limited proteolysis.

The G4R mutation as described in section 1.9 is one of the single point mutations implicated in myoclonus progressive epilepsy of type 1 (EPM1). This mutation from the compact polar and highly conserved glycine to the long basic side chain of arginine is likely to have major implications on the stability of the tentative β-strand 1, which nearly forms a type II loop (see figure 4.1.1) (Stubbs et al., 1990). This mutated β-strand may mediate association of other cystatin B β-sheets by becoming flexible and dissociate from the protein core, it might also act as a tether through which it may associate with other cystatins to form oligomer and/or destabilise the α-helix and increase the propensity to unfold and form larger structures. G4R has previously been shown to remain in the prefibrillar aggregate state for four times longer than the WT in fibrilisation conditions, hence must induce a significant alteration to protein stability (Rabzelj *et al.*, 2005). Single point mutations destabilising amyloidogenic proteins are not unusual with the arctic mutant Aβ (E22G) inducing faster and higher levels of amyloidogenesis (Nilsberth et al., 2001) and the Δ6-β_2-microglobulin which remains in a prefibrillar state for extended periods of time (Eichner et al., 2011).

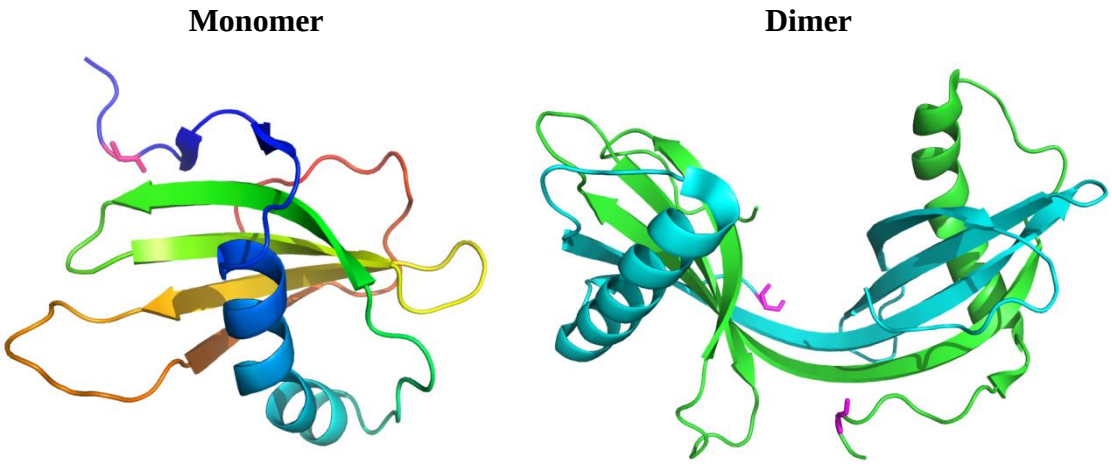


Figure 4.1.1 G4R Mutation Position
The glycine residue position is denoted by a magenta sticks in the monomer and dimer. The near loop in β-strand 1 is also quite clear. PDB: monomer 1STF, dimer 2OCT, graphically rendered in Pymol (DeLano Scientific, CA).

4.2 Materials and Methods

4.2.1 Oligomer Generation from Cystatin B Monomers

The standard cystatin B purification protocol described in section 2.5 was primarily used for most initial protein purification with variations based on a protocol from the Zerovnik group. The Zerovnik group protocol detailed the use of a papain column in the first purification step followed by several rounds of diafiltration then SEC with a preparative sephadex 75 column. The final step was four cycles of freeze-thawing followed by SEC analysis of the oligomeric state. Immediately noted was that only monomeric cystatin B can bind a papain column (Abrahamson, 1993). The monomeric population typically constitutes only ~10% of the total protein content, whilst ~90% is dimer and a few percent tetramer, thus the use of a papain column would dramatically reduce protein yield.

In the first trial a monomer/ dimer/ tetramer mixture of WT E31 cystatin B, purified according to the standard method described in chapter 2.5, was used to replicate the conditions of papain column elution. To replicate the conditions, 9 ml of cystatin B at 1 mg/ml was buffer exchanged into an equal volume with an Amicon ultrafiltration stirred-cell with a Millipore regenerated cellulose 10 000 kDa MWCO filter into 20 mM ethanolamine/100 mM NaCl, pH 10.7. The cystatin B solution was then added drop-

wise into 1 ml 0.02 M $NaPO_4$/1 M NaCl, pH 6.0. The protein solution was then buffer exchanged into 10 mM $NaPO_4$/100 mM NaCl, pH 6.0 buffer. 1 ml aliquots of protein solution were then freeze-thawed for 4 cycles of 1 hour thawing in air at room temperature and a minimum 1 hour at -80°C. Further purification attempts are described below in the results section.

4.2.2 Oligomer Characterisation - SEC HPLC

Samples were taken at each step and the oligomeric state analysed with SEC-HPLC (Shodex KW803 column) for separation of oligomers <700 kDa and SEC with a 100 ml sephacryl-400 column for very high molecular weight oligomers. To remove the possibility that HO was interacting and/or binding to the Shodex KW803 column matrix or stainless steel hardware, a pre-packed analytical superdex 200 column with plastic and glass hardware (Biopharmacia) was also used to analyse the oligomer generation products. Both columns yielded similar chromatograms in all cases.

4.2.3 Oligomer Morphology - TEM

To measure the HO diameters, Digital Micrograph 3 software was used to generate electron intensity histogram profiling to accurately determine and measure the oligomer width. Three diameter measurements were performed for each oligomer, with the average of the three readings being subsequently used for averaging. The three measurements and averaging were performed to account for variance in the approximate disk shape and error caused by edge irregularity and poor contrast of the oligomers. Five different electron microscopy grids produced from the 3 successful preparations were each analysed independently and contributed a maximum 10 values for each oligomer class to ensure no single preparation skewed the results if they proved significantly different. The images used were selected to cover the 3 different preparations, have a high resolution and contrast and extensive oligomer coverage. Images were collected on a Phillips CM100 electron microscope with a 1024 x 1024 pixel Gatan CCD camera using 21 000 x magnification negatively stained electron micrographs, thus each pixel was equal to 1 nm and defines the maximum level of accuracy as ± 1 nm.

4.3 Results

4.3.1 Oligomer Formation

Published Protocols

The replication of the Zerovnik group oligomer production protocol, described above, was attempted with the WT cystatin B (C3S) E31 variant with no success. Freeze-thawing was suggested to be the major step in high molecular weight oligomer (HO) production and proved to increase tetramer formation 5-fold in WT cystatin B (figure 4.3.1) and 3-fold in the V59Q mutation when frozen and defrosted four times, with each cycle carried out with a 1 ml sample over a 1 hour period at room temperature and then at -80°C. Intriguingly, increasing the WT cystatin B protein concentration and freeze-thawing in an identical manner increases the tetramer concentration relative to the dimer population. However, concentration dependence controls without freeze-thawing were not performed. The Q46C-N52C mutant also produces large molecular weight species, but was not followed up, due to potential cysteine cross-linking, resulting in potential amorphous aggregation (see figure 4.3.2) .

Papain Affinity Chromatography

Following the first oligomer production trial, a 100 ml papain column was used to isolate cystatin B monomer from purified solution, in conditions described in the Zerovnik group protocol. The first attempt was unsuccessful with no detectable levels of protein binding the column and subsequently eluting. A further two attempts were made with a higher starting mass of protein and stirring of the papain beads overnight at 4°C in a beaker with suitably buffer exchanged cystatin B and equilibrated papain beads. This stirring method allows monomer to bind to the papain beads, reducing the monomer population in solution which is then replenished dynamically from the dimeric and possibly tetrameric state. Manipulation of the equilibrium in this manner theoretically allows large quantities of monomer to bind to the column and thus result in a high yield. However in this case, the attempts proved unsuccessful and cystatin B did not bind the papain column sufficiently in the conditions described.

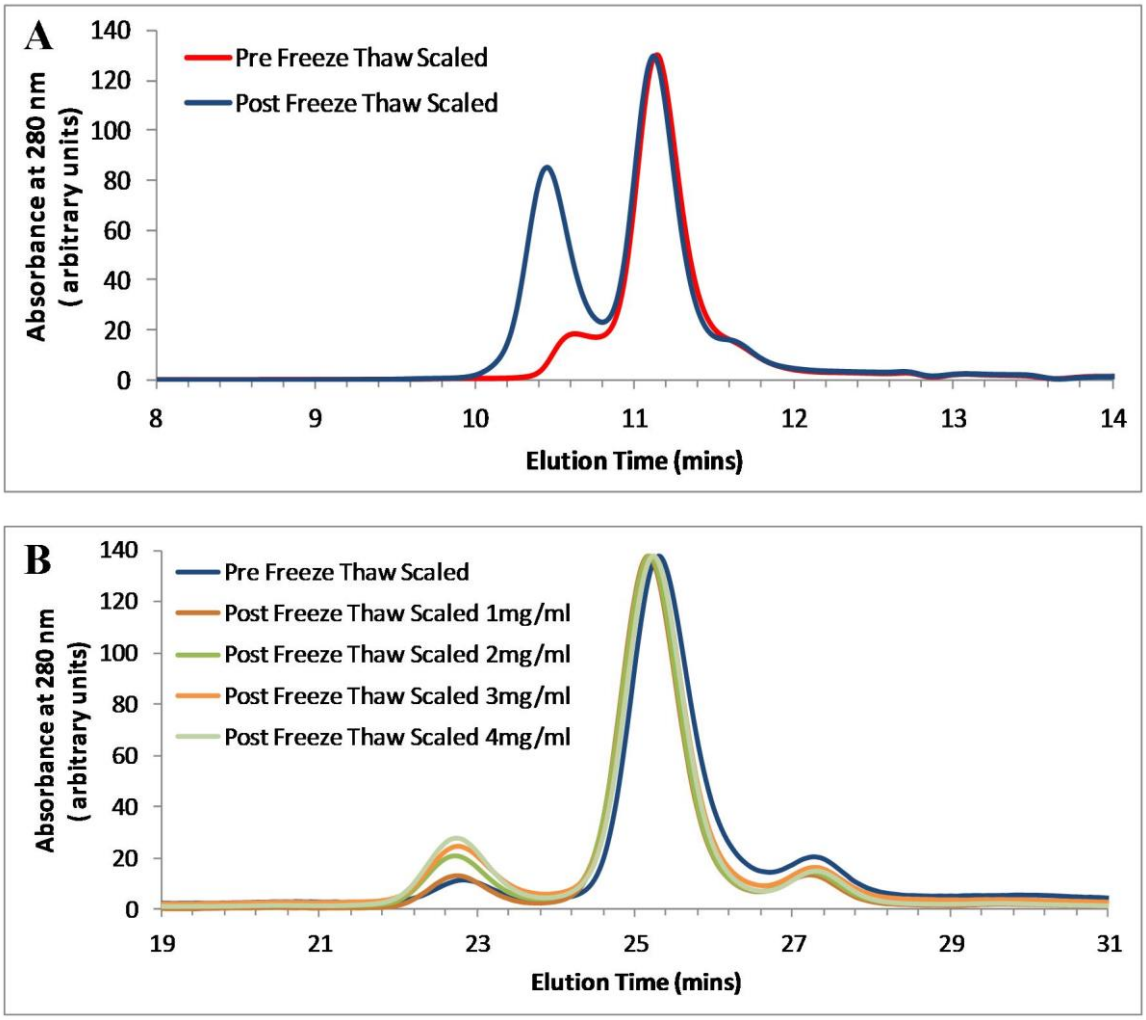


Figure 4.3.1 Freeze-Thawing Increases Tetramer Population in WT Cystatin B

HPLC chromatograms of freeze-thawing have been scaled to the value of the highest dimer peak for comparison. **(A)** Initial freeze-thawing of WT cystatin B ~1 mg/ml showing significant increase of tetramer (Shodex KW803 column) **(B)** Protein concentration dependence of freeze-thawing, with an incremental increase in the ratio of tetramer to dimer as protein concentration increases (superdex 200 column).

Bacterial Expression Strain Trials

Very small peaks observed occasionally in SEC of the standard WT cystatin B preparation were observed and analysed by HPLC-SEC and electron microscopy, although no successful determination of the oligomer state was achieved most likely as a result of low concentration. Following this *E. coli* BL21 (DE3) Gold pLysS cells, the same expression cells utilised by the Zerovnik group, were employed to express WT E31 cystatin B with an induction time of 5 hours in the hope that the different cellular conditions may yield HOs. However, these cells also failed to produce oligomeric species.

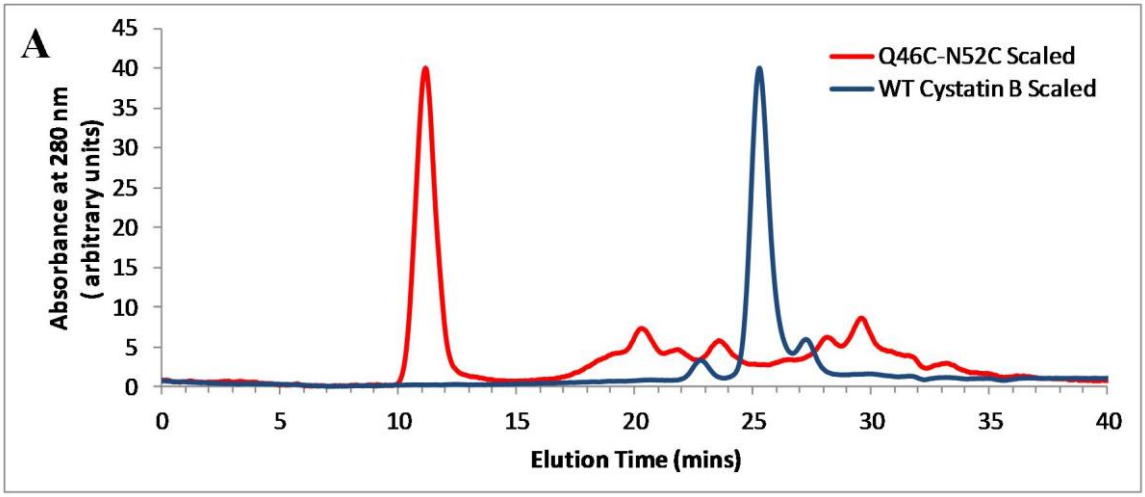


Figure 4.3.2 Q46C-N52C Cystatin B Oligomer

Q46C-N52C cystatin B (red) potentially produces oligomeric or aggregate species which elute on a superdex 200 column near the exclusion volume (11 minutes).

Cystatin B G4R Variant

It was subsequently suggested that the G4R mutant identified in the disease myoclonus epilepsy of type I (EPM1), produces significant amounts of HOs and in fibrilisation conditions remains in the prefibrillar aggregate state for a much greater time span (Rabzelj *et al.*, 2005). The G4R mutant was therefore produced by site-directed mutagenesis into the E31 (C3S) WT background. During initial purification, an error in the anion exchange column purification step, whereby the column was equilibrated with 500 mM NaCl, pH 6.0, 10 mM sodium phosphate gradient elution buffer which prohibited binding of the protein. To rectify, the entire protein eluent was diluted 10-fold in sodium chloride free buffer and loaded onto a correctly equilibrated sodium chloride free column overnight at RT. This treatment produced a low overall protein yield, but allowed clear observation of a small peak of HO species upon elution on a G75 superdex SEC column (see figure 4.3.3). The HO fractions, when examined by EM (see figure 4.3.4), showed an extensive coating of the grid surface with a heterogeneous mixture of oligomeric species.

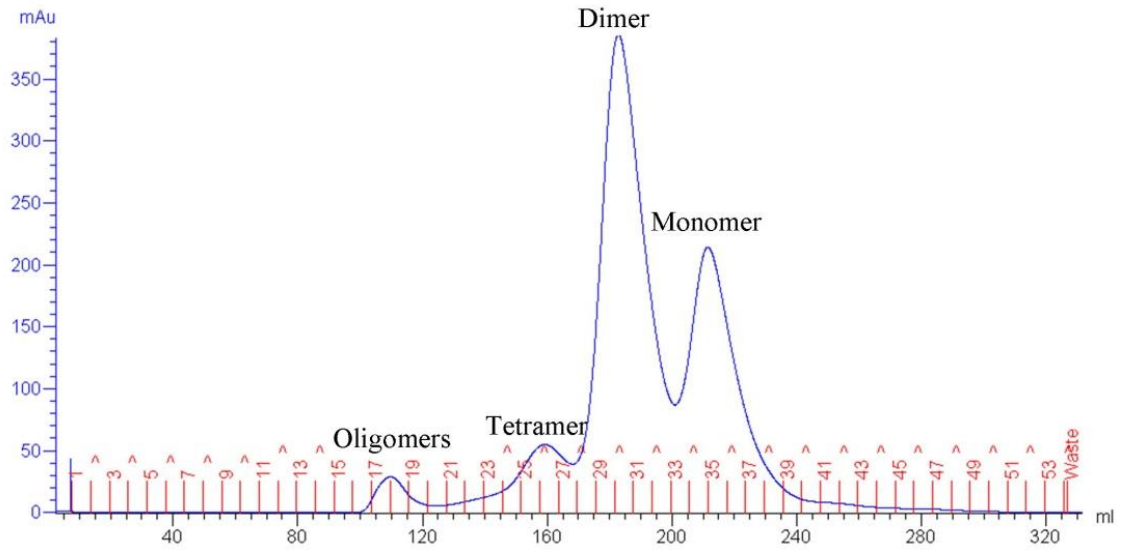


Figure 4.3.3 G4R Oligomer Purification SEC

The oligomeric peak typically eluted into fractions 17-21 (6 ml each), in a 400 ml bed volume superdex G75 column. Some variability occurred in different preparations, such as a shouldering peak between the established oligomer peak and tetramer. In this additional peak, oligomeric species were not observed in EM.

Reproducing the oligomer preparation results was initially not straightforward, with the first two preparations carried out with the standard cystatin B purification methodology (chapter 2.5) failing to yield well defined structures and instead produced amorphous aggregate and/or the oligomer dissociated. A mixture of monomer, dimer and tetramer as well as amorphous aggregate was observed in the same SEC fractions that previously bore distinct oligomer (see figure 4.3.5). The key step proved to be ultrafiltration of the anion exchange protein solution fractions to half the standard concentration (10, not 20 mg/ml) before SEC elution. See figure 4.3.6 for a schematic of the purification, which is nearly identical to the standard method described in section 2.5. Generally, each oligomer preparation has produced low yields (< 1.5 mg/l of LB growth media) and was much lower in M9 minimal media (0.1 mg/l). A range of novel therapeutics such as epigallocatechin-gallate (EGCG) (Ehrnhoefer *et al.*, 2008) and rifamycin SV (Woods et al., 2011) were considered in efforts to enhance oligomer production, but time limitations have forbade continuation of this study.

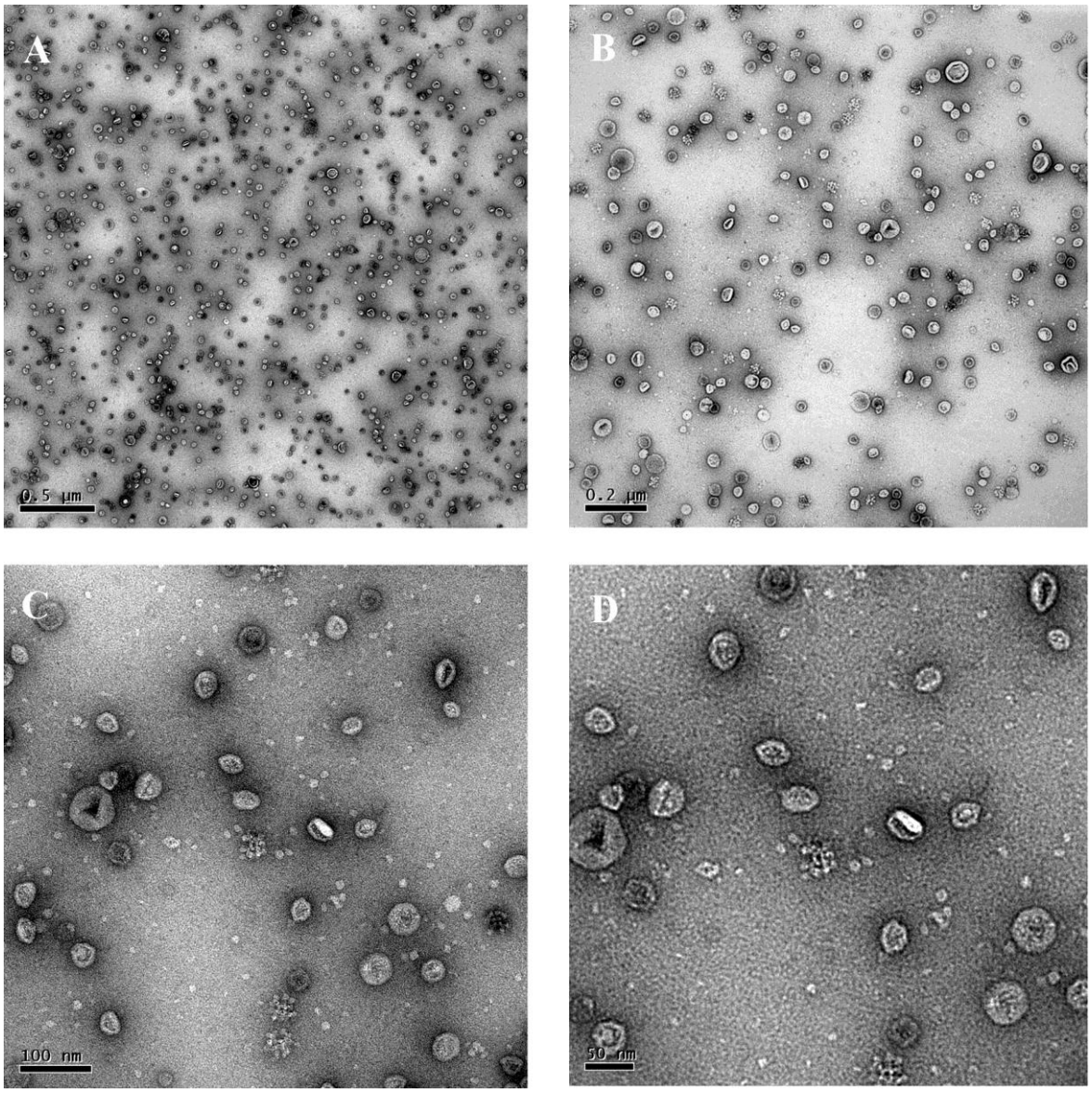


Figure 4.3.4 Initial Successful Preparation of G4R Cystatin B High Molecular Weight Oligomers

Negative Stain Electron Micrographs of purified oligomeric species from the initial purification. The structures were observably coating the entire grid, with a small selection of images shown. (A) 5 200 x magnification (B) 11 500 x (C) 28 500 x (D) 39 000 x. There are several different morphologies present, including the 'cluster' species, see figure 4.3.8 for further analysis of the morphologies.

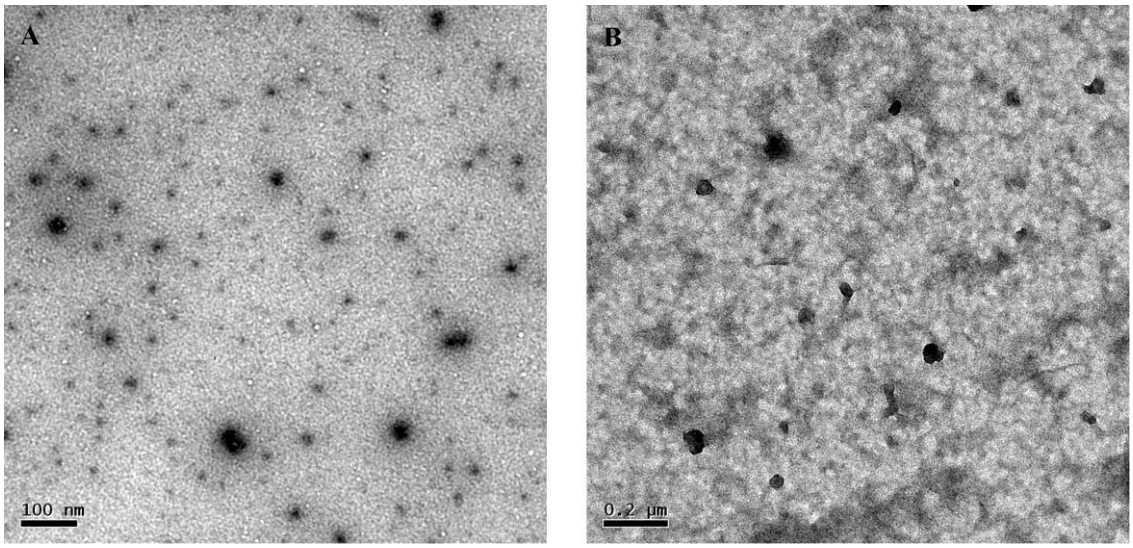


Figure 4.3.5 Cystatin B Oligomer Preparation: High Concentration Aggregation

G4R cystatin B purified by the standard protocol in an attempt to produce a high yield of oligomeric species. The SEC fractions previously containing distinct oligomeric species exhibited an extensive population of oligomer-like aggregate which had a poorly defined structure and could not be adequately focussed upon at higher magnifications in an analogous manner to amorphous aggregate species. **(A)** 21 000 x magnification electron micrograph of oligomer-like aggregate. The oligomer-like dark regions are negatively stained in a similar manner to amorphous aggregate. **(B)** Precipitate formed by G4R upon ultrafiltration prior to SEC. The background is typical of high concentration soluble protein with some heavily stained oligomer-like structures. Some of the precipitate may have resuspended or the soluble protein could be from protein solution remaining with the pellet.

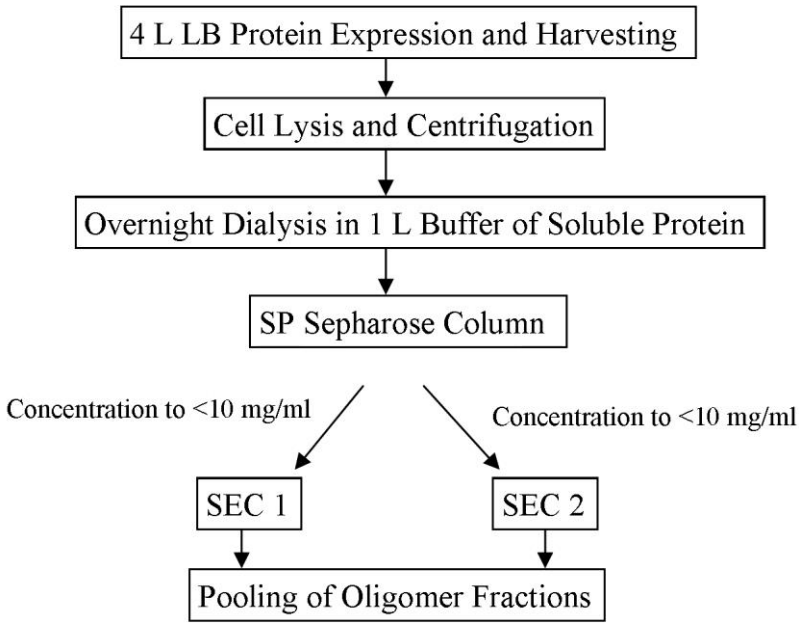


Figure 4.3.6 G4R Oligomer Purification Strategy

The schematic strategy is nearly identical the standard purification method described in section 2.5 with the exception that the protein from the SP sepharose column is divided in two, concentrated to ~ 10 ml (<10 mg/ml) and then passed through the SEC column.

Attempted Production of G4R Higher Molecular Weight Oligomer Species from Purified Monomer, Dimer and Tetramer

Exchanging G4R cystatin B monomer and dimer mixtures (1 mg/ml) into 6 M guanidine hydrochloride and refolding on SEC HPLC in the typical pH 6.0 sodium phosphate buffer (150 mM NaCl, 1 mM NaN_3) did not produce any further oligomeric species. Four cycles of freeze-thawing a 500 µl monomer and dimer mixture at -80°C for 15 minutes and thawing in air at 21°C for 30 minutes also failed to produce oligomeric species. An EM image of the post cell lysis soluble material was obscured by a large amount of amorphously aggregated protein, thus no oligomer species was observed, whether present or absent. High molecular weight oligomer could only be produced directly from *E. coli* protein extractions, with manipulation of the monomer and dimer failing to yield oligomeric species. The reason for this remains unclear, although specific *in vivo* conditions are dissimilar to those *in vitro*.

G4R-Like Oligomers are Observed in WT Preparations Although are Rare

The oligomers generated are similar to observations of rare single oligomers in WT cystatin B fibre preparations. Originally these structures were thought to be artefacts due to the infrequency of observation. HO structures have been observed in a WT cystatin B monomer/ dimer/ tetramer preparation in pH 4.7 sodium acetate buffer (150 mM NaCl, 1 mM NaN_3) which was stored at 4°C for two years and was then studied by EM due to the solution appearing cloudy in a manner analogous to fibre solutions (see figure 4.3.7). The oligomers produced are also similar in appearance to those produced by $A\beta_{1\text{-}42}$ peptide in pH 2.0 aqueous HCl after 4 days at 37°C. This similarity may indicate a comparable overall morphology and therefore it would be worthwhile to test for toxicity. The G4R cystatin B oligomers produced are also similar to those formed by purified monomeric human cystatin C shown in figure 4.3.8. These cystatin C oligomers are observed in equilibrium with fibre in conditions such as pH 2.0.

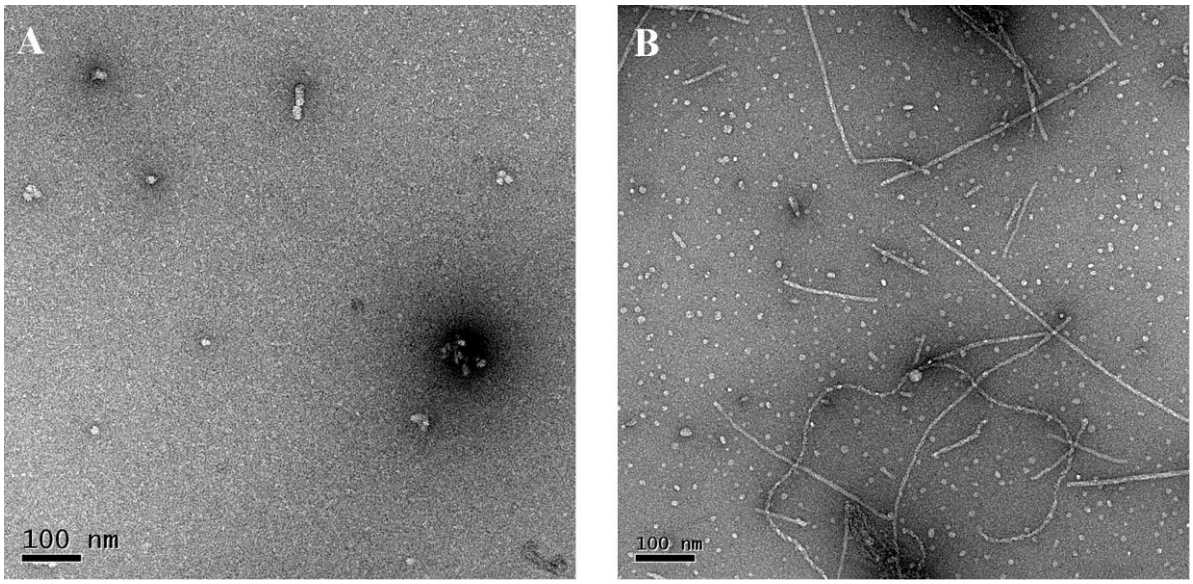


Figure 4.3.7 WT Cystatin B and Aβ Oligomeric Structure Comparison

(A) A cystatin B protein preparation was stored for 2 years in pH 4.7 fibrilisation buffer (minus 10% TFE) at 4°C, after which a significant volume became cloudy in an analogous manner to fibre preparations. Sampling of this solution and EM (21 000 x magnification) indicates these conditions have induced a small proportion of oligomer formation with the electron micrograph representative of the few structures seen. **(B)** 21 000 x magnification of $A\beta_{1-42}$ peptide solution in fibre forming conditions of pH 2.0 aqueous HCl (4 days at 37°C), following ultracentrifugation for 30 minutes at 190 000 x g (RCF_{Max}). The oligomers observed in this image are similar in apparent morphology to the smaller structures formed by cystatin B.

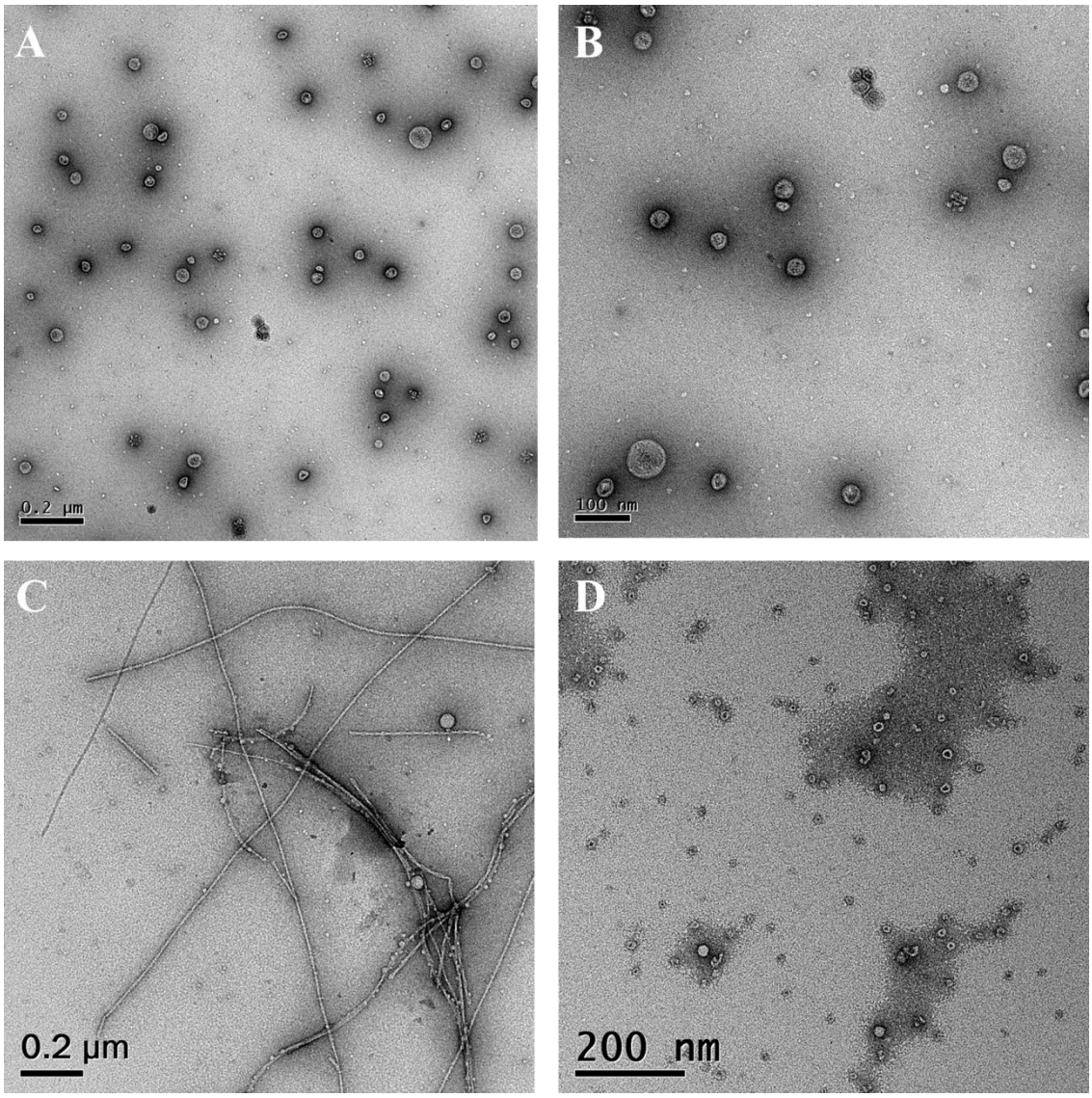


Figure 4.3.8 G4R Cystatin B and WT Cystatin C Oligomer Comparison

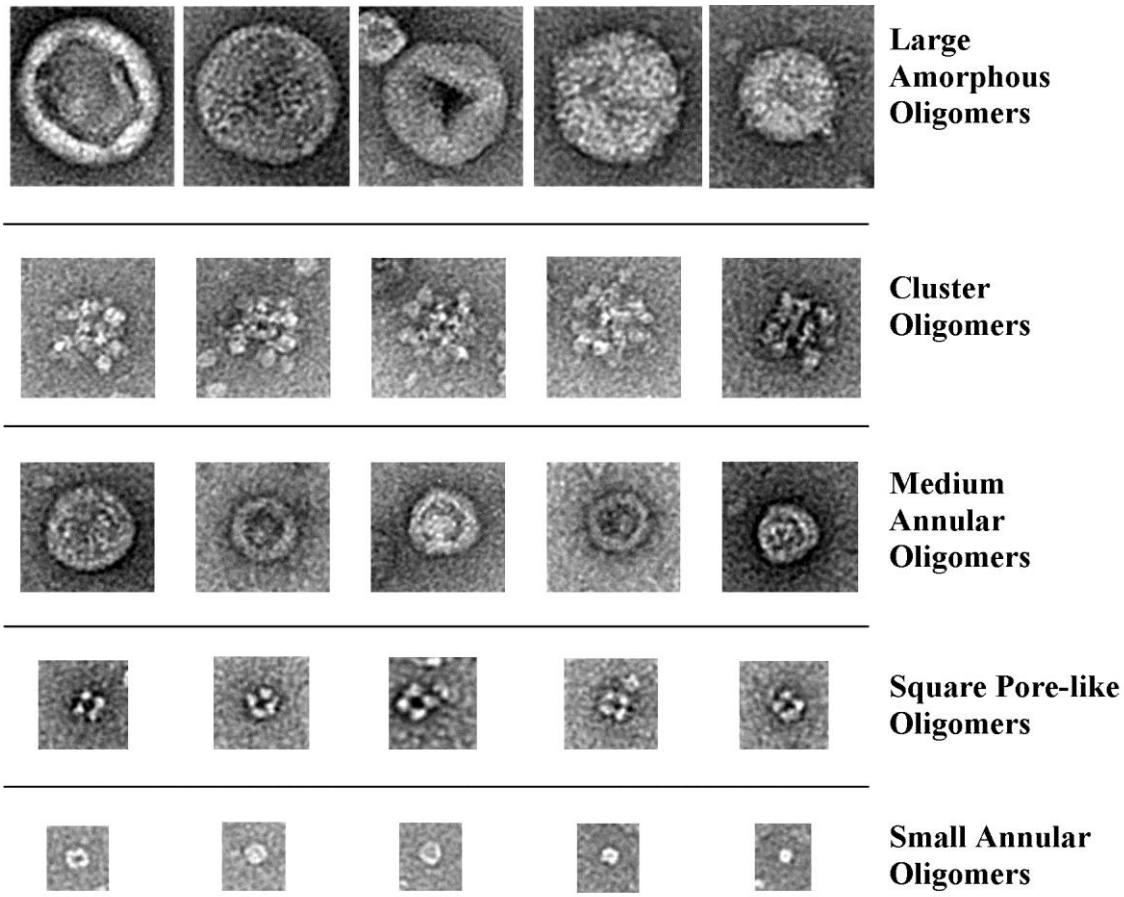
(A and B) Negatively stained electron micrographs of oligomer at 11 500 x (A) and 21 000 x magnification (B). G4R cystatin B oligomer observed after purification from *E. coli*. **(C and D)** Cystatin C oligomer prepared from purified monomeric protein in fibre forming conditions, 11 500 x magnification (C) and 21 000 x magnification (D). During fibre formation these oligomers are observed in equilibrium with fibre and associate with the fibre to a high degree. Cystatin C images kindly supplied by Abigail Williams.

4.3.2 Oligomer Characterisation

Classification of Oligomer Morphology

The G4R cystatin B HO preparations display at least five distinct morphology classes as shown in figure 4.3.9 with the average diameters and statistics. The five classes are; small annular, 'square' pore like, medium biconcave, cluster and large amorphous oligomers. The small annular oligomers are the smallest EM measureable structures and define the limitations of the technique and equipment used. Smaller potentially oligomeric structures were possibly observed in electron micrographs, however the contrast was not high enough to accurately assess the structure, delineate the edges or confidently distinguish them from the background. The 'square' pore-like oligomers are perhaps composed of four small oligomer structures associating together to form a clear central cavity. The medium annular oligomers along with large oligomers exhibit large diameter variation and primarily display an apparent biconcave appearance, although a hollow sphere is also feasible. The cluster oligomers appear to be a conglomeration of very small annular oligomers associating together generally in an apparent disc. The cluster morphology could be a proto-medium annular oligomer, which then matures and reduces in size through compaction. From EM observation and measurement, the oligomers were displayed as flat round disks which appear to bind the grid in a uniform orientation.

The frequency of oligomers and distribution of classes varied from preparation and grid area. For instance the first preparation was devoid of square pore-like oligomers, but contained the clusters, whereas in the third successful preparation this situation was reversed. Overall, in the electron micrographs a qualitative frequency was, small > medium > large > cluster > square pore-like oligomers. A more detailed determination of frequency would likely be a falsehood, due to variations in EM grid binding, sampling of the grid and contrast difficulties which would likely skew the values in favour of the larger structures.



Oligomer Description	Average Diameter (nm)	Number of Particles Measured	Maximum Diameter Observed (nm)	Minimum Diameter Observed (nm)
Small annular	10.6 ±1.4	26	13.1	8.2
Square Pore-like	15.0 ±0.4	6	15.5	14.3
Medium annular	33.5 ±4.7	30	43.7	23.5
Cluster	40.82 ±3.3	8	45.7	36.2
Large Amorphous	67.7 ±13.1	6	88.3	48.3

Figure 4.3.9 G4R Cystatin B Oligomer Structural Classes and Diameter Values

Selected images of each structural class. The 21 000 x negatively stained electron micrographs have been proportionally increased in size by a third for printed clarity. Standard deviation was used to calculate the variance in the average diameter measurements.

Oligomer Species Mass Determination

G4R oligomer mass determination was attempted with SDS-PAGE, SEC HPLC and MS. The high molecular weight oligomer fractions did not migrate readily through 4/20% discontinuous SDS-PAGE gels thus suggesting 1% SDS resistance. The stacking gel interface in oligomer containing lanes did however positively stain for protein with coomassie stain to a small extent, therefore also suggesting proteinaceous content. Elution on a Shodex KW803 SEC HPLC column produced a single peak containing species >700 kDa (>63-mer), which matches the 2 M Da Blue Dextran which elutes at the column exclusion limit (700 kDa) (figure 4.3.10).

Oligomer preparations were analysed by MALDI-TOF and ESI-TOF mass spectroscopy analysis, although were negative for masses corresponding to the G4R protein and also did not contain any other significant mass values which would correspond to phospholipid. The sample mass was possibly too low for detection (<<µM), did not vaporise effectively in the spectrometer or did not fragment sufficiently into smaller species suitable for analysis. Future experimentation to determine the mass, could include the use of a SEC column with a larger mass range, sucrose-gradient analytical ultracentrifugation and electrospray ionisation ion mobility mass spectrometry (ESI-IMS-MS).

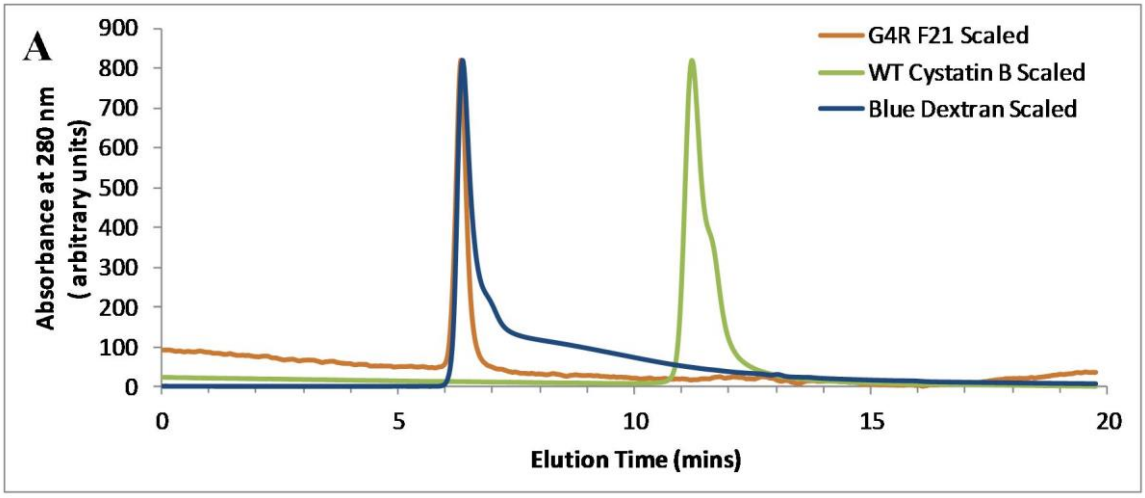


Figure 4.3.10 G4R Oligomer Size Determination by SEC HPLC

The brown peak is fraction 21 from the SEC of G4R purification step which matches the dark blue, blue dextran 200 MDa peak, and indicates the exclusion volume of the column (700 kDa). The green peak is from a typical preparation mixture of dimeric and monomer WT cystatin B.

Immunoassay

To determine the concentration and confirm G4R oligomer identity a dot blot immunoassay was performed (see section 2.6 for the method). The primary antibody used was a polyclonal antibody raised against a full length *E. coli* derived WT cystatin B (R&D systems, AF1408), thus should reliably bind most forms of cystatin B through a variety of available paratopes. This antibody binds WT cystatin B readily up to a 3 µM concentration, however binds soluble G4R 10-fold lower than the WT as shown in figure 4.3.11. Therefore the epitopes present in the antibody are most likely detecting the N-terminal region, which may be less structured in the G4R mutation. The absence of G4R oligomer detection may however be a function of the antibody sensitivity to the low concentration oligomer. This was clearly shown by low sensitivity binding to the G4R monomer/dimer/tetramer and thus does not necessarily reflect the actual oligomer identity. Furthermore, the oligomer may not display a native-like epitope, thus oligomer specific antibodies such as 6E11 raised against Aβ oligomers may be more useful in such an assay.

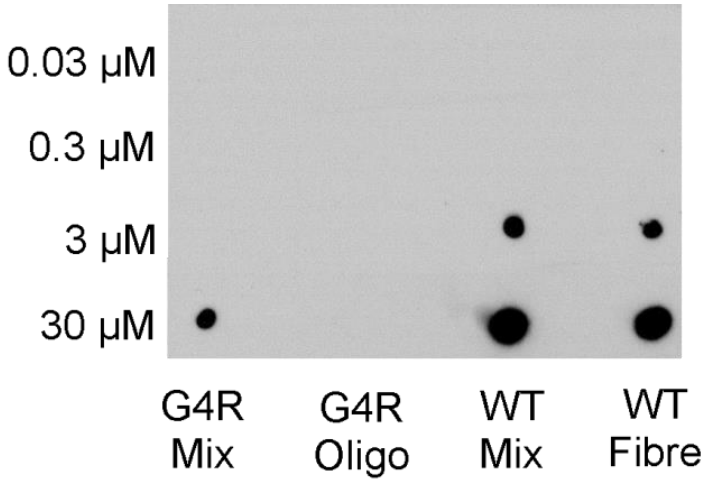


Figure 4.3.11 Dot Blot Immunoassay of G4R and WT Cystatin B

The maximum G4R oligomer concentration was likely to be 10 µM or less and not the stated 30 µM. The other sample concentrations are reliably determined by UV spectroscopy. A 30 second (shown), 5 minute and 19 hour exposure was obtained to ensure sensitive detection.

1D Proton NMR

In an attempt to determine structural information and confirm protein identity, solution state 1D proton NMR was performed and is shown in figure 4.3.12. The spectrum of the oligomeric preparation however yielded no protein specific peaks. This could be due to either insufficient protein concentration (<50 µM) which produces spectra with a high signal to noise ratio and thus is not interpretable. It is also highly likely that the large oligomers will exhibit a low intrinsic tumbling rate in solution which leads to extensive line broadening and therefore no tangible protein signals. In addition, signals from long chain hydrocarbons were not detected which would be indicative of phospholipid presence. Therefore it is very unlikely that the structures observed are lipid based, because even the presence of small amounts would produce a very sharp strong signal.

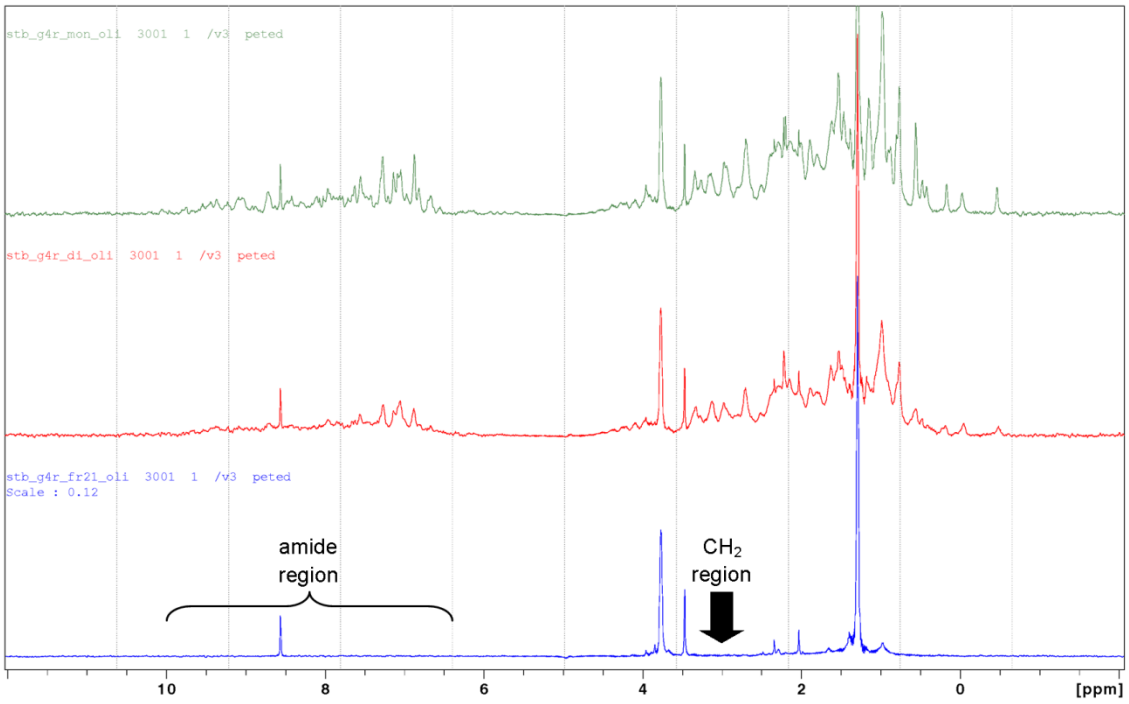


Figure 4.3.12 1D Proton NMR Spectrum of G4R Cystatin B Species

(Green) monomer, **(red)** Dimer and **(blue)** oligomer. The oligomer spectrum shown had a 20 times greater number of scans than the comparative spectra, yet did not present detectable amide signals (10-7.5 ppm) nor signals from long chain hydrocarbons, which would be indicative of phospholipid. Spectra were acquired on a Bruker 600 MHz spectrometer equipped with a TXI cryoprobe at 25°C.

Solvent Dissolution

Following this experiment, the oligomer was lyophilised and dissolved in 95% d_6-DMSO (5% H_2O, dichloroacetic acid, pH 5.0) to determine the feasibility of quenched H/D exchange NMR for structural analysis. Subsequent 1D proton NMR produced no observable proteinaceous NMR signals, which is most likely a function of low concentration. EM of this solution yielded images with small regions of non-uniform aggregated protein and non-distinct oligomer-like structures which suggests either freezing or DMSO dissolution has destroyed the oligomeric conformers (see figure 4.3.13). Addition of either chloroform or hexane in a 1:1 ratio with buffered oligomer preparations also induced amorphous aggregation and furthermore indicates the proteinaceous content of the oligomer preparation.

Dissolution of G4R oligomer was also attempted by solvent exchange to 5.5 M guanidine hydrochloride (unbuffered at pH 5.3) by ultrafiltration. Only the large oligomers remain, suggesting the smaller oligomers are less stable conformers (figure 4.3.13). The highest G4R oligomer concentration determined was 5.8 μM (UV absorbance spectroscopy) and displayed a characteristic protein profile. Further dissolution attempts for limited proteolysis and H/D exchange were performed by SEC of oligomers in 100% acetonitrile (0.1% TFA) which did not work, followed by success with 70% acetonitrile/30% water (0.1% TFA). Elution peaks from at least 4 distinct molecule populations were observed, with the initial peak (lowest amplitude peak) correlating with the elution of WT soluble protein, thus suggesting dissolution in 70% acetonitrile (see figure 4.3.14). The subsequent peaks most likely represent species which are retarded on the column matrix as protein degradation was not observed in other experiments and perhaps relates to the various observed structural classes. A similar occurrence of protein species SEC retardation has been observed with cystatin B fibre dissolution in 6.0 M GuHCl (Sharma, 2009).

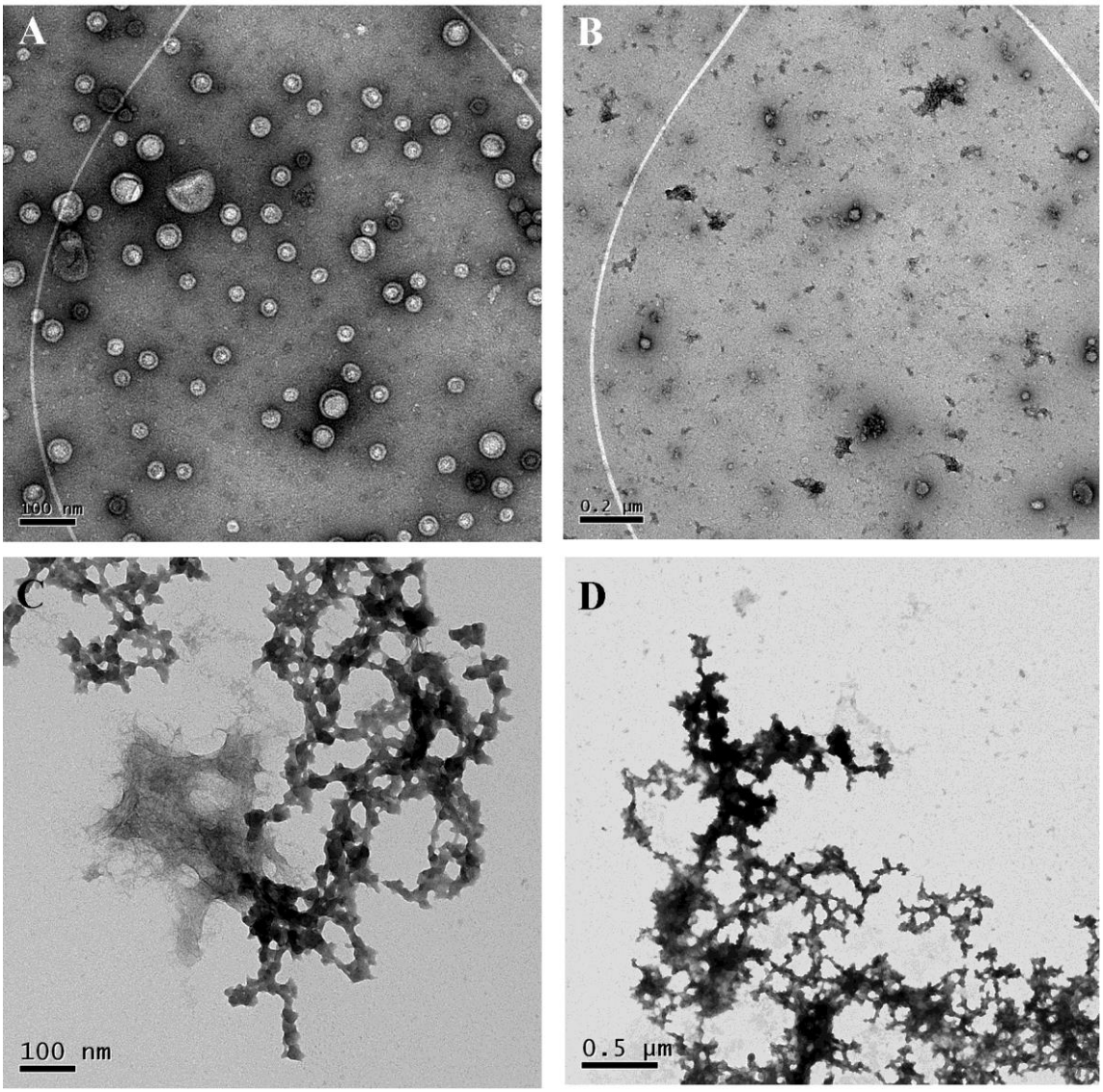


Figure 4.3.13 G4R Cystatin B Oligomer Solvent Dissolution

Electron micrographs of G4R oligomer dissolution attempts. The white ring observed in the first two electron micrographs are camera artefacts caused by uneven cooling, a result of equipment issues at the time. **(A)** 21 000 x magnification electron micrograph of G4R cystatin B oligomer exchanged into a final concentration of 5.5 M guanidine hydrochloride by ultrafiltration. Only the large oligomers remain, suggesting the smaller oligomers are less stable conformers. **(B)** 11 500 x magnification micrograph of G4R HO in 95% d_6-DMSO (5% H_2O, DCA, pH 5.0) showing small regions of aggregated protein and non-distinct oligomer-like structures which suggests either freezing or DMSO dissolution destroys oligomeric conformers. **(C and D)** Addition of either chloroform or hexane in a 1:1 ratio with buffered oligomer preparations induced amorphous aggregation. Electron micrographs are produced from protein in the organic solvent solutions following phase separation. **(C)** 21 000 x magnification of chloroform treated G4R oligomer. **(D)** 5 200 x magnification of hexane treated G4R oligomer.

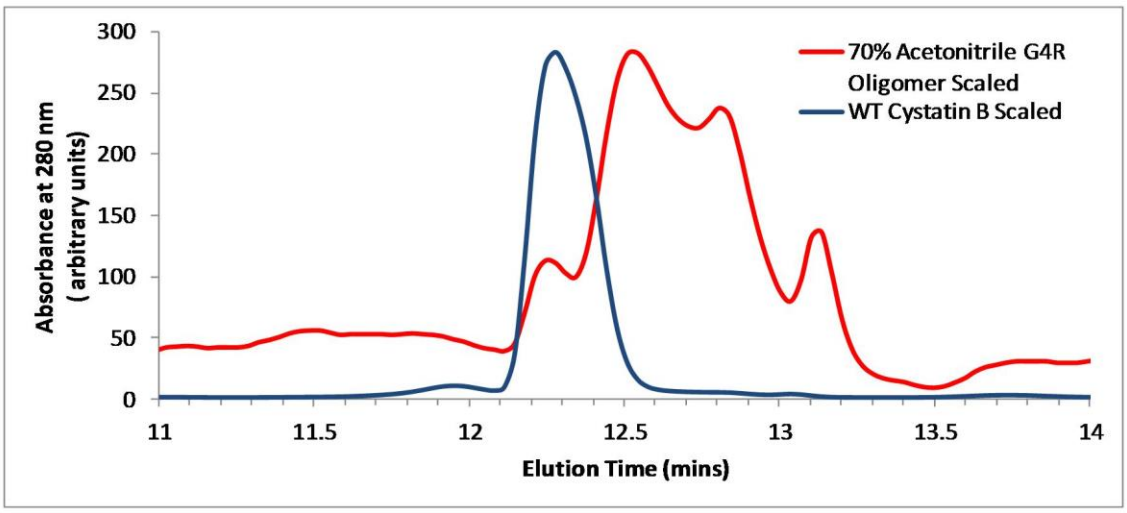


Figure 4.3.14 G4R Cystatin B Oligomer SEC HPLC in 70 % Acetonitrile

SEC of cystatin B G4R oligomers and WT tetramer/dimer/monomer in 70% acetonitrile/30% water (0.1% TFA) (Shodex KW803 column). The chromatogram shows elution peaks from at least 4 distinct molecule populations, with the initial peak correlating with the elution of WT soluble protein and indicating oligomer dissolution. The subsequent peaks most likely represent species which are retarded on the column matrix.

Proteinase K and Lipase Digestion

Limited proteolysis with 1:1000 (monomeric equivalent mass ratio) of proteinase K: G4R oligomer was performed at 25°C over a period of 24 hours using identical methods described in section 3.2. The oligomeric structures however displayed complete proteinase K resistance by both MS and EM (see figure 4.3.15). This resistance suggests that these structures are compact, strongly folded and have a low frequency of local unfolding events in a similar manner to amyloid fibre. A digestion with approximately equimolar lipase, from Aspergillus oryzae (Sigma-Aldrich, 62285), at 37°C for 24 hours was also performed to preclude the possibility that these oligomeric structures are lipid based. The EM results after 24 hours identified that these structures were not disturbed by the robust digestion conditions and are thus unlikely to be phospholipids from the *E. coli* expression cells (see figure 4.3.16).

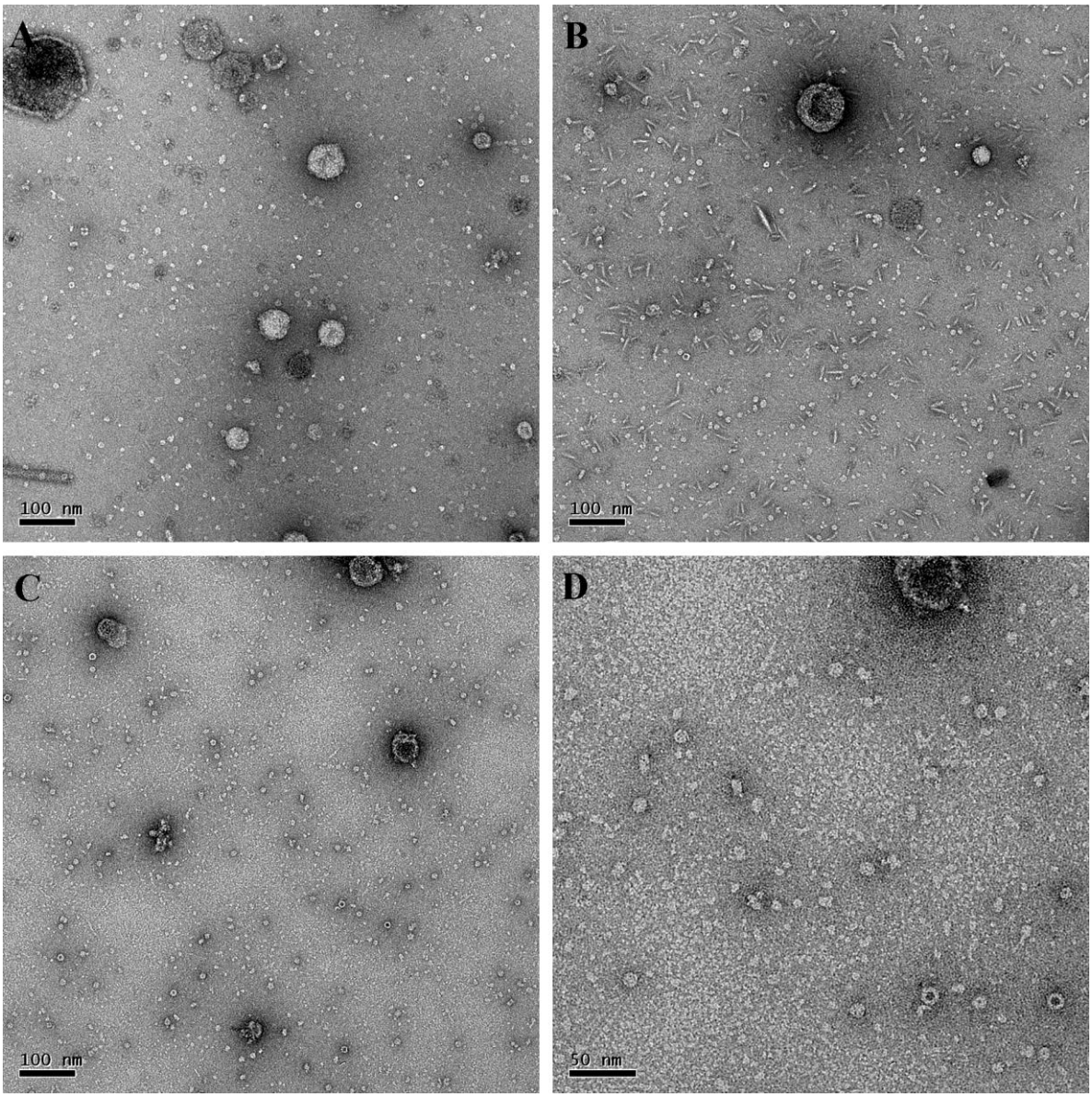


Figure 4.3.15 G4R Oligomer Limited Proteolysis Electron Microscopy

Representative electron micrographs of G4R cystatin B oligomer before and after proteinase K digestion. **(A and B)** 21 000 x magnification of oligomer preparation after concentration and centrifugation and prior to proteolysis. Panel A shows heterogeneous oligomers and a few mature fibres. Panel B is an image from a different section of the same EM grid and shows small protofibrils as well as the same heterogeneous oligomeric species. **(C and D)** 21 000 and 52 000 x respective magnification of oligomer digested for 24 hours with proteinase K. All recorded images showed little change in apparent quantity or morphology of the oligomers following protease digestion. There is however, a distinct absence of protofibrils.

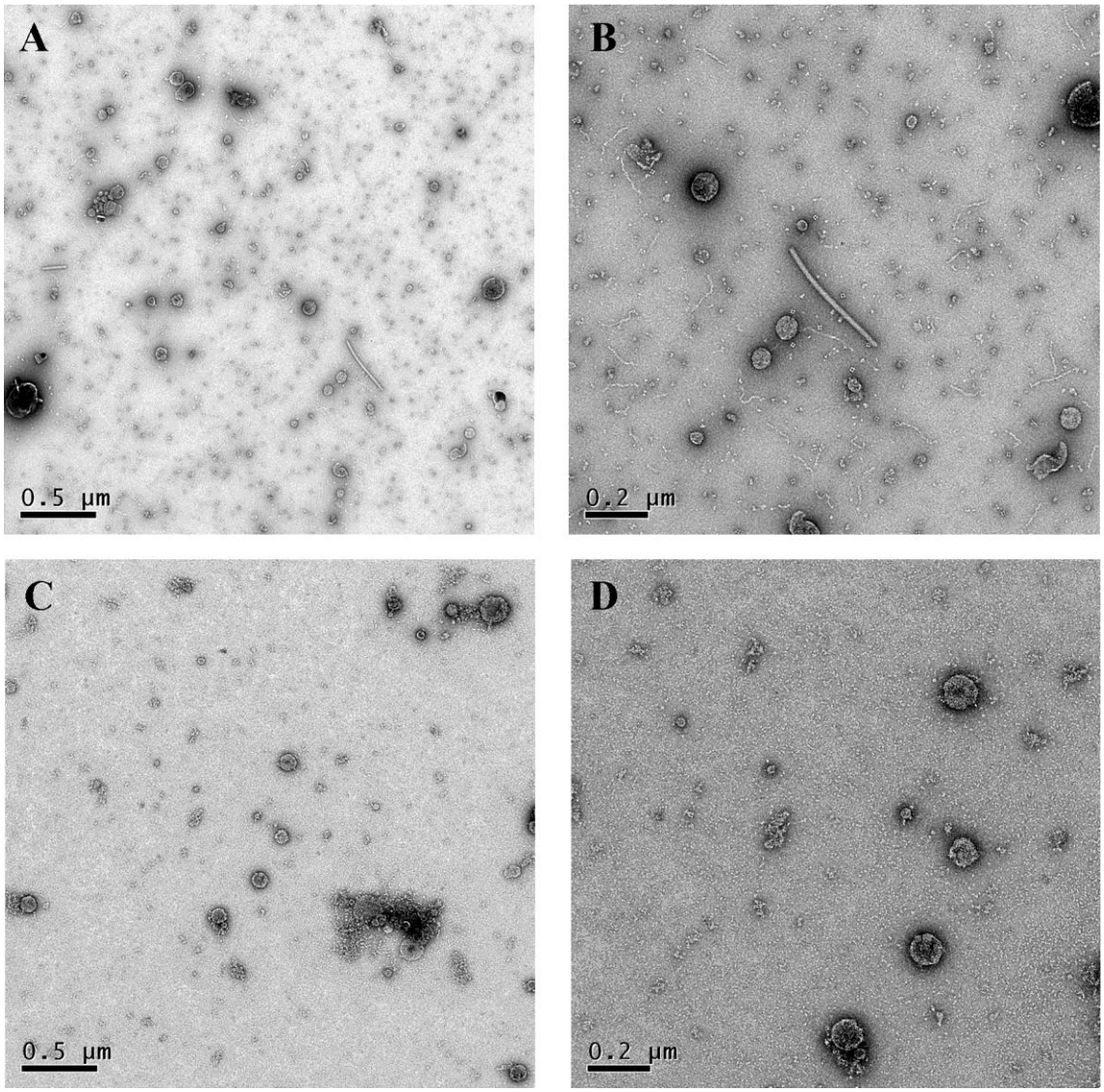


Figure 4.3.16 G4R Cystatin B Oligomer Lipase Digestion

Negatively stained electron micrographs of oligomer at 5 200 x (left) and 11 500 x magnification (right). **(A and B)** Oligomer preparation prior to lipase digestion. **(C and D)** Oligomer preparation following lipase digestion for 24 hours at 37°C.

Circular Dichroism and Thioflavin T Fluorescence

To determine the secondary structure composition of the G4R oligomer, circular dichroism (CD) was performed and compared to WT cystatin B monomer/dimer/tetramer mixtures and amyloid fibres, as shown in figure 4.3.17. The G4R oligomers produced a weak CD signal, perhaps due to low protein concentration, but still yields residue elipticity [Θ]MRW values above the level of signal to noise at 208, 218 and 224 nm (peak value at 224 nm of -159.18x10-3 deg·cm2/dmol-1), thus suggesting the existence of some structural elements. The oligomers may have induced caused some light scattering as indicated by some extra absorbance at >250 nm and red-shift observed around 224 nm. The cystatin B controls were serially diluted 10-fold from 30 µM and both spectra became less reliable and noisy at 3 µM and very poor at 0.03 µM, thus indicating the limitations of this experiment at protein concentrations similar to that expected of the oligomer. This experiment therefore provides tentative evidence of protein secondary structure content. G4R cystatin B oligomer was observed not to induce thioflavin T fluorescence, suggesting the absence of extensive repeating β-sheet structure (see figure 4.3.18).

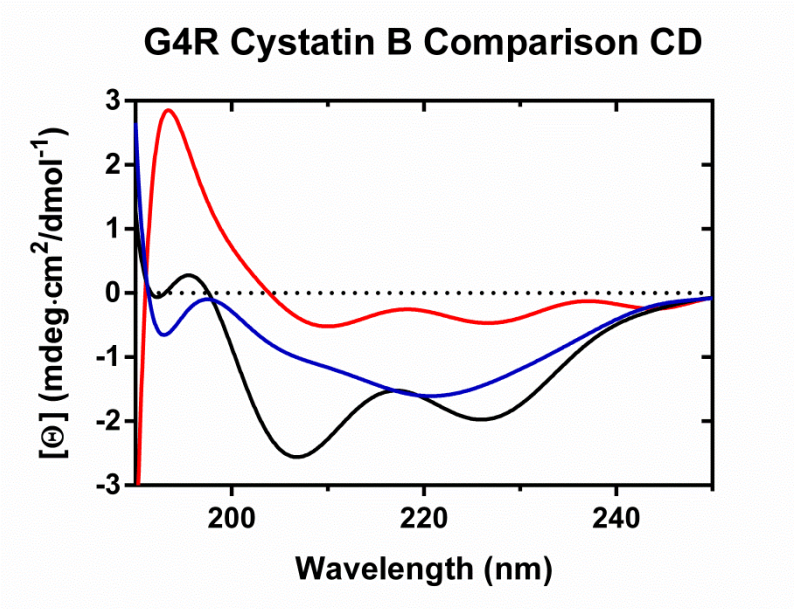


Figure 4.3.17 G4R Cystatin B Oligomer Circular Dichroism

(Red) G4R cystatin B oligomer (~10 µM), **(Black)** WT cystatin B monomer/dimer/tetramer mixture (30 µM), **(Blue)** WT cystatin B fibre (30 µM). Values obtained were subtracted from a buffer baseline.

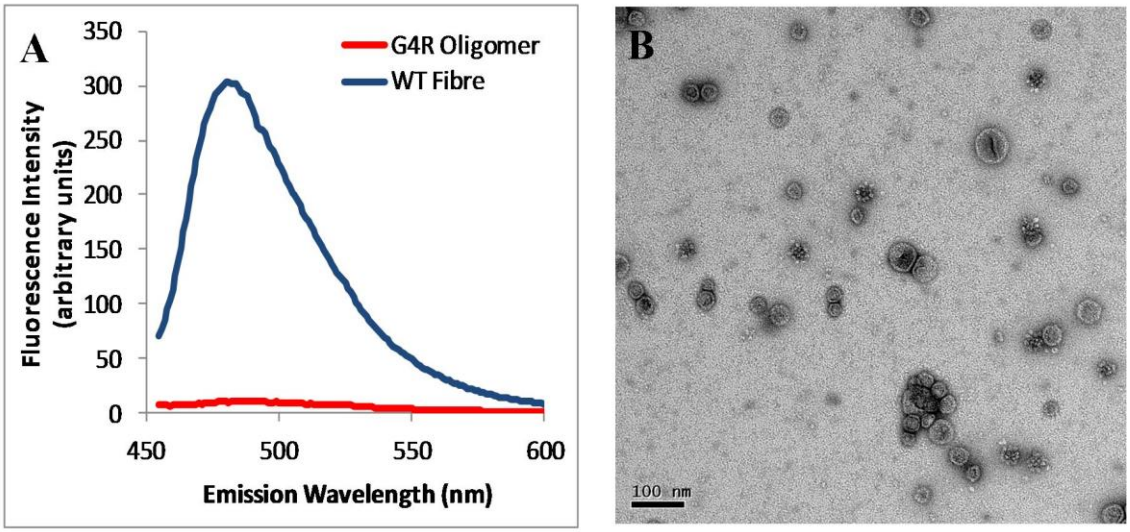


Figure 4.3.18 G4R Cystatin B Thioflavin T Binding

Fluorescence emission spectrum showing absence of G4R oligomer inducing thioflavin T fluorescence. The WT cystatin B fibre preparation was prepared at the same 10 µM concentration as the oligomer preparation and was used as a control. **(B)** 21 000 x magnification electron micrograph of G4R oligomer used in the thioflavin T binding study.

4.4 Discussion

4.4.1 Rationalisation of Oligomerisation

In WT cystatin B, generation of HO structures, even in fibrilisation conditions is a rare occurrence, suggested by the absence in EM time-course experiments (Paramore, 2010). The N-terminal region of cystatin B is therefore suspect in mediating interactions between monomer/ dimer and tetramers with a position 4 arginine improving binding and subsequent HO generation. Furthermore, the G4R mutant in typical pH 4.7 10% TFE fibre forming conditions is suggested to have a much longer lag phase than the WT and accumulates more prefibrillar aggregates (Rabzelj *et al.*, 2005).

4.4.2 Oligomer Formation

Oligomer was produced only in small amounts in soluble protein extractions from *E. coli*, thus suggesting the formation of these structures requires highly specific conditions within the cell, which would perhaps be difficult to reproduce *in vitro*. Yields were typically less than 1.5 mg/l of LB growth media and concentration by ultrafiltration has proved unsuccessful, perhaps due to adhesion to the cell membranes. Perhaps flowing nitrogen over the solution surface could prove successful as a means of protein concentration in future experiments. Interestingly these oligomers remain stable for months at 4 °C which is ideal for experimental manipulation with the caveat that high concentrations of protein >10 mg/ml causes aggregation and/or dissociation. In addition these G4R oligomers failed to produce fibres in the standard fibrilisation conditons, thus suggesting these particular oligomers are stable structures off-pathway to amyloid fibre formation.

4.4.3 Confirmation of Protein Identity

I believe these oligomers are definitively proteinaceous from the range of experiments performed, despite their similarity to vesicles in EM (See table 4.4.1 for an overview of the current oligomer characterisation). The reasons are as follows: characteristic protein UV absorption spectrum, characteristic folded protein spectrum, coomassie blue dye binding in SDS-PAGE, dissolution in polar aprotic solvents, as well as robust lipase resistance. In addition, the similarity to structures formed over time from purified Aβ and cystatin B indicates these structures are similar. Finally, the variance between WT cystatin B rarely forming these structures and G4R reproducibly producing measurable populations also clarifies that these structures are only formed significantly when the cystatin B protein is destabilised in a specific way.

Observation	Experiment	Comment
Low 280 nm absorbance	UV spectrometer, 1 cm pathlength	Reliably detected with liquid chromatography flow cells, thus likely to be at a low concentration (<10 µM for WT cystatin B). Produces typical absorbance spectrum when successful. Large structures may cause light scattering.
Ultrafiltration failed to increase oligomer concentration (5x concentration attempted)	Ultrafiltration (10 kDa MWCO vivaspin 500 centrifugation cell)	Oligomer probably binds the polyethersulfone membranes, and also aggregates/dissociates at high concentrations.
Does not migrate effectively in SDS-PAGE	SDS-PAGE	SDS resistant and low concentration (<5 µg/ml). Oligomer binds coomassie blue stain in the stacking/resolving gel interface.
5.5 M Guanidine Hydrochloride resistant	EM, NMR	Resistant to strong denaturant
Dissolves in 100% DMSO	EM, NMR	Dissolves in polar aprotic solvent
Dissolves in 70% acetonitrile, H_2O, 0.1% TFA mix	SEC HPLC	Dissolves in polar aprotic solvent
Amorphous aggregate identified when mixed in sodium phosphate buffer with chloroform or hexane	EM	Non-polar solvents cause protein aggregation at the solvent interface
Proteinase K Resistant	EM	Resistant to efficient sequence independent protease
Lipase Resistant	EM	24 hour 37°C digest.
Not observable in 1D 1H NMR	NMR	Large size dictates that the tumbling rate is likely to be slow, thus extensive line broadening occurs.
Not observable in MS	ESI-TOF/MALDI-TOF MS	Large oligomers did not sufficiently vapourise or fragment. The presence of lipids was not observed.
No antibody binding	Dot blot immunoassay	The primary antibody binding to the control G4R monomer/dimer/ tetramer was ineffective.

Table 4.4.1 Cystatin B High Molecular Weight Oligomer Characterisation

4.4.4 Oligomer Morphology

G4R Cystatin B produces measurable quantities of higher molecular weight oligomer < 700 kDa in at least five different conformations, which are larger in diameter than WT cystatin B amyloid fibre. The diameter measurements (max and min) range in size from 8-88 nm. The small annular oligomer is close however, at an average 10.6 nm diameter compared to the fibre diameter of 8.6 nm and this is not taking into account the possibility of smaller structures at the limits of detection which may be similar in size.

4.4.5 Oligomer Structure

The absence of thioflavin T binding indicates there is no large extended β-sheet similar to that of amyloid fibre. The circular dichroism data tentatively suggests there is presence of some secondary structure but identification is currently tenuous without producing more oligomer and concentrating it for use in CD and FTIR. The oligomers are highly resistant to strong denaturant such as 5.5 M guanidine hydrochloride, 1% SDS and impervious to hydrolysis by the potent proteinase K enzyme, thus indicating a very compact densely structured moiety. However they can be dissolved by polar aprotic solvents such as 100% DMSO and 70% acetonitrile, thus may be accessible for H/D exchange studies once the yield is improved. At this juncture little is known about the structure of these cystatin B oligomers or those formed by other proteins. There is also no clear consensus in the literature of any particular structural preference, which may indicate an absence of a definitive structural entity present in these oligomers (Fandrich, 2012). Hopefully with future work these rather intriguing structures will be elucidated.

4.4.6 Future Work

Higher yields of HO induced by novel therapeutics in preparations from *E. coli* or from purified monomer/ dimer / tetramer mixes would dramatically improve the study of these interesting structures. Another important step in the study of these structures would be separation of the different conformers, perhaps by specific antibodies, differential centrifugation of filtration. A limitation however, might gaining a suitable concentration of the oligomers for structural methods. EM tomography, FTIR, AFM and gradient analytical ultracentrifugation may provide more detailed structural information. However, for all these experiments a low protein concentration would limit structural methods. An example of this is H/D exchange which typically requires 100 μM preparations as a minimum. Any information produced would be highly valuable with the sensible next step being identification of toxicity or analysis of effects on long term potentiation in neuronal cells (Amijee *et al.*, 2012) alongside structural studies and oligomer specific antibody binding. These oligomers are of specific interest as they appear stable over a period greater than three months, as determined by EM, which therefore increases suitability for structural studies.

Chapter 5: Final Conclusions and Future Perspectives

5.1 Final Conclusions

The initial aim of this project was to develop faster, more generally applicable methods for defining the structure of cystatin amyloids and higher molecular weight oligomeric structures. The methods employed were developed with the aim of using them to probe fibre polymorphism and determine if fibre structure is conserved between cystatin B and C, the former being better characterised. Further to this, the methods were to be used in structural studies of cystatin oligomers.

The use of limited proteolysis, TEM and STEM MPL (mass mapping) on cystatin B fibre, described in this thesis, has generated new information which challenges the currently posited model generated from H/D exchange data (Morgan *et al.*, 2008). Limited proteolysis of cystatin B amyloid fibre has shown that the core structure consists at the maximum, residues 27-80 out of a total 98. The first 27 residues (natively a β-sheet and α-helix) are readily cleaved whilst C-terminal residues 80-98 (natively β-sheet) are digested at a slower rate than the N-terminal region. We therefore propose a new simple working model (see figure 3.4.2) based on the data presented in chapter 3 and H/D exchange data (Morgan *et al.*, 2008). This model posits that the native β-strand 2 and 3 is extended into a single strand stacking in parallel to form the fibre 'backbone' whilst β-strand 4 and 5 forms the second arm of a β-strand arc (hairpin). Furthermore, limited proteolysis on fibres formed at pH 3.3 and 4.0 suggest that the amyloid fibre is not clearly polymorphic when compared to the structures formed at pH 4.7. Furthermore, the experimentally derived fibre core is consistent with predictions from a range of sequence specific amyloidogenic region programs (see section 3.3.4).

Electron microscopy techniques, including mass per unit length measurements, have shown that the fibre is 8.6 nm in width and forms at least four classes of fibre indicating the architecture of the protofibril/fibre association. The class I fibres are most likely

composed of 4 cystatin B molecules per 4 β-strand rise, with subsequent doublets of class I explaining the presence of fibres with 8 and 16 molecules of cystatin B per 4 β-strand rise (Class III and IV). Class II fibres contain 6 molecules per 4-strand rise, which is more difficult to envisage within the reported fibre dimensions, although is reminiscent of proposed models for insulin (Jimenez et al., 1999).

The reliable generation of stable G4R mutant cystatin B oligomer and initial EM and biochemical characterisation has been performed with the aim of initiating future structural analysis of this elusive species. G4R cystatin B oligomers are heterogeneous with at least five distinct morphology classes, with diameters ranging from 8-88 nm. The five classes are; small annular, 'square' pore-like, medium biconcave, cluster and large amorphous oligomers, each having a different morphology and size distribution. These oligomers are very robust and have been shown by EM to be resistant to 24 hour proteinase K digestion (only some classes) and the structures persist in the strong denaturant guanidine hydrochloride (5.5 M). However, these oligomers can potentially be dissolved or aggregated in organic solvents such as DMSO and 70:30% v/v acetonitrile and water, thus enabling the use of quenched H/D exchange. Furthermore, these oligomers remain stable for months when observed by EM, thus are suitable for structural studies. Limitations on the study of these structures include: homogeneity, low yields and difficult manipulation. Hopefully these challenges are not insurmountable.

The investigation of small molecules which induce oligomerisation, such as epigallocatechin gallate (EPCG) were initially of interest in the generation of high yield cystatin oligomer. This initial interest led to the collaborative study of a potential amyloid therapeutic present in *Salvia sclareoides* plant extracts. Examining the efficacy and effects of these extracts was performed by thioflavin T dye binding kinetic assays, EM and additional NMR data. The data collected suggests these extracts can potently destabilise mature fibres, leading to the rapid formation of amorphous aggregates and thus has provided the basis for identification and mechanism elucidation.

5.2 Future Work

The structural work on cystatin B presented in this thesis will aid an on-going SSNMR cystatin B fibre structural study, which will lead to the determination of an atomic resolution model. In addition to this, it may be interesting to explore the removal of the native α-helix region with the α-3C-β mutant or a stable elastase digestion fragment with the primary aim of determining interactions between the unfolded moieties and whether these mediate lateral association. Further to this, an array of truncated mutants described in the appendix could act as a verification of the structural models. The work to date includes the design and production of new plasmid constructs as well as preliminary peptide purification procedures, therefore would require little extra preliminary investigation.

Comparison of the cystatin B and cystatin C amyloid fibre structure could prove very interesting in terms of structural polymorphism between two homologous natively folded proteins, the latter directly associated with cerebral amyloid angiopathy. For the comparison study, limited proteolysis, TEM and STEM MPL measurements, which are relatively quick, may be sufficient for structural assessment. The determination of structural information from oligomeric structures is of keen interest to the amyloid field at present and if successful, would bridge a gap in current understanding and aid future therapeutic development. Furthermore, the discovery of the active compound in Salvia sclareoides, if proven novel, could be a very exciting discovery, as suggested by the very rapid destabilisation of cystatin B amyloid fibre.

References

ABRAHAMSON, M. 1993. Cystatins – Protein inhibitors of papain-like cysteine proteinases. *Ciencia e Cultura (Journal of the Brazilian Association for the Advancement of Science),* 45**,** 299-304.

ALAKURTTI, K., WEBER, E., RINNE, R., THEIL, G., DE HAAN, G. J., LINDHOUT, D., SALMIKANGAS, P., SAUKKO, P., LAHTINEN, U. & LEHESJOKI, A. E. 2005. Loss of lysosomal association of cystatin B proteins representing progressive myoclonus epilepsy, EPM1, mutations. *European Journal of Human Genetics,* 13**,** 208-215.

ALEXANDRESCU, A. T. 2001. An NMR-based quenched hydrogen exchange investigation of model amyloid fibrils formed by cold shock protein A. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing***,** 67-78.

AMIJEE, H., BATE, C., WILLIAMS, A., VIRDEE, J., JEGGO, R., SPANSWICK, D., SCOPES, D. I. C., TREHERNE, J. M., MAZZITELLI, S., CHAWNER, R., EYERS, C. E. & DOIG, A. J. 2012. The N-Methylated Peptide SEN304 Powerfully Inhibits A beta(1-42) Toxicity by Perturbing Oligomer Formation. *Biochemistry,* 51**,** 8338-8352.

ANTZUTKIN, O. N., LEAPMAN, R. D., BALBACH, J. J. & TYCKO, R. 2002. Supramolecular structural constraints on Alzheimer's beta-amyloid fibrils from electron microscopy and solid-state nuclear magnetic resonance. *Biochemistry,* 41**,** 15436-15450.

ARTIMO, P., JONNALAGEDDA, M., ARNOLD, K., BARATIN, D., CSARDI, G., DE CASTRO, E., DUVAUD, S., FLEGEL, V., FORTIER, A., GASTEIGER, E., GROSDIDIER, A., HERNANDEZ, C., IOANNIDIS, V., KUZNETSOV, D., LIECHTI, R., MORETTI, S., MOSTAGUIR, K., REDASCHI, N., ROSSIER, G., XENARIOS, I. & STOCKINGER, H. 2012. ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Research,* 40**,** W597-W603.

ATLAS, D., LEVIT, S., SCHECHTE.I & BERGER, A. 1970. ACTIVE SITE OF ELASTASE - PARTIAL MAPPING BY MEANS OF SPECIFIC PEPTIDE SUBSTRATES. *Febs Letters,* 11**,** 281-&.

BALGUERIE, A., DOS REIS, S., RITTER, C., CHAIGNEPAIN, S., COULARY-SALIN, B., FORGE, V., BATHANY, K., LASCU, I., SCHMITTER, J. M., RIEK, R. & SAUPE, S. J. 2003a. Domain organization and structure-function relationship of the HET-s prion protein of Podospora anserina. *Embo Journal,* 22**,** 2071-2081.

BALGUERIE, A., REIS, S. D., RITTER, C., CHAIGNEPAIN, S., COULARY-SALIN, B., FORGE, V., BATHANY, K., LASCU, I., SCHMITTER, J.-M., RIEK, R. & SAUPE, S. J. 2003b. Domain organization and structure-function relationship of the HET-s prion protein of Podospora anserina. *Embo Journal,* 22**,** 2071-2081.

BALLARD, C., GAUTHIER, S., CORBETT, A., BRAYNE, C., AARSLAND, D. & JONES, E. 2011. Alzheimer's disease. *Lancet,* 377**,** 1019-31.

BARD, F., CANNON, C., BARBOUR, R., BURKE, R. L., GAMES, D., GRAJEDA, H., GUIDO, T., HU, K., HUANG, J. P., JOHNSON-WOOD, K., KHAN, K., KHOLODENKO, D., LEE, M., LIEBERBURG, I., MOTTER, R., NGUYEN, M., SORIANO, F., VASQUEZ, N., WEISS, K., WELCH, B., SEUBERT, P., SCHENK, D. & YEDNOCK, T. 2000. Peripherally administered antibodies against amyloid beta-peptide enter the central nervous system and reduce pathology in a mouse model of Alzheimer disease. *Nature Medicine,* 6**,** 916-919.

BARRETT, A. J., FRITZ, H., GRUBB, A., ISEMURA, S., JARVINEN, M., KATUNUMA, N., MACHLEIDT, W., MULLERESTERL, W., SASAKI, M. & TURK, V. 1986. NOMENCLATURE AND CLASSIFICATION OF THE PROTEINS HOMOLOGOUS WITH THE CYSTEINE-PROTEINASE INHIBITOR CHICKEN CYSTATIN. *Biochemical Journal,* 236**,** 312-312.

BARTELS, T., CHOI, J. G. & SELKOE, D. J. 2011. alpha-Synuclein occurs physiologically as a helically folded tetramer that resists aggregation. *Nature,* 477**,** 107-U123.

BARTOLINI, M., NALDI, M., FIORI, J., VALLE, F., BISCARINI, F., NICOLAU, D. V. & ANDRISANO, V. 2011. Kinetic characterization of amyloid-beta 1-42 aggregation with a multimethodological approach. *Anal Biochem,* 414**,** 215-25.

BAXA, U., TAYLOR, K. L., WALL, J. S., SIMON, M. N., CHENG, N. Q., WICKNER, R. B. & STEVEN, A. C. 2003. Architecture of Ure2p prion filaments - The

N-terminal domains form a central core fiber. *Journal of Biological Chemistry,* 278**,** 43717-43727.

BAXA, U., WICKNER, R. B., STEVEN, A. C., ANDERSON, D. E., MAREKOV, L. N., YAU, W. M. & TYCKO, R. 2007. Characterization of beta-sheet structure in Ure2p(1-89) yeast prion fibrils by solid-state nuclear magnetic resonance. *Biochemistry,* 46**,** 13149-13162.

BENILOVA, I., KARRAN, E. & DE STROOPER, B. 2012. The toxic A beta oligomer and Alzheimer's disease: an emperor in need of clothes. *Nature Neuroscience,* 15**,** 349-357.

BIESCHKE, J., RUSS, J., FRIEDRICH, R. P., EHRNHOEFER, D. E., WOBST, H., NEUGEBAUER, K. & WANKER, E. E. 2010. EGCG remodels mature alpha-synuclein and amyloid-beta fibrils and reduces cellular toxicity. *Proceedings of the National Academy of Sciences of the United States of America,* 107**,** 7710-7715.

BLAKE, C. & SERPELL, L. 1996. Synchrotron X-ray studies suggest that the core of the transthyretin amyloid fibril is a continuous [beta]-sheet helix. *Structure,* 4**,** 989-998.

BUELL, A. K., DOBSON, C. M., KNOWLES, T. P. J. & WELLAND, M. E. 2010. Interactions between Amyloidophilic Dyes and Their Relevance to Studies of Amyloid Inhibitors. *Biophysical Journal,* 99**,** 3492-3497.

BULIC, B., PICKHARDT, M., SCHMIDT, B., MANDELKOW, E. M., WALDMANN, H. & MANDELKOW, E. 2009. Development of Tau Aggregation Inhibitors for Alzheimer's Disease. *Angewandte Chemie-International Edition,* 48**,** 1741-1752.

CARULLA, N., CADDY, G. L., HALL, D. R., ZURDO, J., GAIRI, M., FELIZ, M., GIRALT, E., ROBINSON, C. V. & DOBSON, C. M. 2005. Molecular recycling within amyloid fibrils. *Nature,* 436**,** 554-558.

CERU, S. & ZEROVNIK, E. 2008. Similar toxicity of the oligomeric molten globule state and the prefibrillar oligomers. *Febs Letters,* 582**,** 203-209.

CHANDRASEKARAN, R. & STUBBS, G. 2006. Fibre diffraction. *International Tables for Crystallography.* John Wiley & Sons, Ltd.

CHITI, F., STEFANI, M., TADDEI, N., RAMPONI, G. & DOBSON, C. M. 2003. Rationalization of the effects of mutations on peptide and protein aggregation rates. *Nature,* 424**,** 805-808.

CHOTHIA, C. 1973. CONFORMATION OF TWISTED BETA-PLEATED SHEETS IN PROTEINS. *Journal of Molecular Biology,* 75**,** 295-302.

COHEN, F. E. & KELLY, J. W. 2003. Therapeutic approaches to protein-misfolding diseases. *Nature,* 426**,** 905-9.

CONCHILLO-SOLE, O., DE GROOT, N., AVILES, F., VENDRELL, J., DAURA, X. & VENTURA, S. 2007. AGGRESCAN: a server for the prediction and evaluation of "hot spots" of aggregation in polypeptides. *BioMed Central Bioinformatics,* 8**,** 65.

CONNELLY, S., CHOI, S., JOHNSON, S. M., KELLY, J. W. & WILSON, I. A. 2010. Structure-based design of kinetic stabilizers that ameliorate the transthyretin amyloidoses. *Current Opinion in Structural Biology,* 20**,** 54-62.

CORNWALL, G. A., VON HORSTEN, H. H. & WHELLY, S. 2011. Cystatin-Related Epididymal Spermatogenic Aggregates in the Epididymis. *Journal of Andrology,* 32**,** 679-685.

COSTA, P. P., FIGUEIRA, A. S. & BRAVO, F. R. 1978. AMYLOID FIBRIL PROTEIN RELATED TO PRE-ALBUMIN IN FAMILIAL AMYLOIDOTIC POLYNEUROPATHY. *Proceedings of the National Academy of Sciences of the United States of America,* 75**,** 4499-4503.

CREIGHTON, T. E. (ed.) 1989. *Protein Structure: A Practical Approach*: Oxford University Press.

DE LAURETO, P. P., FRARE, E., BATTAGLIA, F., MOSSUTO, M. F., UVERSKY, V. N. & FONTANA, A. 2005. Protein dissection enhances the amyloidogenic properties of alpha-lactalbumin. *Febs Journal,* 272**,** 2176-2188.

DE LAURETO, P. P., TADDEI, N., FRARE, E., CAPANNI, C., COSTANTINI, S., ZURDO, J., CHITI, F., DOBSON, C. M. & FONTANA, A. 2003. Protein aggregation and amyloid fibril formation by an SH3 domain probed by limited proteolysis. *Journal of Molecular Biology,* 334**,** 129-141.

DEMATTOS, R. B., BALES, K. R., CUMMINS, D. J., DODART, J. C., PAUL, S. M. & HOLTZMAN, D. M. 2001. Peripheral anti-A beta antibody alters CNS and plasma A beta clearance and decreases brain A beta burden in a mouse model of Alzheimer's disease. *Proceedings of the National Academy of Sciences of the United States of America,* 98**,** 8850-8855.

DENG, A., IRIZARRY, M. C., NITSCH, R. M., GROWDON, J. H. & REBECK, G. W. 2001. Elevation of Cystatin C in Susceptible Neurons in Alzheimers' Disease. *Am. J. Pathol.,* 159**,** 1061-1068.

DHULESIA, A., CREMADES, N., KUMITA, J. R., HSU, S. T., MOSSUTO, M. F., DUMOULIN, M., NIETLISPACH, D., AKKE, M., SALVATELLA, X. & DOBSON, C. M. 2010. Local cooperativity in an amyloidogenic state of human lysozyme observed at atomic resolution. *Journal of the American Chemical Society,* 132**,** 15580-8.

DIMANT, H., SHARON, N. & SOLOMON, B. 2009. Modulation effect of filamentous phage on alpha-synuclein aggregation. *Biochemical and Biophysical Research Communications,* 383**,** 491-496.

DIMANT, H. & SOLOMON, B. 2010. Filamentous Phages Reduce alpha-Synuclein Oligomerization in the Membrane Fraction of SH-SY5Y Cells. *Neurodegenerative Diseases,* 7**,** 203-205.

DOBSON, C. M. 1999. Protein misfolding, evolution and disease. *TRENDS in Biochemical Sciences,* 24**,** 329-332.

DOBSON, C. M. 2003. Protein folding and misfolding. *Nature,* 426**,** 884-889.

DUMOULIN, M., LAST, A. M., DESMYTER, A., DECANNIERE, K., CANET, D., LARSSON, G., SPENCER, A., ARCHER, D. B., SASSE, J., MUYLDERMANS, S., WYNS, L., REDFIELD, C., MATAGNE, A., ROBINSON, C. V. & DOBSON, C. M. 2003. A camelid antibody fragment inhibits the formation of amyloid fibrils by human lysozyme. *Nature,* 424**,** 783-788.

EBELING, W., HENNRICH, N., KLOCKOW, M., METZ, H., ORTH, H. D. & LANG, H. 1974. PROTEINASE K FROM TRITIRACHIUM-ALBUM LIMBER. *European Journal of Biochemistry,* 47**,** 91-97.

EHRNHOEFER, D. E., BIESCHKE, J., BOEDDRICH, A., HERBST, M., MASINO, L., LURZ, R., ENGEMANN, S., PASTORE, A. & WANKER, E. E. 2008. EGCG redirects amyloidogenic polypeptides into unstructured, off-pathway oligomers. *Nature Structural & Molecular Biology,* 15**,** 558-566.

EHRNHOEFER, D. E., DUENNWALD, M., MARKOVIC, P., WACKER, J. L., ENGEMANN, S., ROARK, M., LEGLEITER, J., MARSH, J. L., THOMPSON, L. M., LINDQUIST, S., MUCHOWSKI, P. J. & WANKER, E. E. 2006. Green tea (-)-epigallocatechin-gallate modulates early events in huntingtin misfolding and reduces toxicity in Huntington's disease models. *Human Molecular Genetics,* 15**,** 2743-2751.

EICHNER, T., KALVERDA, A. P., THOMPSON, G. S., HOMANS, S. W. & RADFORD, S. E. 2011. Conformational conversion during amyloid formation at atomic resolution. *Mol Cell,* 41**,** 161-72.

FALK, R. H., COMENZO, R. L. & SKINNER, M. 1997. Medical progress - The systemic amyloidoses. *New England Journal of Medicine,* 337**,** 898-909.

FANDRICH, M. 2012. Oligomeric Intermediates in Amyloid Formation: Structure Determination and Mechanisms of Toxicity. *Journal of Molecular Biology,* 421**,** 427-440.

FANDRICH, M., FLETCHER, M. A. & DOBSON, C. M. 2001. Amyloid fibrils from muscle myoglobin - Even an ordinary globular protein can assume a rogue guise if conditions are right. *Nature,* 410**,** 165-166.

FAWZI, N. L., YING, J. F., GHIRLANDO, R., TORCHIA, D. A. & CLORE, G. M. 2011. Atomic-resolution dynamics on the surface of amyloid-beta protofibrils probed by solution NMR. *Nature,* 480**,** 268-U161.

FEHER, M. & SCHMIDT, J. M. 2003. Property distributions: differences between drugs, natural products, and molecules from combinatorial chemistry. *J Chem Inf Comput Sci,* 43**,** 218-27.

FERNANDEZ-ESCAMILLA, A.-M., ROUSSEAU, F., SCHYMKOWITZ, J. & SERRANO, L. 2004. Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat Biotech,* 22**,** 1302-1306.

FRARE, E., MOSSUTO, M. F., DE LAURETO, P. P., DUMOULIN, M., DOBSON, C. M. & FONTANA, A. 2006. Identification of the core structure of lysozyme amyloid fibrils by proteolysis. *Journal of Molecular Biology,* 361**,** 551-561.

GALZITSKAYA, O. V., GARBUZYNSKIY, S. O. & LOBANOV, M. Y. 2006. Prediction of Amyloidogenic and Disordered Regions in Protein Chains. *PLoS Comput Biol,* 2**,** e177.

GEDDES, A. J., PARKER, K. D., ATKINS, E. D. T. & BEIGHTON, E. 1968. CROSS-BETA CONFORMATION IN PROTEINS. *Journal of Molecular Biology,* 32**,** 343-&.

GHISO, J., JENSSON, O. & FRANGIONE, B. 1986. AMYLOID FIBRILS IN HEREDITARY CEREBRAL-HEMORRHAGE WITH AMYLOIDOSIS OF ICELANDIC TYPE IS A VARIANT OF GAMMA-TRACE BASIC-PROTEIN (CYSTATIN-C). *Proceedings of the National Academy of Sciences of the United States of America,* 83**,** 2974-2978.

GIANNINI, S. 2004. *The Amyloidogenesis of Cystatins.* PhD Thesis, University of Sheffield.

GIASSON, B. I., FORMAN, M. S., HIGUCHI, M., GOLBE, L. I., GRAVES, C. L., KOTZBAUER, P. T., TROJANOWSKI, J. Q. & LEE, V. M. Y. 2003. Initiation and synergistic fibrillization of tau and alpha-synuclein. *Science,* 300**,** 636-640.

GLABE, C. G. 2004. Conformation-dependant antibodies target diseases of protein misfolding. *TRENDS in Biochemical Sciences,* 29**,** 542-547.

GLABE, C. G. 2008. Structural Classification of Toxic Amyloid Oligomers. *Journal of Biological Chemistry,* 283**,** 29639-29643.

GLENNER, G. G., EANES, E. D. & PAGE, D. L. 1972. RELATION OF PROPERTIES OF CONGO RED-STAINED AMYLOID FIBRILS TO BETA-CONFORMATION. *Journal of Histochemistry & Cytochemistry,* 20**,** 821-&.

GOLD, R. & SHALITIN, Y. 1975. SPECIFICITY OF PORCINE ELASTASE. *Biochimica Et Biophysica Acta,* 410**,** 421-425.

GOLDSBURY, C., KISTLER, J., AEBI, U., ARVINTE, T. & COOPER, G. J. S. 1999. Watching amyloid fibrils grow by time-lapse atomic force microscopy. *Journal of Molecular Biology,* 285**,** 33-39.

GORDON, D. J., TAPPE, R. & MEREDITH, S. C. 2002. Design and characterization of a membrane permeable N-methyl amino acid-containing peptide that inhibits A beta(1-40) fibrillogenesis. *Journal of Peptide Research,* 60**,** 37-55.

GOTTLIEB, H. E., KOTLYAR, V. & NUDELMAN, A. 1997. NMR chemical shifts of common laboratory solvents as trace impurities. *Journal of Organic Chemistry,* 62**,** 7512-7515.

GOVAERTS, C., WILLE, H., PRUSINER, S. B. & COHEN, F. E. 2004. Evidence for assembly of prions with left-handed beta 3-helices into trimers. *Proceedings of the National Academy of Sciences of the United States of America,* 101**,** 8342-8347.

GRIFFITH, J. S. 1967. SELF-REPLICATION AND SCRAPIE. *Nature,* 215**,** 1043-&.

GUDMUNDSSON, G., HALLGRIMSSON, J., JONASSON, T. A. & BJARNASON, O. 1972. Hereditary Cerebral Haemorrhage with Amyloidosis. *Brain,* 95**,** 387-404.

HAAN, J. & ROOS, R. A. C. 1992. Comparison between the Icelandic and Dutch Forms of Hereditary Cerebral Amyloid Angiopathy. *Clinical Neurology and Neurosurgery,* 94**,** S82-S83.

HAASS, C. & SELKOE, D. J. 2007. Soluble protein oligomers in neurodegeneration: lessons from the Alzheimer's amyloid beta-peptide. *Nature Reviews Molecular Cell Biology,* 8**,** 101-112.

HAMODRAKAS, S. J., LIAPPA, C. & ICONOMIDOU, V. A. 2007. Consensus prediction of amyloidogenic determinants in amyloid fibril-forming proteins. *International Journal of Biological Macromolecules,* 41**,** 295-300.

HARD, T. & LENDEL, C. 2012. Inhibition of Amyloid Formation. *Journal of Molecular Biology,* 421**,** 441-465.

HARPER, J. D., WONG, S. S., LIEBER, C. M. & LANSBURY, P. T. 1997. Observation of metastable A beta amyloid protofibrils by atomic force microscopy. *Chemistry & Biology,* 4**,** 119-125.

HILL, A. F., DESBRUSLAIS, M., JOINER, S., SIDLE, K. C. L., GOWLAND, I., COLLINGE, J., DOEY, L. J. & LANTOS, P. 1997. The same prion strain causes vCJD and BSE. *Nature,* 389**,** 448-450.

HOLMES, D. F., GRAHAM, H. K., TROTTER, J. A. & KADLER, K. E. 2001. STEM/TEM studies of collagen fibril assembly. *Micron,* 32**,** 273-285.

HORTSCHANSKY, P., SCHROECKH, V., CHRISTOPEIT, T., ZANDOMENEGHI, G. & FANDRICH, M. 2005. The aggregation kinetics of Alzheimer's beta-amyloid peptide is controlled by stochastic nucleation. *Protein Sci,* 14**,** 1753-9.

HOSHINO, M., KATOU, H., HAGIHARA, Y., HASEGAWA, K., NAIKI, H. & GOTO, Y. 2002. Mapping the core of the beta(2)-microglobulin amyloid fibril by H/D exchange. *Nature Structural Biology,* 9**,** 332-336.

HOU, L. M., SHAO, H. Y., ZHANG, Y. B., LI, H., MENON, N. K., NEUHAUS, E. B., BREWER, J. M., BYEON, I. J. L., RAY, D. G., VITEK, M. P., IWASHITA, T., MAKULA, R. A., PRZYBYLA, A. B. & ZAGORSKI, M. G. 2004. Solution NMR studies of the A beta(1-40) and A beta(1-42) peptides establish that the met35 oxidation state affects the mechanism of amyloid formation. *Journal of the American Chemical Society,* 126**,** 1992-2005.

HOUSEWEART, M. K., PENNACCHIO, L. A., VILAYTHONG, A., PETERS, C., NOEBELS, J. L. & MYERS, R. M. 2003. Cathepsin B but not cathepsins L or S contributes to the pathogenesis of Unverricht-Lundborg progressive myoclonus epilepsy (EPM1). *Journal of Neurobiology,* 56**,** 315-327.

HOYER, W., GRONWALL, C., JONSSON, A., STAHL, S. & HARD, T. 2008. Stabilization of a beta-hairpin in monomeric Alzheimer's amyloid-beta peptide inhibits amyloid formation. *Proceedings of the National Academy of Sciences of the United States of America,* 105**,** 5099-5104.

HOYER, W. & HARD, T. 2008. Interaction of Alzheimer's A beta peptide with an engineered binding protein - Thermodynamics and kinetics of coupled folding-binding. *Journal of Molecular Biology,* 378**,** 398-411.

HUBBARD, S. J. 1998. The structural aspects of limited proteolysis of native proteins. *Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology,* 1382**,** 191-206.

HUDSON, S. A., ECROYD, H., KEE, T. W. & CARVER, J. A. 2009. The thioflavin T fluorescence assay for amyloid fibril detection can be biased by the presence of exogenous compounds. *Febs Journal,* 276**,** 5960-5972.

HUFF, M., E. , BALCH, W., E. & KELLY, J., W. 2003. Pathological and functional amyloid formation orchestrated by the secretary pathway. *Current Opinion in Structural Biology,* 13**,** 674-682.

HUGHES, E., BURKE, R. M. & DOIG, A. J. 2000. Inhibition of toxicity in the beta-amyloid peptide fragment beta-(25-35) using N-methylated derivatives - A general strategy to prevent amyloid formation. *Journal of Biological Chemistry,* 275**,** 25109-25115.

JAHN, T. R., PARKER, M. J., HOMANS, S. W. & RADFORD, S. E. 2006. Amyloid formation under physiological conditions proceeds via a native-like folding intermediate. *Nat Struct Mol Biol,* 13**,** 195-201.

JANOWSKI, R., KOZAK, M., JANKOWSKA, E., GRZONKA, Z., GRUBB, A., ABRAHAMSON, M. & JASKOLSKI, M. 2001. Human cystatin C, an amyloidogenic protein, dimerizes through three-dimensional domain swapping. *Nature Structural Biology,* 8**,** 316-320.

JAPELJ, B., WALTHO, J. P. & JERALA, R. 2004. Comparison of backbone dynamics of monomeric and domain-swapped stefin A. *Proteins-Structure Function and Genetics,* 54**,** 500-512.

JARRETT, J. T. & LANSBURY JR., P. T. 1993. Seeding "One-Dimensional Crystallization" of Amyloid: A Pathogenic Mechanism in Alzheimer's Disease and Scrapie? *Cell,* 73**,** 1055-1058.

JELINSKA, C., DAVIS, P. J., KENIG, M., ZEROVNIK, E., KOKALJ, S. J., GUNCAR, G., TURK, D., TURK, V., CLARKE, D. T., WALTHO, J. P. & STANIFORTH, R. A. 2011. Modulation of Contact Order Effects in the Two-State Folding of Stefins A and B. *Biophysical Journal,* 100**,** 2268-2274.

JENKO, S., SKARABOT, M., KENIG, M., GUNCAR, G., MUSEVIC, I., TURK, D. & ZEROVNIK, E. 2004. Different propensity to form amyloid fibrils by two homologous proteins - Human stefins A and B: Searching for an explanation. *Proteins-Structure Function and Bioinformatics,* 55**,** 417-425.

JENSSON, O., GUDMUNDSSON, G., ARNASON, A., BLONDAL, H., PETURSDOTTIR, I., THORSTEINSSON, L., GRUBB, A., LOFBERG, H., COHEN, D. & FRANGIONE, B. 1987. HEREDITARY CYSTATIN-C (GAMMA-TRACE) AMYLOID ANGIOPATHY OF THE CNS CAUSING CEREBRAL-HEMORRHAGE. *Acta Neurologica Scandinavica,* 76**,** 102-114.

JERALA, R., TRSTENJAK, M., LENARCIC, B. & TURK, V. 1988. Cloning a synthetic gene for human stefin B and its expression in E. coli. *FEBS Letters,* 239**,** 41-4.

JIMENEZ, J. L., GUIJARRO, J. L., ORLOVA, E., ZURDO, J., DOBSON, C. M., SUNDE, M. & SAIBIL, H. R. 1999. Cryo-electron microscopy structure of an SH3 amyloid fibril and model of the molecular packing. *Embo Journal,* 18**,** 815-821.

JOENSUU, T., LEHESJOKI, A.-E. & KOPRA, O. 2008. Molecular background of EPM1 Unverricht-Lundborg Disease. 49**,** 557-563.

KAMMERER, R. A., KOSTREWA, D., ZURDO, J., DETKEN, A., GARCIA-ECHEVERRIA, C., GREEN, J. D., MULLER, S. A., MEIER, B. H., WINKLER, F. K., DOBSON, C. M. & STEINMETZ, M. O. 2004. Exploring amyloid formation by a de novo design. *Proceedings of the National Academy of Sciences of the United States of America,* 101**,** 4435-4440.

KAYED, R., HEAD, E., THOMPSON, J. L., MCINTIRE, T. M., MILTON, S. C., COTMAN, C. W. & GLABE, C. G. 2003. Common structure of soluble amyloid oligomers implies common mechanism of pathogenesis. *Science,* 300**,** 486-489.

KELLY, J. W. 2000. Mechanisms of amyloidogenesis. *Nature Structural Biology,* 7**,** 824-826.

KENIG, M., BERBIC, S., KRIJESTORAC, A., KROON-ZITKO, L., TUSEK, M., POMPE-NOVAL, M. & ZEROVNIK, E. 2004. Differences in aggregation properties of three site-specific mutants of recombinant human stefin B. *Protein Science,* 13**,** 63-70.

KHETERPAL, I., WILLIAMS, A., MURPHY, C., BLEDSOE, B. & WETZEL, R. 2001. Structural features of the A beta amyloid fibril elucidated by limited proteolysis. *Biochemistry,* 40**,** 11757-11767.

KODALI, R. & WETZEL, R. 2007. Polymorphism in the intermediates and products of amyloid assembly. *Current Opinion in Structural Biology,* 17**,** 48-57.

KOKALJ, S. J., GUNCAR, G., STERN, I., MORGAN, G., RABZELJ, S., KENIG, M., STANIFORTH, R. A., WALTHO, J. P., ZEROVNIK, E. & TURK, D. 2007. Essential role of proline isomerization in stefin B tetramer formation. *Journal of Molecular Biology,* 366**,** 1569-1579.

KOKKONI, N., STOTT, K., AMIJEE, H., MASON, J. M. & DOIG, A. J. 2006. N-methylated peptide inhibitors of beta-amyloid aggregation and toxicity. Optimization of the inhibitor structure. *Biochemistry,* 45**,** 9906-9918.

KOO, E. H., LANSBURY, P. T. & KELLY, J. W. 1999. Amyloid diseases: Abnormal protein aggregation in neurodegeneration. *Proceedings of the National Academy of Sciences of the United States of America,* 96**,** 9989-9990.

KOPITAR-JERALA, N. 2006. The role of cystatins in cells of the immune system. *Febs Letters,* 580**,** 6295-6301.

LALIOTI, M. D., MIROTSOU, M., BURESI, C., PEITSCH, M. C., ROSSIER, C., OUAZZANI, R., BALDYMOULINIER, M., BOTTANI, A., MALAFOSSE, A. & ANTONARAKIS, S. E. 1997. Identification of mutations in Cystatin B, the gene responsible for the Unverricht-Lundborg type of progressive myoclonus epilepsy (EPM1). *American Journal of Human Genetics,* 60**,** 342-351.

LANG, A. E. & LOZANO, A. M. 1998. Parkinson's disease - First of two parts. *New England Journal of Medicine,* 339**,** 1044-1053.

LASHUEL, H. A., HARTLEY, D., PETRE, B. M., WALZ, T. & LANSBURY, P. T. 2002. Neurodegenerative disease - Amyloid pores from pathogenic mutations. *Nature,* 418**,** 291-291.

LASHUEL, H. A. & LANSBURY, P. T., JR. 2006. Are amyloid diseases caused by protein aggregates that mimic bacterial pore-forming toxins? *Q Rev Biophys,* 39**,** 167-201.

LENDEL, C., BERTONCINI, C. W., CREMADES, N., WAUDBY, C. A., VENDRUSCOLO, M., DOBSON, C. M., SCHENK, D., CHRISTODOULOU, J. & TOTH, G. 2009. On the Mechanism of Nonspecific Inhibitors of Protein Aggregation: Dissecting the Interactions of alpha-Synuclein with Congo Red and Lacmoid. *Biochemistry,* 48**,** 8322-8334.

LEVINE, H. 1997. Stopped-flow kinetics reveal multiple phases of thioflavin T binding to Alzheimer beta(1-40) amyloid fibrils. *Archives of Biochemistry and Biophysics,* 342**,** 306-316.

LEVY, E., LOPEZOTIN, C., GHISO, J., GELTNER, D. & FRANGIONE, B. 1989. Stroke in Icelandic Patients with Hereditary Amyloid Angiopathy Is Related to a Mutation in the Cystatin C Gene, an Inhibitor of Cysteine Proteases. *Journal of Experimental Medicine,* 169**,** 1771-1778.

LEVY, E., SASTRE, M., KUMAR, A., GALLO, G., PICCARDO, P., GHETTI, B. & TAGLIAVINI, F. 2001. Codeposition of cystatin C with amyloid-beta protein in the brain of Alzheimer disease patients. *Journal of Neuropathology and Experimental Neurology,* 60**,** 94-104.

LOPEZ DE LA PAZ, M. & SERRANO, L. 2004. Sequence determinants of amyloid fibril formation. *Proceedings of the National Academy of Science,* 101**,** 87-92.

LUCA, S., YAU, W. M., LEAPMAN, R. & TYCKO, R. 2007. Peptide conformation and supramolecular organization in amylin fibrils: Constraints from solid-state NMR. *Biochemistry,* 46**,** 13505-13522.

LUHESHI, L. M., HOYER, W., DE BARROS, T. P., HARD, I. V., BRORSSON, A. C., MACAO, B., PERSSON, C., CROWTHER, D. C., LOMAS, D. A., STAHL, S., DOBSON, C. M. & HARD, T. 2010. Sequestration of the A beta Peptide Prevents Toxicity and Promotes Degradation In Vivo. *Plos Biology,* 8.

LUHRS, T., RITTER, C., ADRIAN, M., RIEK-LOHER, D., BOHRMANN, B., DOELI, H., SCHUBERT, D. & RIEK, R. 2005. 3D structure of Alzheimer's amyloid-

beta(1-42) fibrils. *Proceedings of the National Academy of Sciences of the United States of America,* 102**,** 17342-17347.

MAJI, S. K., PERRIN, M. H., SAWAYA, M. R., JESSBERGER, S., VADODARIA, K., RISSMAN, R. A., SINGRU, P. S., NILSSON, K. P. R., SIMON, R., SCHUBERT, D., EISENBERG, D., RIVIER, J., SAWCHENKO, P., VALE, W. & RIEK, R. 2009. Functional Amyloids As Natural Storage of Peptide Hormones in Pituitary Secretory Granules. *Science,* 325**,** 328-332.

MAKIN, O. S., ATKINS, E., SIKORSKI, P., JOHANSSON, J. & SERPELL, L. C. 2005. Molecular basis for amyloid fibril formation and stability. *Proceedings of the National Academy of Sciences of the United States of America,* 102**,** 315-320.

MARGITTAI, M. & LANGEN, R. 2008. Fibrils with parallel in-register structure constitute a major class of amyloid fibrils: molecular insights from electron paramagnetic resonance spectroscopy. *Quarterly Reviews of Biophysics,* 41**,** 265-297.

MARSHALL, K. E., HICKS, M. R., WILLIAMS, T. L., HOFFMANN, S. V., RODGER, A., DAFFORN, T. R. & SERPELL, L. C. 2010. Characterizing the assembly of the Sup35 yeast prion fragment, GNNQQNY: structural changes accompany a fiber-to-crystal switch. *Biophys J,* 98**,** 330-8.

MARSHALL, K. E., MORRIS, K. L., CHARLTON, D., O'REILLY, N., LEWIS, L., WALDEN, H. & SERPELL, L. C. 2011. Hydrophobic, Aromatic, and Electrostatic Interactions Play a Central Role in Amyloid Fibril Formation and Stability. *Biochemistry,* 50**,** 2061-2071.

MARTIN, J. R., CRAVEN, C. J., JERALA, R., KROONZITKO, L., ZEROVNIK, E., TURK, V. & WALTHO, J. P. 1995. THE 3-DIMENSIONAL SOLUTION STRUCTURE OF HUMAN STEFIN-A. *Journal of Molecular Biology,* 246**,** 331-343.

MAURER-STROH, S., DEBULPAEP, M., KUEMMERER, N., DE LA PAZ, M. L., MARTINS, I. C., REUMERS, J., MORRIS, K. L., COPLAND, A., SERPELL, L., SERRANO, L., SCHYMKOWITZ, J. W. H. & ROUSSEAU, F. 2010. Exploring the sequence determinants of amyloid structure using position-specific scoring matrices. *Nat Meth,* 7**,** 237-242.

MAURER-STROH, S., DEBULPAEP, M., KUEMMERER, N., LOPEZ DE LA PAZ, M., MARTINS, I. C., REUMERS, J., COPLAND, A., SERPELL, L., SERRANO, L., ROUSSEAU, F. & SCHYMKOWITZ, J. 2009. An exploration of the sequence determinants of amyloid formation and the development of the amyloid specific prediction algorithm WALTZ. *As yet unpublished, probably Nature Methods***,** 54.

MCDONALD, M., BOX, H., BIAN, W., KENDALL, A., TYCKO, R. & STUBBS, G. 2012. Fiber Diffraction Data Indicate a Hollow Core for the Alzheimer's A beta 3-Fold Symmetric Fibril. *Journal of Molecular Biology,* 423**,** 454-461.

MCKOY, A. F., CHEN, J., SCHUPBACH, T. & HECHT, M. H. 2012. A Novel Inhibitor of Amyloid beta (A beta) Peptide Aggregation FROM HIGH THROUGHPUT SCREENING TO EFFICACY IN AN ANIMAL MODEL OF ALZHEIMER DISEASE. *Journal of Biological Chemistry,* 287**,** 38992-39000.

MIAKE, H., MIZUSAWA, H., IWATSUBO, T. & HASEGAWA, M. 2002. Biochemical characterization of the core structure of alpha-synuclein filaments. *Journal of Biological Chemistry,* 277**,** 19213-19219.

MILANESI, L., SHEYNIS, T., XUE, W. F., ORLOVA, E. V., HELLEWELL, A. L., JELINEK, R., HEWITT, E. W., RADFORD, S. E. & SAIBIL, H. R. 2012. Direct three-dimensional visualization of membrane disruption by amyloid fibrils. *Proceedings of the National Academy of Sciences of the United States of America,* 109**,** 20455-20460.

MIROY, G. J., LAI, Z. H., LASHUEL, H. A., PETERSON, S. A., STRANG, C. & KELLY, J. W. 1996. Inhibiting transthyretin amyloid fibril formation via protein stabilization. *Proceedings of the National Academy of Sciences of the United States of America,* 93**,** 15051-15056.

MORGAN, G. 2006. *Structural Studies on the Cystatin Amyloid Fibril.* PhD, University of Sheffield.

MORGAN, G. J., GIANNINI, S., HOUNSLOW, A. M., CRAVEN, C. J., ZEROVNIK, E., TURK, V., WALTHO, J. P. & STANIFORTH, R. A. 2008. Exclusion of the Native [alpha]-Helix from the Amyloid Fibrils of a Mixed [alpha]/[beta] Protein. *Journal of Molecular Biology,* 375**,** 487-498.

MULLER, S. A., GOLDIE, K. N., BURKI, R., HARING, R. & ENGEL, A. 1992. FACTORS INFLUENCING THE PRECISION OF QUANTITATIVE SCANNING-TRANSMISSION ELECTRON-MICROSCOPY. *Ultramicroscopy,* 46**,** 317-334.

MURPHY, R. M. 2007. Kinetics of amyloid formation and membrane interaction with amyloidogenic proteins. *Biochimica Et Biophysica Acta-Biomembranes,* 1768**,** 1923-1934.

MYERS, S. L., THOMSON, N. H., RADFORD, S. E. & ASHCROFT, A. E. 2006. Investigating the structural properties of amyloid-like fibrils formed in vitro from beta 2-microglobulin using limited proteolysis and electrospray ionisation mass spectrometry. *Rapid Communications in Mass Spectrometry,* 20**,** 1628-1636.

NAIKI, H. & GEJYO, F. 1999. [20] Kinetic analysis of amyloid fibril formation. *In:* RONALD, W. (ed.) *Methods in Enzymology.* Academic Press.

NAIKI, H., HIGUCHI, K., HOSOKAWA, M. & TAKEDA, T. 1989. Fluorometric determination of amyloid fibrils in vitro using the fluorescent dye, thioflavine T. *Analytical Biochemistry,* 177**,** 244-249.

NEEDLEMAN, S. B. & WUNSCH, C. D. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol,* 48**,** 443-53.

NELSON, R. & EISENBERG, D. 2006. Structural models of amyloid-like fibrils. *In:* KAJAVA, A., SQUIRE, J. M. & PARRY, D. A. D. (eds.) *Fibrous Proteins: Amyloids, Prions and Beta Proteins.* San Diego: Elsevier Academic Press Inc.

NELSON, R., SAWAYA, M. R., BALBIRNIE, M., MADSEN, A. O., RIEKEL, C., GROTHE, R. & EISENBERG, D. 2005. Structure of the cross-beta spine of amyloid-like fibrils. *Nature,* 435**,** 773-778.

NILSBERTH, C., WESTLIND-DANIELSSON, A., ECKMAN, C. B., CONDRON, M. M., AXELMAN, K., FORSELL, C., STENH, C., LUTHMAN, J., TEPLOW, D. B., YOUNKIN, S. G., NASLUND, J. & LANNFELT, L. 2001. The 'Arctic' APP mutation (E693G) causes Alzheimer's disease by enhanced Abeta protofibril formation. *Nat Neurosci,* 4**,** 887-93.

NORD, K., GUNNERIUSSON, E., RINGDAHL, J., STAHL, S., UHLEN, M. & NYGREN, P. A. 1997. Binding proteins selected from combinatorial libraries of an alpha-helical bacterial receptor domain. *Nature Biotechnology,* 15**,** 772-777.

NORIO, R. & KOSKINIEMI, M. 1979. PROGRESSIVE MYOCLONUS EPILEPSY - GENETIC AND NOSOLOGICAL ASPECTS WITH SPECIAL REFERENCE TO 107 FINNISH PATIENTS. *Clinical Genetics,* 15**,** 382-398.

NOWAK, R. J., CUNY, G. D., CHOI, S., LANSBURY, P. T. & RAY, S. S. 2010. Improving binding specificity of pharmacological chaperones that target mutant superoxide dismutase-1 linked to familial amyotrophic lateral sclerosis using computational methods. *J Med Chem,* 53**,** 2709-18.

OLOFSSON, A., SAUER-ERIKSSON, A. E. & OHMAN, A. 2006. The solvent protection of alzheimer amyloid-beta-(1-42) fibrils as determined by solution NMR spectroscopy. *J Biol Chem,* 281**,** 477-83.

ONO, K., HASEGAWA, K., NAIKI, H. & YAMADA, M. 2004. Curcumin has potent anti-amyloidogenic effects for Alzheimer's β-amyloid fibrils in vitro. *Journal of Neuroscience Research,* 75**,** 742-750.

OSTNER, G., LINDSTROM, V., CHRISTENSEN, P. H., KOZAK, M., ABRAHAMSON, M. & GRUBB, A. 2013. Stabilization, Characterization, and Selective Removal of Cystatin C Amyloid Oligomers. *Journal of Biological Chemistry,* 288**,** 16438-16450.

PALMER, A. M. 2011. Neuroprotective therapeutics for Alzheimer's disease: progress and prospects. *Trends in Pharmacological Sciences,* 32**,** 141-147.

PALSDOTTIR, A., ABRAHAMSON, M., THORSTEINSSON, L., ARNASON, A., OLAFSSON, I., GRUBB, A. & JENSSON, O. 1988. MUTATION IN CYSTATIN-C GENE CAUSES HEREDITARY BRAIN HEMORRHAGE. *Lancet,* 2**,** 603-604.

PANDE, V. S. 2004. A universal TANGO? *Nat Biotech,* 22**,** 1240-1241.

PANZA, F., FRISARDI, V., SOLFRIZZI, V., IMBIMBO, B. P., LOGROSCINO, G., SANTAMATO, A., GRECO, A., SERIPA, D. & PILOTTO, A. 2012. Immunotherapy

for Alzheimer's disease: from anti-beta-amyloid to tau-based immunization strategies. *Immunotherapy,* 4**,** 213-238.

PARAMORE, R. 2010. *Studies on the Mechanism of Amyloid Formation by Cystatin B.* PhD PhD, University of Sheffield.

PARAMORE, R., MORGAN, G. J., DAVIS, P. J., SHARMA, C.-A., HOUNSLOW, A., TALER-VERCIC, A., ZEROVNIK, E., WALTHO, J. P., CLIFF, M. J. & STANIFORTH, R. A. 2012. Mapping local structural perturbations in the native state of stefin B (cystatin B) under amyloid forming conditions. *Frontiers in molecular neuroscience,* 5**,** 94.

PARAVASTU, A. K., LEAPMAN, R. D., YAU, W. M. & TYCKO, R. 2008. Molecular structural basis for polymorphism in Alzheimer's beta-amyloid fibrils. *Proceedings of the National Academy of Sciences of the United States of America,* 105**,** 18349-18354.

PENNACCHIO, L. A., BOULEY, D. M., HIGGINS, K. M., SCOTT, M. P., NOEBELS, J. L. & MYERS, R. M. 1998. Progressive ataxia, myoclonic epilepsy and cerebellar apoptosis in cystatin B-deficient mice. *Nature Genetics,* 20**,** 251-258.

PERUTZ, M. F. 1999. Glutamine repeats and neurodegenerative diseases: molecular aspects. *Trends in Biochemical Sciences,* 24**,** 58-63.

PERUTZ, M. F., JOHNSON, T., SUZUKI, M. & FINCH, J. T. 1994. GLUTAMINE REPEATS AS POLAR ZIPPERS - THEIR POSSIBLE ROLE IN INHERITED NEURODEGENERATIVE DISEASES. *Proceedings of the National Academy of Sciences of the United States of America,* 91**,** 5355-5358.

PERVUSHIN, K., RIEK, R., WIDER, G. & WUTHRICH, K. 1997. Attenuated T-2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proceedings of the National Academy of Sciences of the United States of America,* 94**,** 12366-12371.

PETKOVA, A. T., ISHII, Y., BALBACH, J. J., ANTZUTKIN, O. N., LEAPMAN, R. D., DELAGLIO, F. & TYCKO, R. 2002. A structural model for Alzheimer's beta-amyloid fibrils based on experimental constraints from solid state NMR. *Proceedings of the National Academy of Sciences of the United States of America,* 99**,** 16742-16747.

PETKOVA, A. T., LEAPMAN, R. D., GUO, Z. H., YAU, W. M., MATTSON, M. P. & TYCKO, R. 2005. Self-propagating, molecular-level polymorphism in Alzheimer's beta-amyloid fibrils. *Science,* 307**,** 262-265.

PETKOVA, A. T., YAU, W. M. & TYCKO, R. 2006. Experimental constraints on quaternary structure in Alzheimer's beta-amyloid fibrils. *Biochemistry,* 45**,** 498-512.

PORAT, Y., ABRAMOWITZ, A. & GAZIT, E. 2006. Inhibition of amyloid fibril formation by polyphenols: Structural similarity and aromatic interactions as a common inhibition mechanism. *Chemical Biology & Drug Design,* 67**,** 27-37.

PRLIC, A., BLIVEN, S., ROSE, P. W., BLUHM, W. F., BIZON, C., GODZIK, A. & BOURNE, P. E. 2010. Pre-calculated protein structure alignments at the RCSB PDB website. *Bioinformatics,* 26**,** 2983-5.

PRUSINER, S. B. 1982. NOVEL PROTEINACEOUS INFECTIOUS PARTICLES CAUSE SCRAPIE. *Science,* 216**,** 136-144.

PUL, R., DODEL, R. & STANGEL, M. 2011. Antibody-based therapy in Alzheimer's disease. *Expert Opinion on Biological Therapy,* 11**,** 343-357.

RABZELJ, S., TURK, V. & ZEROVNIK, E. 2005. In vitro study of stability and amyloid-fibril formation of two mutants of human stefin B (cystatin B) occurring in patients with EPM1. *Protein Science,* 14**,** 2713-2722.

RAUTER, A. P., DIAS, C., MARTINS, A., BRANCO, I., NENG, N. R., NOGUEIRA, J. M., GOULART, M., SILVA, F. V. M., JUSTINO, J., TREVITT, C. & WALTHO, J. P. 2012. Non-toxic Salvia sclareoides Brot. extracts as a source of functional food ingredients: Phenolic profile, antioxidant activity and prion binding properties. *Food Chemistry,* 132**,** 1930-1935.

RAY, S. S., NOWAK, R. J., BROWN, R. H. & LANSBURY, P. T. 2005. Small-molecule-mediated stabilization of familial amyotrophic lateral sclerosis-linked superoxide dismutase mutants against unfolding and aggregation. *Proceedings of the National Academy of Sciences of the United States of America,* 102**,** 3639-3644.

REUMERS, J., MAURER-STROH, S., SCHYMKOWITZ, J. & ROUSSEAU, F. 2009. Protein sequences encode safeguards against aggregation. *Human Mutation,* 30**,** 431-437.

REVESZ, T., GHISO, J., LASHLEY, T., PLANT, G., ROSTAGNO, A., FRANGIONE, B. & HOLTON, J. L. 2003. Cerebral Amyloid Angiopathies: A Pathologic, Biochemical and Genetic View. *Journal of Neuropathology and Experimental Neurology,* 62**,** 885-898.

RITTER, C., MADDELEIN, M.-L., SIEMER, A. B., LUHRS, T., ERNST, M., MEIER, B. H., SAUPE, S. J. & RIEK, R. 2005. Correlation of structural elements and infectivity of the HET-s prion. *Nature,* 435**,** 844-848.

SAMBASHIVAN, S., LIU, Y. S., SAWAYA, M. R., GINGERY, M. & EISENBERG, D. 2005. Amyloid-like fibrils of ribonuclease A with three-dimensional domain-swapped and native-like structure. *Nature,* 437**,** 266-269.

SAMBROOK, J., FRITSCH, E. F. & MANIATIS, T. 1989. *Molecular Cloning: A Laboratory Manual,* Cold Spring Harbour Laboratory Press.

SANCHEZ, L., MADURGA, S., PUKALA, T., VILASECA, M., LOPEZ-IGLESIAS, C., ROBINSON, C. V., GIRALT, E. & CARULLA, N. 2011. A beta 40 and A beta 42 Amyloid Fibrils Exhibit Distinct Molecular Recycling Properties. *Journal of the American Chemical Society,* 133**,** 6505-6508.

SANDBERG, A., LUHESHI, L. M., SOLLVANDER, S., DE BARROS, T. P., MACAO, B., KNOWLES, T. P. J., BIVERSTAL, H., LENDEL, C., EKHOLM-PETTERSON, F., DUBNOVITSKY, A., LANNFELT, L., DOBSON, C. M. & HARD, T. 2010. Stabilization of neurotoxic Alzheimer amyloid-beta oligomers by protein engineering. *Proceedings of the National Academy of Sciences of the United States of America,* 107**,** 15595-15600.

SCHAGGER, H. 2006. Tricine-SDS-PAGE. *Nature Protocols,* 1**,** 16-22.

SELKOE, D. J. & SCHENK, D. 2003. Alzheimer's disease: Molecular understanding predicts amyloid-based therapeutics. *Annual Review of Pharmacology and Toxicology,* 43**,** 545-584.

SERIO, T. R., CASHIKAR, A. G., KOWAL, A. S., SAWICKI, G. J., MOSLEHI, J. J., SERPELL, L., ARNSDORF, M. F. & LINDQUIST, S. L. 2000. Nucleated conformational conversion and the replication of conformational information by a prion determinant. *Science,* 289**,** 1317-1321.

SERPELL, L. C. 2000. Alzheimer's amyloid fibrils: structure and assembly. *Biochimica Et Biophysica Acta-Molecular Basis of Disease,* 1502**,** 16-30.

SERPELL, L. C., SUNDE, M., BENSON, M. D., TENNENT, G. A., PEPYS, M. B. & FRASER, P. E. 2000. The protofilament substructure of amyloid fibrils. *Journal of Molecular Biology,* 300**,** 1033-1039.

SHARMA, C.-A. 2009. *Elucidating the Topology of Cystatin B Amyloid Fibrils.* PhD, University of Sheffield.

SHEWMAKER, F., WICKNER, R. B. & TYCKO, R. 2006. Amyloid of the prion domain of Sup35p has an in-register parallel beta-sheet structure. *Proceedings of the National Academy of Sciences of the United States of America,* 103**,** 19754-19759.

SIPE, J. D., BENSON, M. D., BUXBAUM, J. N., IKEDA, S., MERLINI, G., SARAIVA, M. J. M. & WESTERMARK, P. 2010. Amyloid fibril protein nomenclature: 2010 recommendations from the nomenclature committee of the International Society of Amyloidosis. *Amyloid-Journal of Protein Folding Disorders,* 17**,** 101-104.

SIPE, J. D. & COHEN, A. S. 2000. Review: history of the amyloid fibril. *J Struct Biol,* 130**,** 88-98.

SMITH, A. 2003. Protein misfolding. *Nature,* 426**,** 883-883.

SOTO, C., SIGURDSSON, E. M., MORELLI, L., KUMAR, R. A., CASTANO, E. M. & FRANGIONE, B. 1998. beta-sheet breaker peptides inhibit fibrillogenesis in a rat brain model of amyloidosis: Implications for Alzheimer's therapy. *Nature Medicine,* 4**,** 822-826.

STANIFORTH, R. A., GIANNINI, S., HIGGINS, L. D., CONROY, M. J., HOUNSLOW, A. M., JERALA, R., CRAVEN, C. J. & WALTHO, J. P. 2001. Three-

dimensional domain swapping in the folded and molten-globule states of cystatins, an amyloid-forming structural superfamily. *Embo Journal,* 20**,** 4774-4781.

STEFANI, M. 2010. Biochemical and biophysical features of both oligomer/fibril and cell membrane in amyloid cytotoxicity. *FEBS J,* 277**,** 4602-13.

STINE, W. B., JUNGBAUER, L., YU, C. & LADU, M. J. 2011. Preparing synthetic Abeta in different aggregation states. *Methods Mol Biol,* 670**,** 13-32.

STUBBS, G. 1999. Developments in fiber diffraction. *Curr Opin Struct Biol,* 9**,** 615-9.

STUBBS, M. T., LABER, B., BODE, W., HUBER, R., JERALA, R., LENARCIC, B. & TURK, V. 1990. The Refined 2.4a X-Ray Crystal-Structure of Recombinant Human Stefin-B in Complex with the Cysteine Proteinase Papain - a Novel Type of Proteinase-Inhibitor Interaction. *Embo Journal,* 9**,** 1939-1947.

TARTAGLIA, G. G., PAWAR, A. P., CAMPIONI, S., DOBSON, C. M., CHITI, F. & VENDRUSCOLO, M. 2008. Prediction of Aggregation-Prone Regions in Structured Proteins. *Journal of Molecular Biology,* 380**,** 425-436.

TAUPIN, P., RAY, J., FISCHER, W. H., SUHR, S. T., HAKANSSON, K., GRUBB, A. & GAGE, F. H. 2000. FGF-2-responsive neural stem cell proliferation requires CCg, a novel autocrine/paracrine cofactor. *Neuron,* 28**,** 385-397.

THOMPSON, M. J., SIEVERS, S. A., KARANICOLAS, J., IVANOVA, M. I., BAKER, D. & EISENBERG, D. 2006. The 3D profile method for identifying fibril-forming segments of proteins.

THOMPSON, R. C. & BLOUT, E. R. 1973. RESTRICTIONS ON BINDING OF PROLINE-CONTAINING PEPTIDES TO ELASTASE. *Biochemistry,* 12**,** 51-57.

TJERNBERG, L. O., NASLUND, J., LINDQVIST, F., JOHANSSON, J., KARLSTROM, A. R., THYBERG, J., TERENIUS, L. & NORDSTEDT, C. 1996. Arrest of beta-amyloid fibril formation by a pentapeptide ligand. *Journal of Biological Chemistry,* 271**,** 8545-8548.

TURK, V. & BODE, W. 1991. The Cystatins - Protein Inhibitors of Cysteine Proteinases. *Febs Letters,* 285**,** 213-219.

TURK, V., TURK, B. & TURK, D. 2001. Lysosomal cysteine proteases: facts and opportunities. *Embo Journal,* 20**,** 4629-4633.

TYCKO, R. 2011. Solid-State NMR Studies of Amyloid Fibril Structure. *In:* LEONE, S. R., CREMER, P. S., GROVES, J. T. & JOHNSON, M. A. (eds.) *Annual Review of Physical Chemistry, Vol 62.* Palo Alto: Annual Reviews.

VINTERS, H. V., NISHIMURA, G. S., SECOR, D. L. & PARDRIDGE, W. M. 1990. Immunoreactive-A4 and Gamma-Trace Peptide Colocalization in Amyloidotic Arteriolar Lesions in Brains of Patients with Alzheimers Disease. *Am. J. Pathol.,* 137**,** 233-240.

WALL, J. S., SIMON, M. N., LIN, B. Y. & VINOGRADOV, S. N. 2008. Mass mapping of large globin complexes by scanning transmission electron microscopy. *In:* POOLE, R. K. (ed.) *Globins and Other Nitric Oxide-Reactive Proteins, Pt A.* San Diego: Elsevier Academic Press Inc.

WALSH, D. M., LOMAKIN, A., BENEDEK, G. B., CONDRON, M. M. & TEPLOW, D. B. 1997. Amyloid beta-protein fibrillogenesis - Detection of a protofibrillar intermediate. *Journal of Biological Chemistry,* 272**,** 22364-22372.

WALSH, P., SIMONETTI, K. & SHARPE, S. 2009. Core structure of amyloid fibrils formed by residues 106-126 of the human prion protein. *Structure,* 17**,** 417-26.

WALTHO, J. P., MARTIN, J. R., CRAVEN, C. J., JERALA, R., ZEROVNIK, E. & TURK, V. 1995. THE REFINED SOLUTION STRUCTURE OF HUMAN STEFIN-A. *Journal of Cellular Biochemistry***,** 40-40.

WANKER, E. E., SCHERZINGER, E., HEISER, V., SITTLER, A., EICKHOFF, H. & LEHRACH, H. 1999. Membrane filter assay for detection of amyloid-like polyglutamine-containing protein aggregates. *Amyloid, Prions, and Other Protein Aggregates,* 309**,** 375-386.

WASMER, C., LANGE, A., VAN MELCKEBEKE, H., SIEMER, A. B., RIEK, R. & MEIER, B. H. 2008. Amyloid fibrils of the HET-s(218-289) prion form a beta solenoid with a triangular hydrophobic core. *Science,* 319**,** 1523-1526.

WHITE, H. E., HODGKINSON, J. L., JAHN, T. R., COHEN-KRAUSZ, S., GOSAL, W. S., MULLER, S., ORLOVA, E. V., RADFORD, S. E. & SAIBIL, H. R. 2009. Globular tetramers of beta(2)-microglobulin assemble into elaborate amyloid fibrils. *J Mol Biol,* 389**,** 48-57.

WOLFE, L. S., CALABRESE, M. F., NATH, A., BLAHO, D. V., MIRANKER, A. D. & XIONG, Y. 2010. Protein-induced photophysical changes to the amyloid indicator dye thioflavin T. *Proceedings of the National Academy of Sciences of the United States of America,* 107**,** 16863-16868.

WOODS, L. A., PLATT, G. W., HELLEWELL, A. L., HEWITT, E. W., HOMANS, S. W., ASHCROFT, A. E. & RADFORD, S. E. 2011. Ligand binding to distinct states diverts aggregation of an amyloid-forming protein. *Nature Chemical Biology,* 7**,** 730-739.

XUE, W.-F., HELLEWELL, A. L., GOSAL, W. S., HOMANS, S. W., HEWITT, E. W. & RADFORD, S. E. 2009. Fibril Fragmentation Enhances Amyloid Cytotoxicity. *Journal of Biological Chemistry,* 284**,** 34272-34282.

YAMAGUCHI, K. I., KATOU, H., HOSHINO, M., HASEGAWA, K., NAIKI, H. & GOTO, Y. 2004. Core and heterogeneity of beta(2)-microglobulin amyloid fibrils as revealed by H/D exchange. *Journal of Molecular Biology,* 338**,** 559-571.

YAMAMOTO, T. 2010. Molecular Dynamics of Reversible and Irreversible Melting in Chain-Folded Crystals of Short Polyethylene-like Polymer. *Macromolecules,* 43**,** 9384-9393.

YE, Y. & GODZIK, A. 2003. Flexible structure alignment by chaining aligned fragment pairs allowing twists. *Bioinformatics,* 19 Suppl 2**,** ii246-55.

ZAHN, R., LIU, A. Z., LUHRS, T., RIEK, R., VON SCHROETTER, C., GARCIA, F. L., BILLETER, M., CALZOLAI, L., WIDER, G. & WUTHRICH, K. 2000. NMR solution structure of the human prion protein. *Proceedings of the National Academy of Sciences of the United States of America,* 97**,** 145-150.

ZANDOMENEGHI, G., KREBS, M. R. H., MCCAMMON, M. G. & FANDRICH, M. 2004. FTIR reveals structural differences between native beta-sheet proteins and amyloid fibrils. *Protein Science,* 13**,** 3314-3321.

ZEROVNIK, E., POMPE-NOVAK, M., SKARABOT, M., RAVNIKAR, M., MUSEVIC, I. & TURK, V. 2002a. Human stefin B readily forms amyloid fibrils in vitro. *Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology*, 1594**,** 1-5.

ZEROVNIK, E., ZAVASNIK-BERGANT, V., KOPITAR-JERALA, N., POMPE-NOVAK, M., SKARABOT, M., GOLDIE, K., RAVNIKAR, M., MUSEVIC, I. & TURK, V. 2002b. Amyloid fibril formation by human stefin B in vitro: Immunogold labelling and comparison to stefin A. *Biological Chemistry*, 383**,** 859-863.

ZHANG, Z., CHEN, H. & LAI, L. 2007. Identification of amyloid fibril-forming segments based on structure and residue-based statistical potential.

ZHU, M., RAJAMANI, S., KAYLOR, J., HAN, S., ZHOU, F. & FINK, A. L. 2004. The Flavonoid Baicalein Inhibits Fibrillation of α-Synuclein and Disaggregates Existing Fibrils. *Journal of Biological Chemistry*, 279**,** 26846-26857.

Appendix 1: X-ray Fibre Diffraction

A1.1 Introduction

Partially aligned cystatin B fibre has previously been observed to diffract at 4.7Å on the meridional plane and at 10 Å on the equatorial plane, (Jenko et al., 2004) which is indicative of cross-β structure (Geddes *et al.*, 1968) and is diagnostic of amyloid fibres (Blake and Serpell, 1996). The meridional reflections are produced by the repeated interstrand hydrogen bond spacing whilst the orthogonal equatorial reflections are generated by intersheet interactions. A completely aligned cystatin B fibre preparation would allow determination of further cystatin B structural information, such as helical pitch/helical repeating unit (115.5 Å/ 24 β-strands in transthyretin (Blake and Serpell, 1996)). Fibre diffraction can only image molecular details in exceptionally favourable cases (Chandrasekaran and Stubbs, 2006, Stubbs, 1999) but can often provide accurate overall shapes and dimensions (McDonald *et al.*, 2012).

A1.2 Methodology, Results and Discussion

To generate additional fibre structural constraints, several attempts to align cystatin B fibres have been made in collaboration with Louise Serpell. However, each attempt failed to result in suitably aligned fibres, perhaps due to the substantial lateral association found in cystatin B fibre preparations. Initially the fibres were prepared in the standard pH 4.7 (15 mM NaAc, 10% TFE, 150 mM NaCl) conditions (described in 2.6.1), before centrifugation of the insoluble fibres and 2 wash steps in d.H_2O, ready for fibre alignment and diffraction. The first fibre preparation produced short fibres, caused by mechanical shearing in the washing steps which subsequently failed to align. A second attempt, with gentler handling yielded less shearing but equally did not align.

The previous partial fibre alignment was achieved by fibre formation for 4 months in a 2.3 tesla magnetic field, thus fibrilisation within a 20 tesla field was attempted by incubating a fibre preparation at 21°C centrally underneath an unshielded 800 MHz NMR spectrometer to maximise the effective magnetic flux. After 9 months, fibre formation had not occurred, therefore, a 20% (v/v) fibre 'seed' preparation generated from 10 minute sonication of mature fibres was added to initiate fibre elongation, which requires future follow-up study. Fibre formation with increased sodium chloride (0.5-1.0 M) may in the future yield fibres with lower lateral association due to the

interruption of ionic interactions between the fibres. Low salt conditions (10 mM sodium phosphate buffer) were tested but did not reduce lateral association. In addition, cystatin B fibre formed from a truncated protein with all regions extraneous to the fibre core removed may identify if association is mediated by the fibre core or unfolded residues. Further attempts to align fibres should be followed up to produce a better model of cystatin B fibre.

Appendix 2 Cystatin B Truncate Mutant Study

A2.1 Introduction

The sequence based amyloid prediction programs inspired the study of a range of truncate mutant proteins and peptides based on human cystatin B. These truncate mutants were to be used in determining whether the short amino acid sequences drive amyloid fibre formation in a full length folded protein and ascertain whether the formed amyloid fibres retain the same structure. Three mutant proteins and two peptides were designed from the amyloid hotspot prediction software results and disease implicated mutants.

An alpha-helix truncate E31 cystatin B mutant with a 3C protease site insert after residue 35 (α-3C-β) (see figure A2.6.1) was designed to provide insight into helix functionality in amyloidosis, due to the specific cleavable nature of the inserted 3C-protease site thereby allowing accurate removal of the α-helix. The D61X truncate mutant has β-strands 4 and 5 removed to elucidate the function of these strands in amyloid formation, study the amyloid formation of strands 2 and 3 and analyse fibre polymorphism. R68X is similar to D61X and has β-strands 5 and half of β-strand 4 removed to elucidate the function of these strands in a similar manner to D61X. R68X has the additional interest as this mutant, and others have been found in sufferers of myoclonus epilepsy of type 1 (EPM1), a rare progressive and degenerate epilepsy (Rabzelj *et al.*, 2005). Two peptides with the sequences, KAVSFKSQVVAGTNYFIKV and GTNYFIKV, which are cystatin B residues 39-59 and 50-57 were designed to provide insight into the main amyloidogenic and domain swapping β-strand 2 and 3 and the highly amyloidogenic β-strand 3.

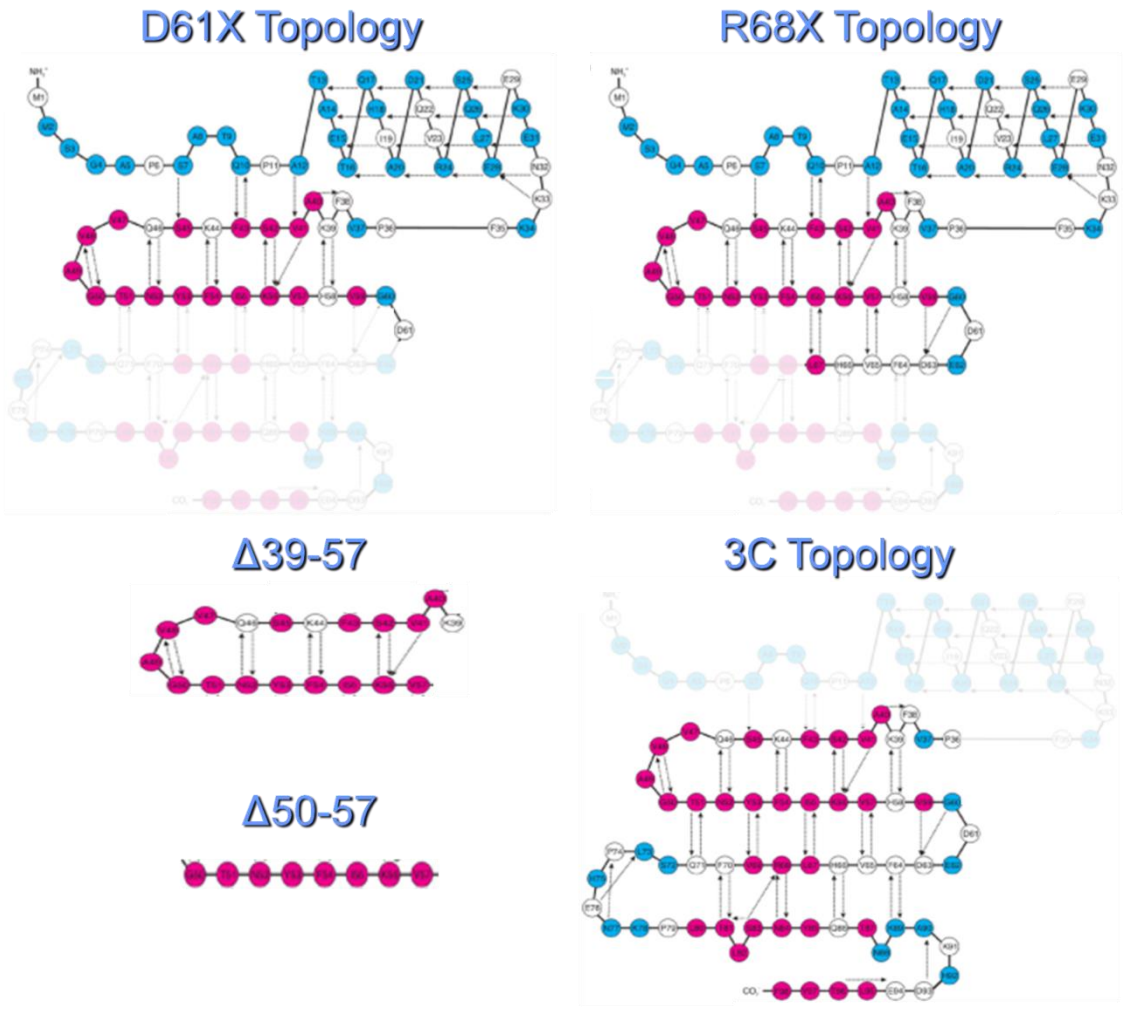


Figure A.2.6.1 Cystatin B Truncate Mutant Topology

The truncated region of these mutants is masked to indicate the remaining structures and the extant of the truncation. The topology diagram colouring is from H/D exchange data of cystatin B amyloid fibre (Morgan *et al.*, 2008) with red indicating protection, blue no protection and white undetermined.

A2.2 Materials and Methods

The alpha-helix truncate E31 cystatin B with a 3C protease site insert after residue 35 (α-3C-β) was produced by nucleic acid de novo synthesis and sub-cloning of the gene into a pET-11a vector by GeneArt. D61X and R68X truncate E31 cystatin B mutants were produced by stop codon insertion via site-directed mutagenesis of the WT in pET-11a. Summer project student, Steve Harbourne, produced the D61X and R61X mutants by site-directed mutagenesis addition of two stop codons into the (C3S), E31 WT cystatin B sequence. The two peptides with the sequences, KAVSFKSQVVAGTNYFIKV and GTNYFIKV were *de novo* synthesised by Biotec (USA) and Arthur Moir (University of Sheffield).

Purification

Several attempts to produce the three truncate proteins in the soluble cell lysate failed, including low temperature (25°C) expression, use of *E.coli* BL21 (DE3) Gold PLys-S cells and shorter induction times to reduce the cellular protein concentration and avoid aggregation. The most successful purification strategy used a purification method similar to that described in section 2.5, modified by dissolving the high yield inclusion bodies in 4-5 M urea in which the anion exchange was performed followed by refolding and urea removal. See table A2.6.2 for buffer conditions used and the physic-chemical characteristics table in section 2.5.

The range of buffers required for dissolving and purifying the truncate mutants was based on the predicted protein pI value and a urea concentration dissolution assay probed by centrifugation and SDS-PAGE. Each specific buffer was used throughout the purification steps with appropriate addition of sodium chloride when required. Any urea containing buffer was freshly made on or closely before the day of use to reduce the decomposition of urea into uric acid, which amidates protein C-terminal carboxylate groups. During truncate purification, the protein was exchanged into new solution at least every two days.

Cystatin B Protein	pH	Buffer (10mM)	Hydrogen Donor	Hydrogen Acceptor	Additives
WT E31	6.0	Sodium Phosphate	NaH_2PO_4	Na_2HPO_4	N/A
α -3C-β	6.0	Sodium Phosphate	NaH_2PO_4	Na_2HPO_4	4M urea, 1mM EDTA
D61X	7.0	Sodium Phosphate	NaH_2PO_4	Na_2HPO_4	5M urea, 1mM EDTA
R68X	5.0	Sodium Acetate	AcOH	CH_3COONa	5M urea, 1mM EDTA

Table A2.6.2. Buffer Table for WT and Truncate Mutant Purification

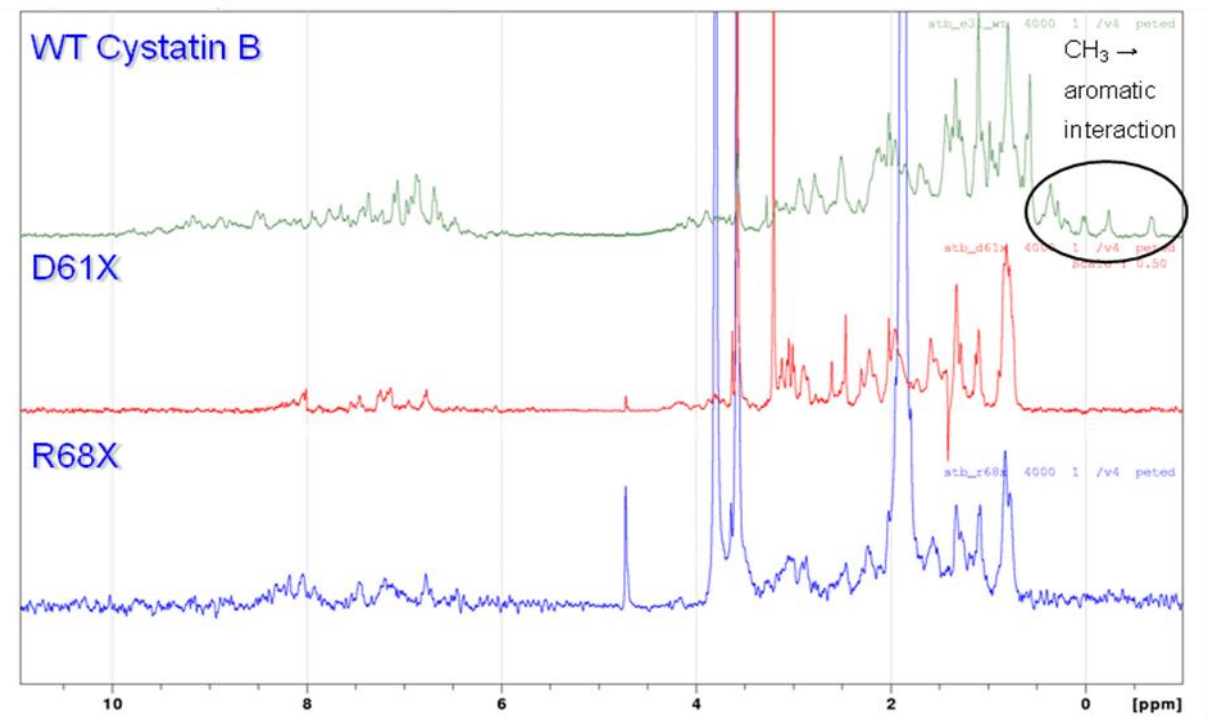
The modified protocol starts by resuspending the insoluble cell lysis pellet in 4-5 M buffered urea solution and sonication repeated in identical manner to cell lysis. The resuspension was then dialysed overnight against 1 litre of appropriate urea containing buffer at 4°C in suitable 6-8 000 or 3 000 Da MWCO Spectra/Por dialysis tubing (Spectrum Laboratories) to remove excess DNA fragments which interfere with ion-exchange chromatography. Following dialysis, the resuspension was centrifuged at 24 500 rpm at 4°C for 15 minutes to remove cell debris and insoluble protein. Subsequent anion-exchange and refolding of the protein in a urea-free equilibrated size-exclusion chromatography proved to be the most successful procedure. Rapid dilution refolding was attempted but failed to produce sufficient concentrations of folded protein and proved to be rather slow when concentrating the protein from several hundred millilitres of protein.

α-3C-β additionally, once purified and soluble in biological buffer requires proteolysis with the highly specific 3C-protease to remove the α-helix, this however has not been attempted due to poor yields in the first purification step.

A3.3 Results

The truncate mutants proved difficult to work with due to intrinsic instability and were readily degraded, with a usable lifetime of less than 2 weeks. The mutants have been purified to a low yield (~1 mg per litre of growth medium). A fibrilisation trial was unsuccessful. The most likely reason for this failure was the use of an old stock bottle of accelerant trifluoroethanol (TFE), which potentially had high a moisture content. The structure of D61X and R68X was probed by 1D H-NMR (see figure A2.6.2) and identified that the side chain methyl groups were not structurally constrained, which can be indicative of unfolded protein or molten-globule state. Circular dichroism spectroscopy (CD) was then performed on R68X and the α-3C-β mutant to obtain structural information of the protein backbone. CD identified that both proteins had significant backbone secondary structure, thus suggesting that both mutants were in the molten-globule state. The D61X mutant was also tested but appeared to be degraded and produced a noisy CD spectrum.

1D 1H-NMR



Circular Dichroism

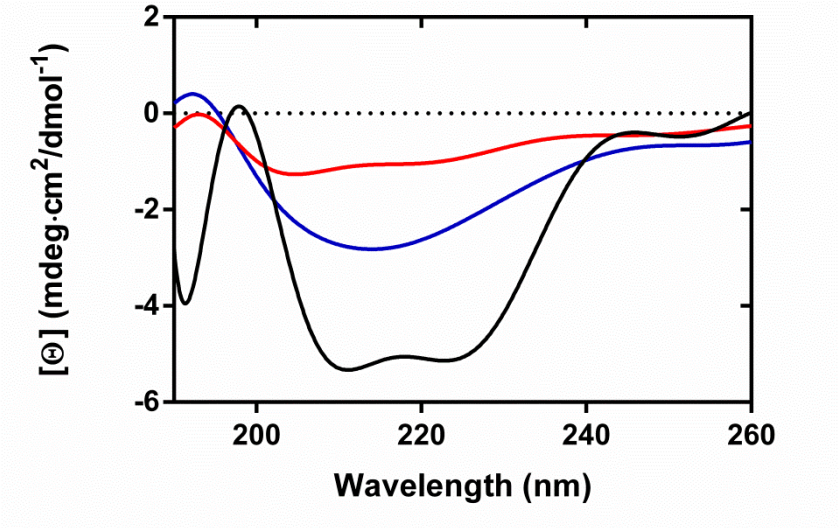


Figure A.2.6.2 1D H-NMR and CD of WT, D61X and R68X Cystatin B

Top: 1D H-NMR; WT cystatin B is a typically folded protein with <0.5 ppm NMR resonances produced from side-chain methyl aromatic residue interactions which is indicative of protein folding. D61X and R68X are devoid of these resonances suggesting a potentially unfolded state. **Bottom:** Circular dichroism spectra showing that both the WT (black), R68X (blue) and α-3C-β (red) amide backbone is folded. The signal from α-3C-β could be weak due to either a very low concentration and/or weak signal generated by β-sheet in the instance of α-helical destabilisation.

A2.3 Discussion

This study was followed up no further due to difficulties encountered in purification and time constraints. Limited proteolysis of WT cystatin B was pursued instead with the aim of then using the technique to elucidate structural information about the truncate mutant amyloid fibres formed and ascertain whether the formed amyloid fibres retain a similar structure. Thioflavin T fibrilisation assays to measure the rate of fibre formation compared to the mutants and the WT could also be performed with the aim of determining whether short amino acid sequences drive fibre formation in a full length folded protein with the two *de novo* synthesised peptides of particular interest.

Appendix 3: Study of Novel Amyloid Therapeutics

A3.1 Introduction

The development of novel therapeutics to treat amyloid disease is of paramount importance. *Salvia* plant species are reputed for their medicinal properties and have been regularly used in European and Chinese folk medicine as an antiseptic, and as treatment for memory loss and insomnia (Rauter *et al.*, 2012). *Salvia* species are thus an excellent plant family for natural product searches to treat amyloid-related diseases. *Salvia sclareoides* is a native Portuguese aromatic herb related to common sage and has been studied in this work and that of others. The preliminary studies on S*alvia sclareoides* n-butanol extract shows it contains components shown to interact with human prion protein (91-231 amino acid construct, huPrP$_{91-231}$) (Rauter et al., 2012). In that HSQC-NMR experiment, attenuation of amino acid resonances of 93-112, 130-187 and 205-225, which represent regions in the unstructured N-terminal region, two strands of the β-sheet and the three α-helices, indicates potential extensive binding and associated conformational changes. This attenuation of resonances is typically associated with ligand exchange dynamics, and is commonly observed under non-saturating conditions for ligands with dissociation constants in the μM range (Rauter *et al.*, 2012).

Further to this study, persistently infected scrapie mouse SMB cell lines were tested for reduced levels of scrapie form prion protein (PrP^{SC}) with only the *Salvia* water extract attenuating the PrP^{SC} concentration significantly (EC_{50} = 5 µg/ml) (Jennifer Louth, personal communication). All *Salvia sclareoides* extracts were tested along with the subsequently listed pure compounds from 1-50 µM: rosmarinic acid, ferulic acid, vanillic acid, syringic acid, p-coumaric acid, catechin, rutin, gentisic acid and ursolic acid. The majority of extracts studied have proven to be non-toxic in cell culture, thus viable for further therapeutics research.

In addition to the prion studies, *Salvia* extracts have been shown to destabilise cystatin B dimer, which exchanges to a species not detectable by NMR, most likely amorphous aggregate (work by Nadia Makaya and Matthew Cliff). To extend these preliminary results, a range of S*alvia sclareoides* extracts and pure compounds identified in the extracts have been screened for binding to soluble cystatin B species by HSQC NMR, and the effects upon fibrilisation and preformed fibre stability were probed. Nadia Makaya and Catarina Dias (Masters level students) performed these experiments wholly or partly whilst I designed, assisted or completed the experiments and analysed the results. The study objectives were to investigate the interaction of S*alvia sclareoides* extracts, determine the active compound and identify the underlying modal activity.

A3.2 Materials and Methods

***Salvia sclareoides* Extracts; Component Compounds and Extraction Procedures**

The *Salvia sclareoides* extracts used herein were produced by Prof. Amelia Rauter at the University of Lisbon, with identification of some of the compounds present. These extracts contain phenolic compounds and flavenoids such as (+)-catechin, (see table A3.2.1) (Rauter *et al.*, 2012) which are currently of great interest in AD drug discovery (Hard and Lendel, 2012). Figure A3.2.1 illustrates the extraction processes and highlights that all the extracts have had the extraction solvent removed by evaporation. 10% DMSO buffered aqueous solvent was used to resuspend the majority of the extracts and act as a general organic solvent to solvate organic active compounds (infusion and water extract were suspended in buffer). DMSO was used due to its high solvating qualities and suitability for NMR and fluorescence spectroscopy experiments.

	Extracts			
Phenolic Compound	**Methanol**	**Ethanol**	**Butanol**	**Acetone**
Gallic Acid	0.8	0.8	1.2	1.7
(+)-Catechin	-	-	0.5	-
Gentisic Acid	-	1.0	-	3.6
Chlorogenic Acid	1.1	-	1.3	-
Caffeic Acid	-	0.8	1.4	2.4
Vanillic Acid	2.5	1.3	2.8	6.9
Syringic Acid	-	-	-	2.1
Coumaric Acid	1.3	0.1	0.6	0.8
Ferulic Acid	0.8	-	0.5	0.1
Rutin Hydrate	-	-	-	1.6
Kaempferol glucoside	-	-	0.7	-
Quercetin dehydrate	-	-	1.4	-

Table A3.2.1 Phenolic Compounds in *Salvia sclareoides* Extracts

Relative percentages of phenolics determined by HPLC. Data from (Rauter *et al.*, 2012).

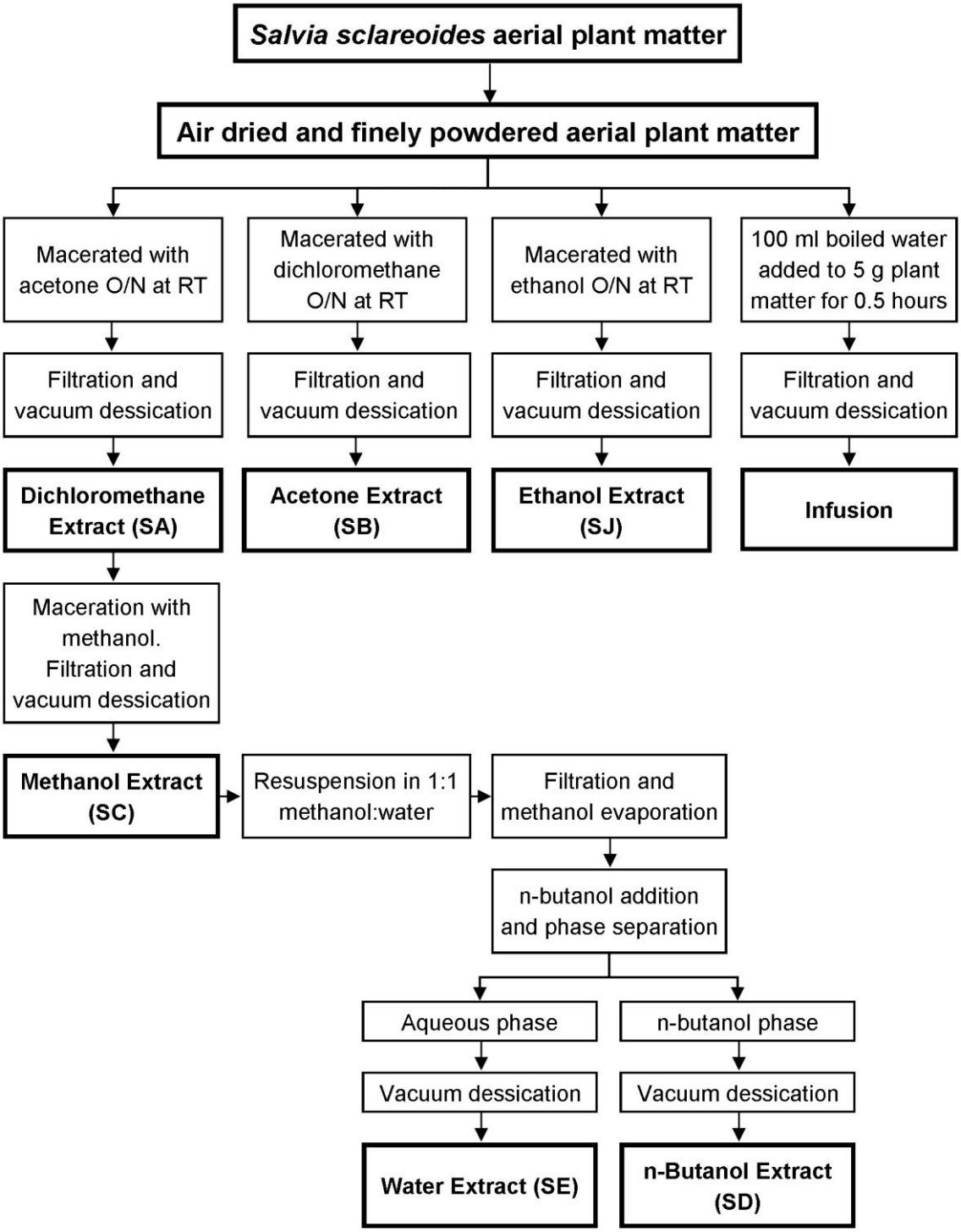


Figure A3.2.1 *Salvia sclareoides* Natural Product Extraction Procedures

The natural product extraction procedure is performed on the plant aerial parts which are above ground. The abbreviation O/N and RT are short for overnight and room temperature. Vacuum dessication produces a powdered extract which was then typically resuspended in a 10% DMSO/water solution.

Cystatin B Defibrilisation Thioflavin T Assays with Addition of *Salvia* Extracts

The continuous defibrilisation assays were performed at 37°C in a similar manner to the fibrilisation assay standard method described in section 2.6. Mature amyloid fibres produced by the method described in section 2.6 were centrifuged at 14 000 g for 15 minutes, washed with pH 6.0, 11 mM sodium phosphate buffer (165 mM NaCl, 1 mM NaN_3) centrifuged again and the procedure repeated to wash the fibres and ensure the correct and reproducible pH value was obtained. Shortly before assay initiation 10% DMSO was added to the resuspended fibre in aqueous buffer. 10 µl/ml of *Salvia* extract or pure compound was added to the protein solutions in all the experiments described herein and the relative concentrations of the primarily investigated extracts and compounds are shown in table A3.3.2 The discontinuous assay method was similar to the continuous assay, except that several millilitres of the fibre and extract solution were incubated without Thiflavin T (ThT), then at specific time intervals 1 ml of the solution was aliquotted and ThT (10 µM final concentration) was added before quickly measuring fluorescence.

The defibrilisation assays were performed at pH 6.0 with the rationale for this being that the soluble protein is maximally stable at this pH (predicted pI 7.0 and molten globule state at pH 4.7), thus reducing the amyloid fibres dynamic 'on' rate, and therefore allowing clear observation of the extract effects. The expected effects of the extract upon amyloid fibre were uncertain thus the experiment was designed to maximise potential observable changes. In addition, pH 6.0 is closer to physiological conditions than the pH 4.7 amyloid forming milieu, thus altering the protonation state of protein and extract molecule closer to that encountered as a therapeutic agent *in vivo*. Cystatin B amyloid fibres, by EM and H/D exchange, had been previously reported as stable up to pH 9.0 for a period of months (Morgan *et al.*, 2008), therefore the fibres were expected to be suitably stable. It was thought that performing the experiment at pH 4.7, with or without 10% Trifluoroethanol (TFE) would not work due to competition with fibre formation, thus a presumed stable pH was chosen.

Salvia Extract	Preparation Concentration (mg/ml)	Final Concentration (µg/ml)	DMSO Percentage
Acetone (SB)	12.5	62.5	10
Butanol (SD)	25	125	10
Water Extract (SE)	12.5	62.5	0
Infusion	25	125	0
Gentisic Acid	3.1 (20 mM)	30.8 (200 µM)	0
Rosmarinic Acid	3.6 (10 mM)	36.0 (100 µM)	10

Table A3.2.2 Primary Tested Extract Concentrations

Cystatin B Fibrilisation Thioflavin T Assays with Addition of *Salvia* Extracts

The *Salvia* extract fibrilisation assay used the standard fibrilisation conditions described in section 2.6. The samples were incubated at 37°C, with the addition of 10% DMSO and 10 µl of extracts performed rapidly before commencement of data collection resulting in an experimental dead-time of 9 minutes with up to four samples at a time.

A3.3 Results

Cystatin B Defibrilisation Thioflavin T Assays with Addition of *Salvia* Extracts

The initial cystatin B continuous defibrilisation thioflavin T (ThT) fluorescence assay with acetone and butanol *S. sclareoides* extracts and a 10% DMSO fibre control (figure A3.3.1) showed a rapid decrease in ThT fluorescence in all the samples tested. This included the fibre control, which indicates reduction of amyloid fibre. Attenuation of the amyloid fibre control threw the results into question and indicated fibre destabilisation, which could have been caused by the presence of 10% DMSO, pH 6.0 solvent and/or ThT. However, the rate of fluorescence emission reduction to a matter of minutes with the addition of the butanol extract and within hours in the presence of acetone extract suggested this experiment should be repeated with more controls due to significantly faster attenuation rates in the extract samples. To ensure fluorescence quenching was not occurring, absorption spectra of the extracts in identical buffer conditions and concentration as the assay were performed from 220-600 nm. These spectra proved the extracts did not absorb above signal-to-noise levels at or near 482 nm (data not shown). Furthermore, EM confirmed conversion of mature fibres to amorphous aggregate.

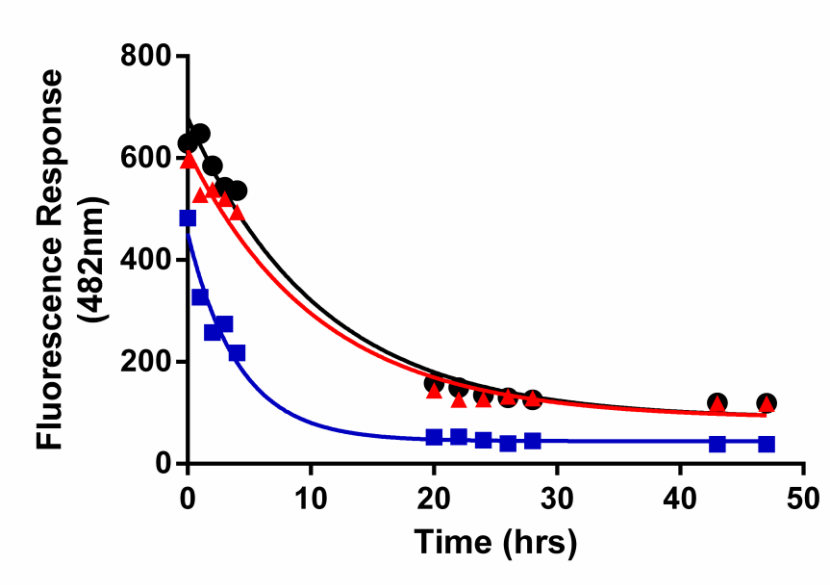


Figure A3.3.1 First Cystatin B Defibrilisation with *Salvia* Extracts

Defibrilisation experiment with the addition of acetone extract (red), butanol (blue) and a 10% DMSO (black) fibre control. Data fitted with a single exponential. All samples were incubated at 37°C and contain 1 ml of 30 µM mature cystatin B fibre in pH 6.0 sodium phosphate buffer, 10% DMSO, 10 µM thioflavin T and a 10 µl addition of extract. The control fibre (10% DMSO) and *S. sclareoides* extract preparations all exhibit rapid defibrilisation suggesting the assay conditions as well as the extracts are responsible for fibre reduction.

A subsequent defibrilisation experiment repeated the previous conditions with replicate samples, and a mature fibre control with and without 10% DMSO or 10% TFE, all at pH 6.0 (see figure A3.3.2). All the samples tested, including the controls, once again indicated a decrease in ThT fluorescence. The reduction of the controls indicated that at pH 6.0, the addition of ThT significantly destabilises the mature amyloid fibre and the additional DMSO solvent was not responsible for the previous control fibre destabilisation. Against expectations, 10% TFE in these conditions did not promote fibre stability, thus suggesting that neither soluble protein nor amorphous aggregate is suitably destabilised in these conditions to maintain the fibre population. Once again acetone and butanol *S. sclareoides* extracts caused rapid defibrilisation, and are thus pharmacologically interesting.

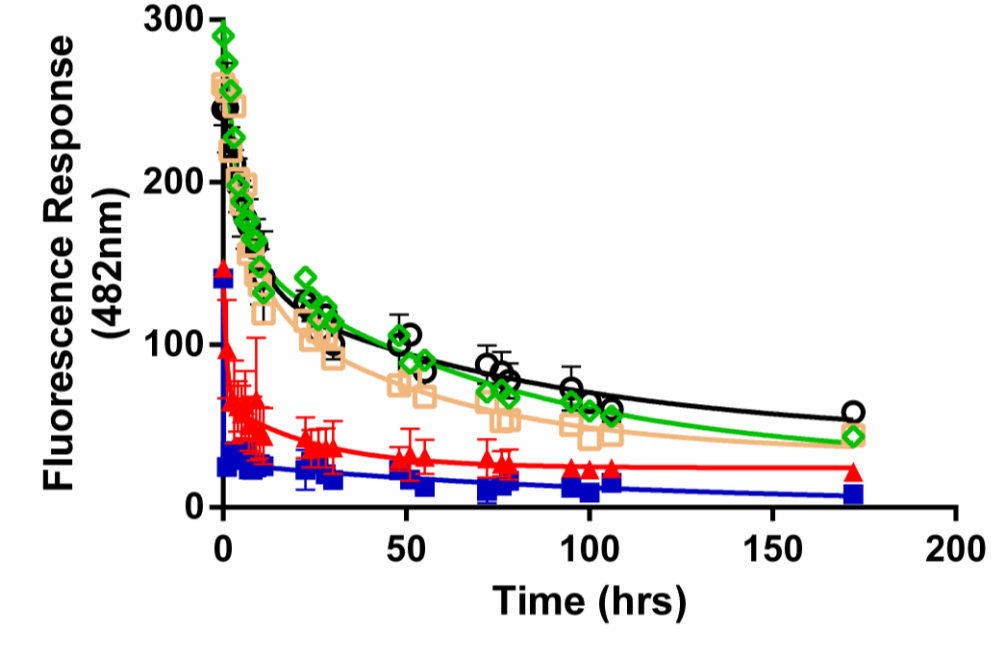


Figure A3.3.2 Second Cystatin B Defibrilisation Assay with *Salvia* Extracts

Defibrilisation experiment with acetone (red) and butanol (blue) *S. sclareoides* extracts and fibre controls with and without 10% DMSO and 10% TFE (black, green and orange respectively). The raw fluorescence response of the samples has been baselined with ThT-extract controls, which were minimal except for the acetone extract which exhibited some intrinsic fluorescence at 482 nm. The curves are fitted with a double exponential which was required for fitting, because a single exponential fitted the data poorly. All samples were incubated at 37°C and contain 1 ml of 30 µM mature cystatin B fibre in pH 6.0 sodium phosphate buffer, 10% DMSO, 10 µM thioflavin T and a 10 µl addition of extract. The controls such as the fibre blank, fibre with 10% TFE and fibre with 10% DMSO all show a similar reduction in fluorescence. However, the preparations with *salvia* extracts show a marked and rapid decrease in fibre compared to the slower decrease shown in the controls.

Two subsequent discontinuous defibrilisation assays were performed (see figure A3.3.3), whereby incubated preformed fibre controls or samples with *Salvia* extract were aliquotted at specific intervals from an incubated solution, ThT added and the fluorescence measured within a minute. The first (top) assay identifies a reduction in cystatin B amyloid fibre within minutes of acetone and butanol extract addition. The second (bottom) assay shows cystatin B amyloid fibre dissociation within a matter of hours with additions of infusion, water extract and rosmarinic acid. Both fibre and 10% DMSO fibre controls produce a high, flat fluorescence signal indicating that the fibres are stable in the conditions used, particularly in the absence of ThT until shortly before the measurements were taken. This verifies that ThT is destabilising under these experimental conditions. The discontinuous assays show that all the *Salvia* extracts tested and rosmarinic acid induce fibre dissociation in the conditions tested. The fibre

controls remained stable and fibrous throughout the experiments due to the absence of ThT. In all the defibrilisation assays perfomed, EM of the final solutions with extracts or pure compounds proved mostly negative for amyloid fibre and displayed amorphous aggregate structures as shown in figure A3.3.4.

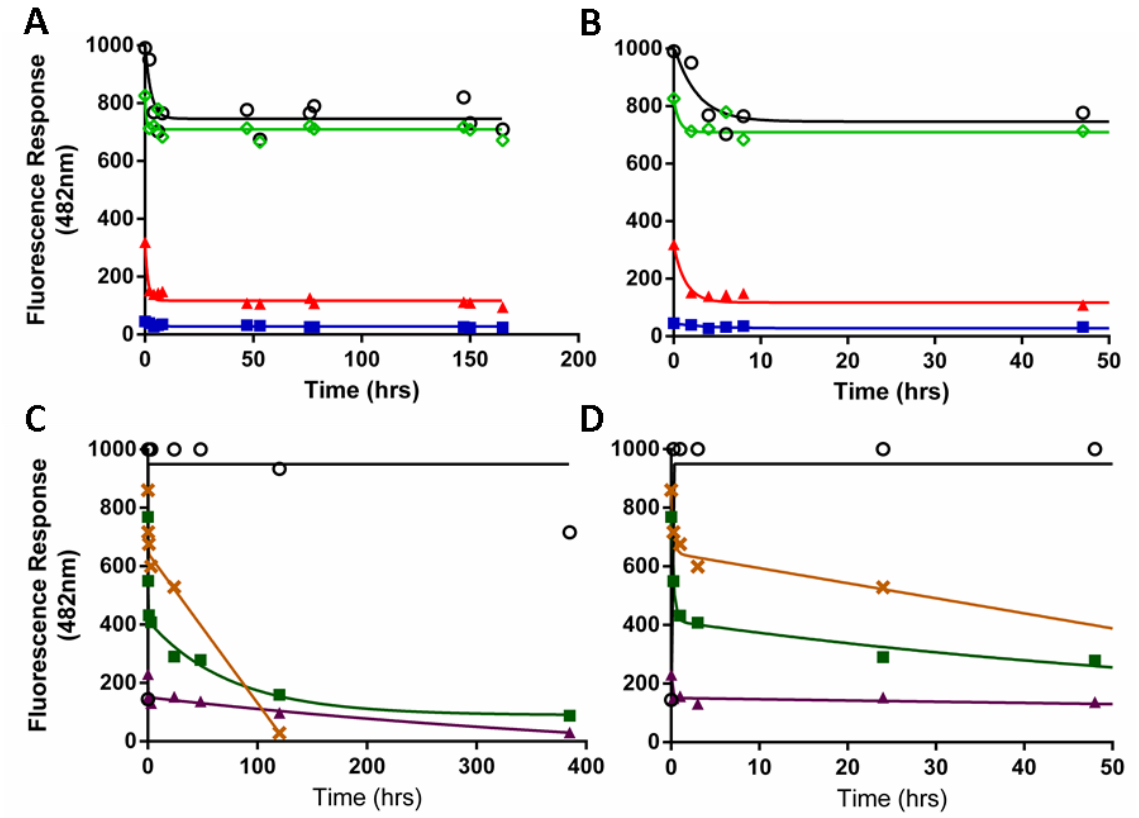


Figure A3.3.3 Discontinuous Cystatin B Defibrilisation Assays with *Salvia* Extracts

(A and B) The first assay conducted with acetone extract (red triangles) and butanol extract (blue squares) both show significant decreases in fibre content. The fibre control (green diamonds) and a 10% DMSO fibre control (black circles) show relatively flat profiles thus indicating that in the discontinuous assay conditions the fibre remains stable. **(C and D)** Second assay with water soluble extractions which include; rosmarinic acid (purple triangles), infusion (green squares) and water extract [SE] (orange crosses). The fibre control remains relatively stable, however a value of 1000 is the detection limit of the instrument, so the actual value may have been higher. All samples were incubated at 37°C and contain 1 ml of 30 µM mature cystatin B fibre in pH 6.0 sodium phosphate buffer, 10% DMSO, 10 µM thioflavin T and a 10 µl addition of extract.

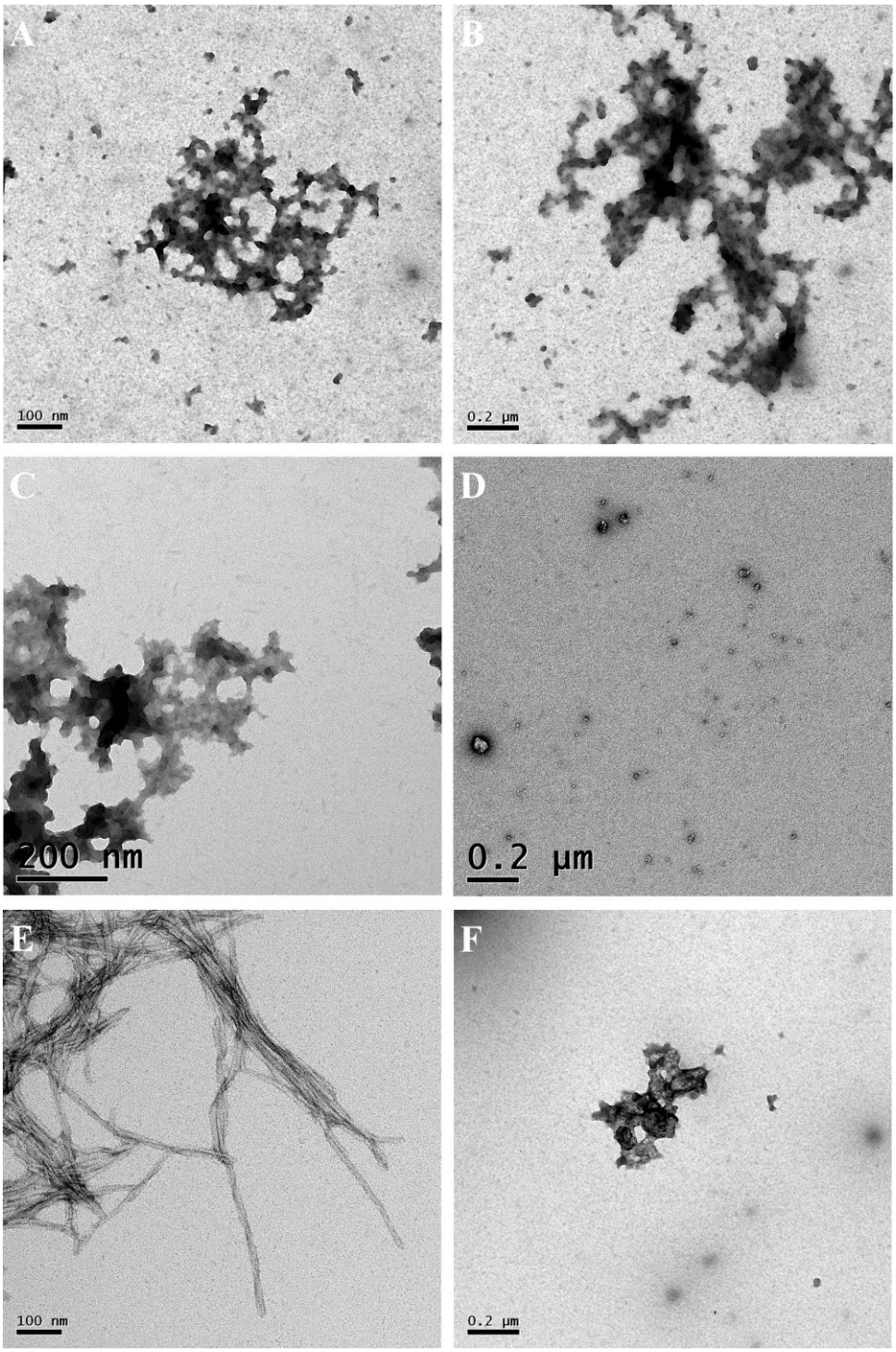


Figure A3.3.4 Selected Defibrilisation Electron Micrographs

Electron micrographs of fibre solutions following defibrilisation assays which shows conversion of amyloid fibre into amorphous aggregate and potential oligomer species. All images are magnified 11 500 x except for C and E which is at 21 000 x magnification. WT cystatin B fibre destabilised with the addition of; **(A)** acetone extract (SB), **(B)** butanol extract (SD), **(C)** water extract (SE), **(D)** rosmarinic acid which shows potential oligomer formation and little amorphous aggregation. Example control images include; **(E)** fibre control with 10% TFE at pH 6.0 and **(F)** Extract B which identifies small globular and protein aggregate structures. These are however much smaller, often have a different morphology and are less extensive than the aggregate products found in extract and fibre preparations.

In an attempt to capture the early defibrilisation event, which was mainly missed with the discontinuous assays, a short continuous 25 minute assay was performed with sampling every 0.5 minutes, a dead-time of 24 seconds and gentle mixing of the sample every 2 minutes by hand. This rapid defibrilisation assay was performed with butanol extract (figure A3.3.5) and proved successful, although the initial exponential phase was most likely missed as the typical fluorescence response of this fibre preparation was about 800 units. Improvement of this experiment by diluting the extract and reducing the temperature (conducted by Amy Irvine) indicates that the defibrilisation rate remains very fast and cannot be measured without resorting to stopped flow methods.

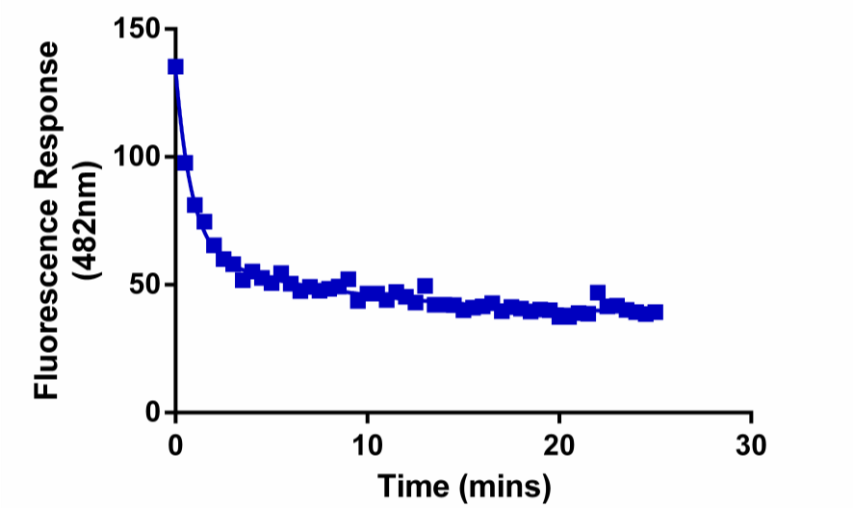


Figure A3.3.5 Rapid Cystatin B Defibrilisation Assay with *Salvia* Butanol Extract

This rapid continuous assay with ThT present allowed clearer observation of the defibrilisation event caused by addition of Salvia sclareoides butanol extract to mature cystatin B fibre. Data fitted to a double exponential curve.

WT Cystatin B Thioflavin T Fibrilisation Assay with Addition of *Salvia* Extracts

The aim of this experiment was to observe an expected reduction in fibrilisation rate caused by *S. sclareoides* extracts. The fibrilisation experiments were performed in standard fibrilisation conditions at pH 4.7 and the thioflavin T fluorescence time courses are shown in figure A3.3.6. Compared to the control, the fibrilisation rate is either equal or enhanced with the addition of the extracts. EM of these samples at the end of the experiment identified that all the samples had significant quantities of amyloid fibre present.

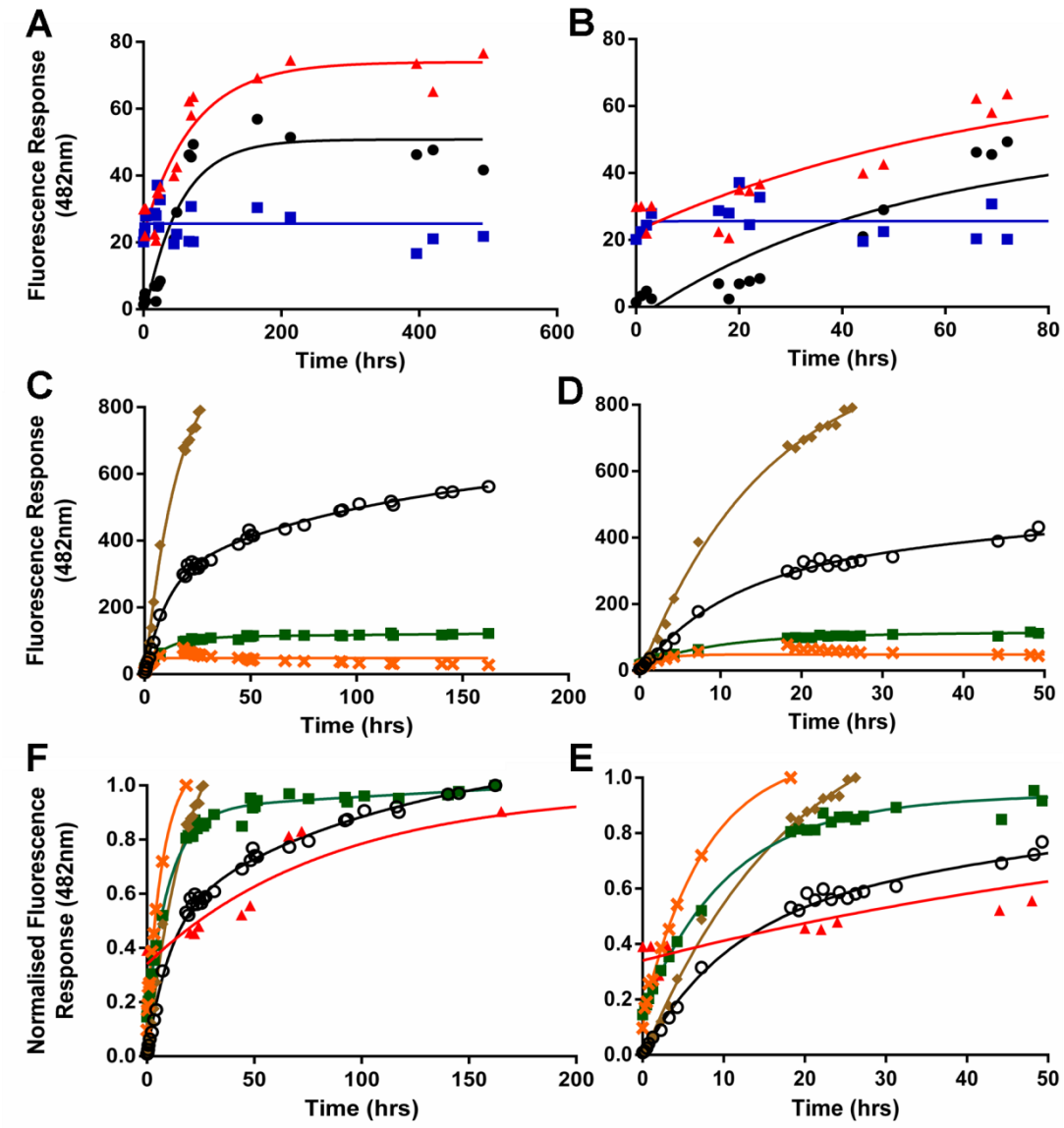


Figure A3.3.6 Cystatin B Fibrilisation with *S. sclareoides* Extracts

(A and B) The first assay conducted with acetone extract (red triangles), butanol extract (blue squares) and 10% DMSO/TFE Control. The acetone extract does not impede fibrillogenesis with a marginally increased rate. The butanol extract either forms fibre very rapidly or inhibits fibre formation with the former most likely as EM shows mature fibre formation in all samples at the end of each incubation period. **(C and D)** The second assay used water soluble extractions which include; gentisic acid (brown diamonds), infusion (green squares), water extract [SE] (orange crosses) and a 10% TFE fibre control (black circles). The reaction containing gentisic acid is clearly faster than the control and rapidly forms fibres. The absolute fluorescence values of reactions with infusion and water extract however are lower than the control but when observed comparatively in (E and F) the rates of formation are quicker than the control reaction. **(E and F)** Comparison of a selection of fibrilisation reactions with *S. sclareoides* extracts. The data when normalised suggests that all the extracts form fibre faster than the control. The overall fluorescence intensities with extract present are typically much lower than the controls as shown in panels A-D. All samples were incubated at 37°C and contain 1 ml of 30 µM mature cystatin B fibre in pH 6.0 sodium phosphate buffer, 10% DMSO, 10 µM thioflavin T and a 10 µl addition of extract.

A3.4 Discussion

It has been determined that 10 μM ThT in combination with the pH 6.0 conditions destabilises mature cystatin B amyloid fibre over a period of hours. In comparison, 10 μM ThT at pH 4.7 has been shown not to alter the fibrilisation rate (Giannini, 2004). Thioflavin T binding at pH 6.0 clearly destabilises the amyloid fibre most likely through its binding interactions to the continuous β-sheet. Thioflavin T binding is stronger at pH 6.0 than at lower pH values such as pH 4.7 (Naiki *et al.*, 1989, LeVine, 1997). Equally, the binding of ThT and the environment (pH 6.0) could combine to sufficiently stabilise the soluble cystatin B species and disturb the equilibrium and/or the fibre in these conditions becomes less robust, which leads to fibre dissociation. If significant, this stability could perhaps be studied by CD or FTIR with thermal or chemical denaturation. The discontinuous assays and EM showed that in the absence of ThT, the fibre controls remained stable and fibrous throughout the experiments.

From the two discontinuous assays it is clear that all the *Salvia* extracts and pure compounds tested induce fibre dissociation as detected by ThT fluorescence and EM. An example of tested pure compound is rosmarinic acid which has been shown to produce a very similar effect to that of the *Salvia* extracts, has been identified in the extracts as a major component (Catarina Dias, personal communication) and is observed to destabilise and inhibit Aβ fibre in ThT assays and EM (Ono *et al.*, 2004). Cystatin B and Aβ fibres in stable conditions would ideally be studied further in similar discontinuous ThT assays or through the use of other techniques such as CD, FTIR or quantitative dot blot assays (Wanker *et al.*, 1999) using fibre specific antibodies (Kayed *et al.*, 2003). The presence of ThT in the assays has been shown to produce false positives and negatives due to competition for binding, (Lendel et al., 2009, Buell et al., 2010, Hudson et al., 2009) and although useful must also be followed up with other experimental methods.

The rapid defibrilisation assay with butanol extract proved successful with measurement of the end stages of the defibrilisation reaction, which had not been observed previously with this highly effective extract. Over the 25 minute period, the effect of 10 μM ThT and pH 6.0 on fibre destabilisation was not measurable. Therefore ThT presence was deemed not to affect these results unduly and was necessary for the regular measurements required. Extracts exhibiting slower kinetics over a period of hours may have problems with the ThT effect, thus adequate controls would have to be trialled before screening the remaining extracts and pure compounds in this manner. Without using a stopped flow system this is likely to be the best ThT based assay method possible, without the manipulation of the conditions such as reducing the temperature and extract concentration.

The fibrilisation results indicate that compared to the control, the fibrilisation rate is either equal or slightly enhanced with the addition of the extracts. This could be due to increased amorphous aggregate formation, upon extract addition, which from cystatin B fibre forming EM time-course experiments appears to be on-pathway to fibre formation (Paramore, 2010). Escalating the amorphous aggregate population may increase the rate of nucleation or elongation of the amyloid fibres at pH 4.7 by presenting randomly folded and orientated cystatin B. Equally, given that the results are within experimental error, it is possible that the true effect is missed under the conditions used. The work should be repeated at 30°C, to slow the kinetics of the reaction for clearer observation of the lag and elongation phases.

Potential binding of soluble *Salvia* extracts or pure compound to Cystatin B was tested by HSQC NMR with only butanol and 'water' extracts indicating marked changes in spectra, whereby the dimer species resonances disappeared without increase in monomer or tetramer peaks. The pure compound which causes dimer conversion to what is most likely amorphous aggregate is yet to be identified. This result along with the prion protein studies indicate that the most interesting extracts to study are the butanol and water extracts (SD and SE). In addition, the aqueous solubility of the water extract would make the active compound an ideal therapeutic agent in terms of meeting ADMET conditions.

A3.5 Conclusions

10 µM Thioflavin T at pH 6.0 causes mature cystatin B fibre formed at pH 4.7 to dissociate over a period of hours, thus suggesting the soluble protein and or/fibre stability is altered. All the extracts tested for defibrilisation show strong activities at comparable concentrations. The pure compound which causes the dimer conversion to what is most likely amorphous aggregate is yet to be tested or identified, thus requires further study. The most interesting extracts to study are the butanol and water extracts (SD and SE). The rapid defibrilisation assay proves to be the most successful method employed, thus with alteration of the conditions it could prove to be an excellent screening method. To conclude, the effects of *Salvia sclareoides* extracts is worth pursuing further due to the positive preliminary results presented herein.