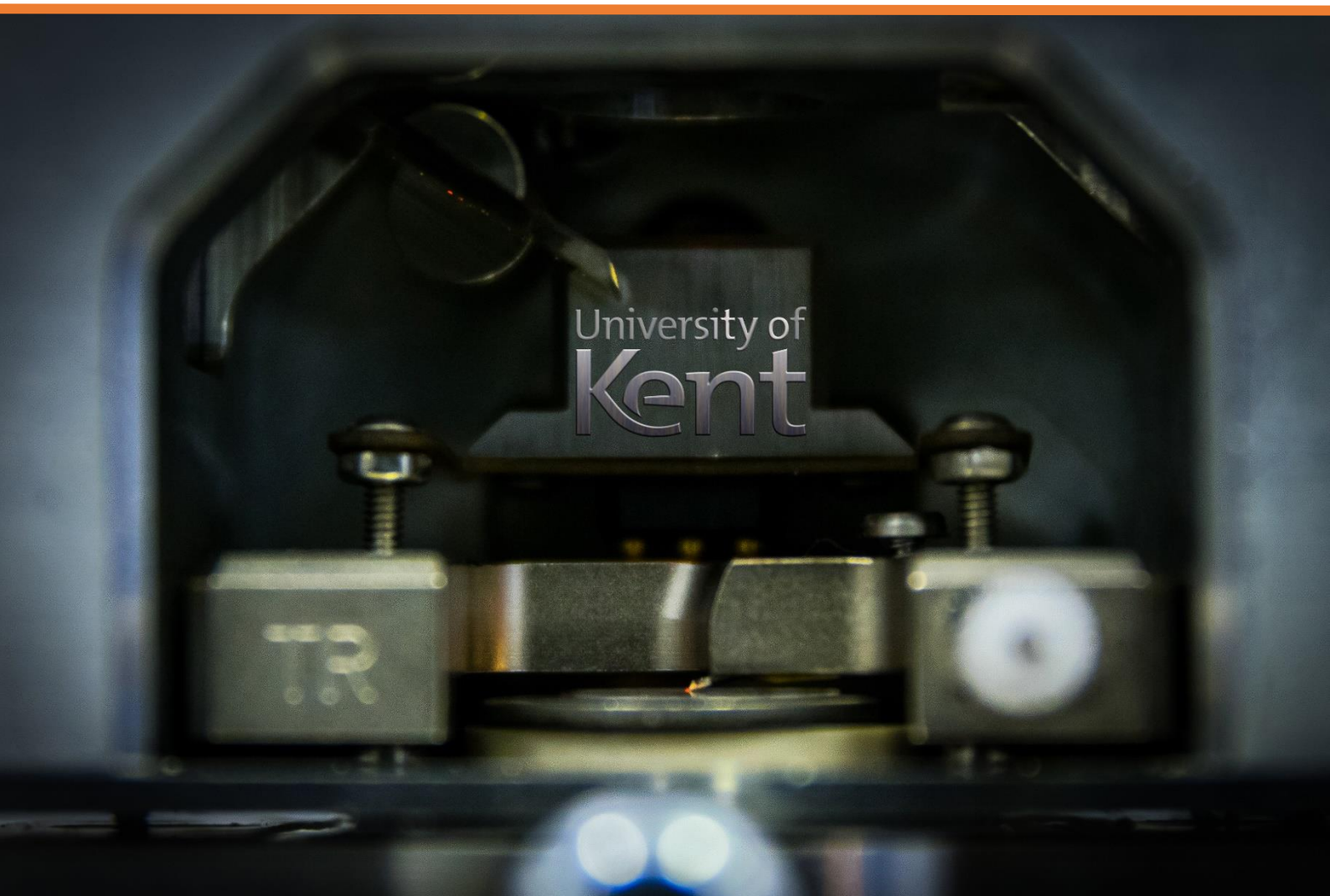
The polymorphic nature of amyloid assembly:

Exploring fibril morphology and the structural relationship towards mechanical stability

Ben J.F. Blakeman

A thesis submitted for the degree of PhD in Biochemistry



Declaration

I hereby declare that this thesis has not been submitted in part or in full towards the completion of any other degree or qualification at the University of Kent, or any other institute of learning.

Ben Blakeman

May 2017

Acknowledgements

I would like to acknowledge and thank those that have supported me in my journey towards completing my PhD.

First, I would I would like to thank Dr Wei-Feng Xue and Dr Louise serpell in providing me with the opportunity of working on this project. I would like to further acknowledge and thank Dr Wei-Feng Xue in his endless support and guidance as both my supervisor and friend. In many regards, you have become a role model to me. Also, thank you for inspiring me to pursue photography and providing me the opportunity to surpass you at squash (I wish).

To everyone in the Kent fungal group (KFG), I would like to thank you all for creating a work environment that I have thoroughly enjoyed. Many of you have been available to provide support when needed both professionally and personally and I am glad to call you my friends. I would especially like to thank Dr David Beal, who has unequivocally provided me with invaluable support, from before I even begun my PhD to the very end. Dan, Elliot, Gemma and Dave I will always miss our daily banter; it made being in the lab a constant joy with endless laughs. I have learned so much from you all, and consider you models in many regards. Dan and Elliot, through our many shared interests, you both helped make the lab a truly enjoyable place to work. I am also glad to have finished my time in the KFG as the reigning squash champion.

Outside of the KFG I would like to thank Ian Brown for always being there to provide help, ideas and general conversation. Also, making the AFM less lonely with your singing. I would also like to thank my good friend Chris Miller, for both a collaboration on his project but mainly for being there to provide an escape from frequent monotony. You are one of the most interesting and intelligent people I have ever met and I will miss our frequent discussions about anything and everything, I am sure some of your knowledge rubbed off on me!

Finally, I would like to thank my family who have supported me emotionally, financially and professionally through my PhD. None of this would have been possible without them. My Fiancé Sarah has been my rock throughout my PhD, regarding both professional and personal life. I would not have been able to achieve this without her constant love and support. She is the reason I have strived to achieve my PhD.

Abstract

The polymorphic nature of amyloid fibrils is important in the understanding of structural based relationships, such as a morphology influence on cytotoxicity and disease progression. The work reported here uses Atomic force microscopy (AFM) to enhance the understanding of fibril morphology in addition to the relationship between structure and stability towards breakage. A novel quantitative cluster analysis was developed here to identify the vast range of fibril morphologies present within a population. Using fibrils formed from three peptide sequences identified by the WALTZ algorithm, we have characterised the polymorphism displayed by each fibril population and provided structural models to predict the likely filament arrangements accessible to each. The range of fibril polymorphism also conveys mechanical differences, defined here by persistence length values for each respective population. These mechanical differences subsequently affect fibrils stability towards breakage, quantified here using AFM and subsequent image analysis.

Additionally, using AFM, a structural comparison was performed between Sup35NM amyloid fibrils formed *in vitro* and those formed *in situ* using a synthetic biology approach with the Curli-dependent amyloid generator (C-DAG) in *E. Coli*. Structural similarities between fibrils formed using this system and those formed in vitro is of great value given the importance of a sequence-structure relationship. The work in this thesis expands on possible fibril morphologies and the related mechanical properties, which has implications in the understanding of disease enhancing structural motifs and the utilisation of amyloid fibrils in a biotechnology role.

Abbreviations

AFM	Atomic force microscopy
αEB	α-Synuclein ion exchange buffer
αEIB	α-Synuclein ion exchange elution buffer
αLB	α-Synuclein lysis buffer
TM-AFM	Tapping mode - Atomic force microscopy
APP	Amyloid precursor protein
CJD	Creutzfeldt-Jakob disease
CD	Circular dichroism
C-DAG	Curli – dependent amyloid generator
DNA	Deoxyribonucleic acid
DMSO	Dimethyl sulfoxide
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
kDa	Kilodaltons
L_p	Persistence length
L_c	Contour length
mQ	Milli-Q
Ni-NTA	Nickel-Nitrilotriacetic acid
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
PBS	Phosphate buffered saline
SLB	Sup35NM lysis buffer
SLBG	Sup35NM lysis buffer denaturing buffer
SLBGE	Sup35NM lysis buffer elution
SFFB	Sup35NM fibril forming buffer
syFFB	α-Synuclein fibril forming buffer
Rpm	Revolutions per minute
TAE	Tris base, acetic acid and EDTA buffer

Contents

Chapter 1

An introduction to amyloid: Diseases, structures, and methods

Protein mis-folding and the resulting amyloid associated disorders

Amyloid aggregates are thought to be able to form from any soluble protein. This occurs following a protein mis-folding event and subsequent self-assembly. Longitudinal hydrogen bonds are formed between monomer units adjacent to the direction of protein or peptide β-strands , forming a polymer of repeated monomer units with indefinite length (Chiti & Dobson, 2006). The mis-folding events generally result in a partially ordered but unstable conformation prior to self-assembly into fibrils. Globular proteins associated with amyloid disease such as lysozyme and transthyretin or peptides such as $A\beta_{1-42}$ (Bucciantini et al., 2004b) can suffer destabilising mutations, rendering them amyloidgenic. Proteins that are not usually amyloidgenic *in vivo* have been shown to undergo fibril formation, following denaturing conditions that may mimic what would happen to a protein that is affected by destabilizing mutations (Schuler, 1999). Furthermore, certain amyloidgenic proteins are intrinsically disordered. Such unfolded proteins can form into intermediate conformers that contain both α-helix and β sheet structures. For the case of IAPP, through self-association the conformer stabilises into a nucleus rich in β sheet (Walsh, 1999). In the case of Aβ the amyloid precursor protein (APP) is sequentially cleaved through proteolysis, by β and γ secretase. Cleavage occurs at either the 1st or 11^{th} residue and then followed by cleavage at the 99^{th} residue, by β and γ secretase respectively. This results in the release of the Aβ peptide that will originally be unfolded, but through Fibrillogenesis will develop partial conformation that will be stabilised through oligomerization, forming an amyloid 'nucleus' (Han Zhang, Qilin Ma, Yun-wu Zhang, 2012)(J C Rochet & Lansbury, 2000). Such denaturing or mis-folding events result in the ability of the protein to aggregate through nucleation dependent fibrillation. The resulting structure, can be defined as a long unbranched fibril characterised by a cross β sheet of indefinite length. The cross β sheet structure

has been characterised, and been shown to be a shared conformation adopted by all amyloid fibrils (Fitzpatrick, Debelouchina, Bayro, Clare, & Caporini, 2013; Karen E Marshall & Serpell, 2009; L C Serpell & Smith, 2000). Amyloid fibrils resulting from nucleation dependent fibrillation, have been linked to various human diseases.

Amyloid and disease: The amyloid cascade hypothesis in Alzheimer's

Amyloid related diseases can be associated with separate groups based on location of pathology, for example the neurodegenerative conditions, in which aggregation occurs in the brain. Non- neuropathic localised amyloidosis in which the aggregation occurs in a single type of tissue that is not in the brain, and non-neuropathic systemic amyloidosis in which aggregation can occur in multiple tissues. Certain conditions, such as Parkinson's and Alzheimer's can either be sporadic (predominantly) but can also be hereditary such as with mutations that increase the endogenous levels of γ secretase in the case of Alzheimer's. While other conditions are predominantly hereditary in their onset, such as fibrinogen amyloidosis. Other diseases, while also sporadic and hereditary, are also transmissible between hosts. These are amyloid diseases characterised by the development of prions, which are highly infectious amyloidgenic proteins such as for the case of Transmissible spongiform encephalopathies, for example Creutzfeldt–Jakob disease.

Research into amyloid is an important field, in large due to the link to numerous human diseases. Various diseases can be linked to specific species of amyloid, for example Parkinson's which can be linked to Alpha-synuclein and Diabetes mellitus type 2 which involves amylin. Alzheimer's disease (AD) is the most common of the fatal neurodegenerative diseases associated with amyloid. The progressive pathology of the disease is possibly due to the cytotoxic nature of amyloid fibrils, and the disease progression has been suggested through the amyloid cascade hypothesis. This was

proposed in 1991 by John Hardy and David Allsop and has been reviewed multiple times, for example by Eric Karren et al. 2011 (Karran, Mercken, & De Strooper, 2011) and by Christiane Reitz 2012 (Reitz, 2012). Modifications since its initial proposal are a result of varying views in the exact pathology of AD. As a simplified overview, the cascade culminates the stages from proteolytic cleavage of APP to the onset of dementia, the key stages being what happens between these start and end points and is still a topic of intense debate. Initially the hypothesis proposed that the aggregation of $A\beta_{1-42}$ results in interfibrillar association and deposition, hence the formation of senile plaques in the brains grey matter (SP). Senile plaques have been found to contain $A\beta_{1-40}$, $A\beta_{1-42}$ and $A\beta_{1-38}$, initially suggesting these plaques were the cause of the neuronal cell death. Amyloid has been presented as potent synaptotoxins which block proteasome function (Tseng, Green, Chan, Blurton-Jones, & LaFerla, 2008), inhibit mitochondrial activity and alter intracellular Ca2+ levels (Bucciantini et al., 2004b), stimulating the inflammatory processes, which leads to dementia. It was also proposed that senile plaques lead to the formation of neurofibrillary tangles (NFT) and these also contributed to neuronal cell death (Hardy, 2009). Considering that senile plaques are composed of Aβ aggregates, it has been previously shown that Aβ interacts with signalling pathways that are responsible for the phosphorylation of the microtubule associated protein tau. This results in hyper phosphorylation of tau, disrupting its native function and leading to the accumulation of toxic tau species which aggregate and deposit as NFT (Busciglio, Lorenzo, Yeh, & Yankner, 1995). Therefore, supporting the idea that senile plaque formation leads to NFT formation. An additional mechanism in which Aβ can aid in the accumulation of tau is through the inhibition of the proteasome (Tseng et al., 2008) which would usually degrade the hyper phosphorylated tau.

However It has been suggested that SP and NFT are a result of the pathology of AD and are not in fact a causative factor (Herrup, 2015). While it is certain that Aβ contributes to AD pathology there is now doubt that plaques which consist of fibrillar deposits are the cause of neuronal cell death, with a weak correlation between plaque formation and AD pathology. Down syndrome patients with elevated levels of Aβ displayed early onset of AD (Goldgaber D, Lerman MI, McBride OW, Saffiotti U, 1987) and in various hereditary AD cases, the mutation was in the gene encoding APP or the presenilin genes that codes for γ secretase (Suzuki N, Cheung TT, Cai XD, Odaka A, Otvos L Jr, Eckman C, Golde TE, 1994). This suggests that Aβ fibril formation is involved in AD pathology, however the stage of fibrillation that produces the toxic species is of huge debate, not only for AD but for fibril pathogenesis and cytotoxicity in general. It has been suggested that Aβ dimers are the toxic species (Mc Donald et al., 2010), also shown to be highly toxic are Aβ fibril monomers and oligomers, of varying numbered repeats (Greenwald & Riek, 2010). These cytotoxic species may cause neuronal cell death during the fibrillation process, and the resulting plaques and NFT seen in AD may be a result of the cell death caused by these toxic intermediates. Therefore, while there is much debate as to what stage of the amyloid cascade neuronal cell death occurs which causes dementia, AD is a good example of how amyloid can cause disease and it shows the complexity of diseases related to amyloid.

The amyloid cascade hypothesis is a theory that expanded on the understanding of the involvement of amyloid in associated diseases. It is met with opposition towards its validity and has even been called for rejection by the community in search of another possible cause of Alzheimer's disease (Herrup, 2015). While there seems no doubt in the correlation between the presence of amyloid material and the Alzheimer's pathology, some would argue that it is not representative of causation and the amyloid

build up may be a result of disease progression. The key motive for rejection addresses suggestions of the cascade hypothesis; an individual with amyloid deposits in their brain should display the disease. However, it has been shown that individuals with substantial deposits in their brain, can display little to no clinical signs of dementia (Villemagne et al., 2011). In addition, various factors provide a discrepancy between amyloid and disease causation. For example, the isolated expression of Aβ in transgenic mice is not sufficient for disease onset and cognitive impairment despite plaque formation (Kim et al., 2013). Additionally it has been suggested that inhibition of γ-secretase in clinical models may in fact enhance AD symptoms when given its proposed role in the cascade; would be assumed to reduce symptoms (Herrup, 2015). The γ-secretase which is responsible for cutting of the APP, is presenilin protein encoded by either PSEN1 or PSEN2. Highlighted as a reason for rejection is the observation that in transgenic animal models with PSEN upregulated, there is no pathology of the disease (Herrup, 2015). Overall, while there is little doubt that AD is an amyloid associated disease; many scientists in the field suggest that Aβ or Tau alone may not be sufficiently responsible for the complexity AD pathology. Proposing additional avenues for causation is difficult given the rigid implementation of Aβ into AD pathology, however evidence has been provided for such. For example, a strong case has been made for the excitotoxic loss of Ca^{2+} homeostasis as the cause of AD (Supnet & Bezprozvanny, 2010). Regardless, substantial evidence strongly relates amyloid to AD amongst other diseases, therefore its thorough understanding is vital in either confirming or rejecting its role as the cause of AD.

Understanding the relationship between fibril size and cytotoxicity

While other amyloid associated diseases result from pathways that are possibly different to the Alzheimer's cascade hypothesis, the cytotoxic involvement of amyloid

species is a likely shared mechanism. The study of how amyloid species carry out their membrane disruption is an important area of amyloid research. While it was originally thought that the fibrillar form of amyloid is responsible for the cytotoxicity of amyloid, through the year's evidence has grown for the accountability of smaller, oligomeric structures that amyloid adopts as an intermediate stage. In addition to this it has been suggested that dimers can also represent the toxic species (W. Xue, Hellewell, Hewitt, & Radford, 2010). Monomers and membrane bound oligomers have also been indicated as such (Bai, Zhang, Liu, & Li, 2015; Williams & Serpell, 2011). While small oligomeric species accumulate during early stages of Amyloid formation, it has also been shown that small fibrils (~400nm) that result from the fragmentation of longer fibrils also display cytotoxicity due to their small size. The size of fragmented fibrils is likely similar to that of pre-fibrillar oligomers. Therefore, it seems that amyloid particles of reduced length are in fact highly cytotoxic species. Fragmentation of full length fibrils enhances the cytotoxic potential of a population by an increase in the number of individual particles (W. Xue et al., 2010). The physical dimensions and inter-filament interactions of fibrils seem to contribute to the stability towards fragmentation, which is what makes mechanical properties such as rigidity important factors in fragmentation behaviour. The interactions between filaments; within fibrils, and the balance of these with the electrostatic repulsion may be a contributing factor towards fibril stability.

While the potential cytotoxicity may be a size dependent mechanism, cytotoxicity is still likely due to a variety of species, from oligomers to mature fibrils (Lewis, Walden, & Hydrophobic, 2014; Tipping, van Oosten-Hawle, Hewitt, & Radford, 2015). Extensive research has been conducted on β_2 microglobulin regarding the distribution of fibril length and the relation of the length to both cytotoxicity and the seeding of new fibrils post-fragmentation (W. F. Xue, Hellewell, et al., 2009). The length distribution was

characterised using tapping-mode atomic force microscopy (TM-AFM) and single particle image analysis(W. F. Xue, Homans, & Radford, 2009). As well as this, the ability of fibrils to seed the growth of new fibrils was analysed through the controlled fragmentation of preformed fibrils. It was shown that fragmentation of fibrils enhanced the seeding of new fibrils and that smaller fibrils acted as fibril seeds that nucleate the growth of new fibrils efficiently; reducing the initial lag phase seen in monomer-initiated fibril formation. As fragmentation continues the seeding efficiency of new fibrils increases in a manner that is directly proportional; as fibrils fragment it effectively creates an increased number of fibril particles to act as seeds for subsequent fibrils(W. F. Xue, Homans, et al., 2009). Research by the same team showed that the addition of the fragmented, short fibrils to liposomes that contained 50nm carboxyflourescein, resulted in effective membrane disruption that was detected through the release of the encapsulated dye. The longer fibrils (>1000nm) used did not disrupt the liposome membrane as effectively as the short fibrils did (<400nm), suggesting that short fibrils are more cytotoxic than the longer species(W. F. Xue, Hellewell, et al., 2009). One explanation for this, is that due to the higher number of individual fibrils in a population of short fibrils, the overall surface area of fibril ends and surfaces along the fibril axis results in a higher number of interactions with the targeted membrane (W. Xue, Hellewell, Hewitt, & Radford, 2010) further experiments looking at cell viability using MTT assays adhered to the idea that shorter fibrils display enhanced cytotoxicity compared to longer fibrils. While it has been shown that shorter fibrils are very cytotoxic, the exact method for this is not fully understood with various mechanisms proposed. However generally it is agreed that the mechanism which does take place involves the disruption of cellular membranes (Glabe, 2006). Disruption of such membranes may occur permeabilising them as well as rapidly increasing the

intracellular cytosolic calcium ions. One such mechanism of this is defined by the creation of pores via membrane thinning and is termed the “channel hypothesis” (Glabe, 2006). Some fibrils are developed intracellularly while others are extracellular, both of which may disrupt intracellular membranes, with the extracellular fibrils needing to undergo fragmentation so they are small enough to be internalised by the target cells. Both internal and externally produced fibrils can target the plasma membrane, which provides a further target for fibril cytotoxicity. Once the membrane is permeabilised, the sudden and rapid increase in calcium ions is not blocked by cobalt. This suggests that the influx is not due to activation of the cells endogenous ca^{2+} channels and therefore an invasive mechanism orchestrated by created by amyloid particles. One Hypothesis for the increase in Ca^{2+} ions, is through the liberation of sequestered Ca^{2+} (Glabe, 2006).

The common cross-β core structure of amyloid

Amyloid fibrils share a common structure in the form of cross β architecture (Jahn et al., 2010a). The cross-β architecture is a repeating quaternary structure that is composed of β strands that are orientated perpendicular to the long fibril axis, the β strands stack together to form a β sheet of indefinite length. Therefore, amyloid fibrils are composed of a highly-ordered arrangement of many repeating units of a protein or peptide. These β sheets will then associate laterally with one another to form protofilaments, which are the constituent units of amyloid fibrils. This architecture has been supported by structural data from experiments using x-ray diffraction. Amyloid fibrils give rise to a characteristic X-ray fibre diffraction pattern with a meridional diffraction pattern at ~4.7Å which displays the interstrand spacing between the stacked repeating β strands. An equatorial diffraction pattern at ~11Å corresponds to the lateral spacing between associated β sheets (Makin, Atkins, Sikorski, Johansson, & Serpell,

2005; Morris et al., 2013; X. Wang & Chapman, 2008). This Information can then be used on well-aligned fibrils to model low resolution structures of the fibrils, using the additional information that all amyloid fibrils have a translational symmetry which is parallel to the fibril axis, and most have a further helical symmetry. Determining the 3D structure of amyloid is challenging. X-ray crystallography is only viable for crystal like fibrils of short amyloidgenic peptide sequences based on amylin, tau, Aβ and Sup35. An extensively studied example of a solid-state NMR 3D structure based on crystal like fibrils of a complete amyloid fibril, is that of the prion HET-s (Wasmer et al., 2009). Research done using X-ray crystallography on the sequence GNNQQNY (Sup35)(Van Der Wel, Lewandowski, & Griffin, 2007) supports the idea of laterally associated β sheets, through a "steric zipper" model (Nelson et al., 2005)(Greenwald & Riek, 2010). This model is characterised by interfaces between β sheets that are devoid of water molecules, so are dubbed "dry" interfaces, and it involves the interdigitation of complementary side chains resulting in a high peptide packing density. The alternate model of interaction is a "wet interface" (Nelson et al., 2005) and this involves hydrogen bonding between side chains and the interface includes water molecules, which is a type of interaction that is familiar in protein crystal structures. Both these interactions occur between a pair of β sheets, and the interactions can be polar with hydrogen bonding and polar stacking and non-polar with aromatic stacking and van der Waals and peptides can have a combination of both wet and dry interactions (Nelson et al., 2005).

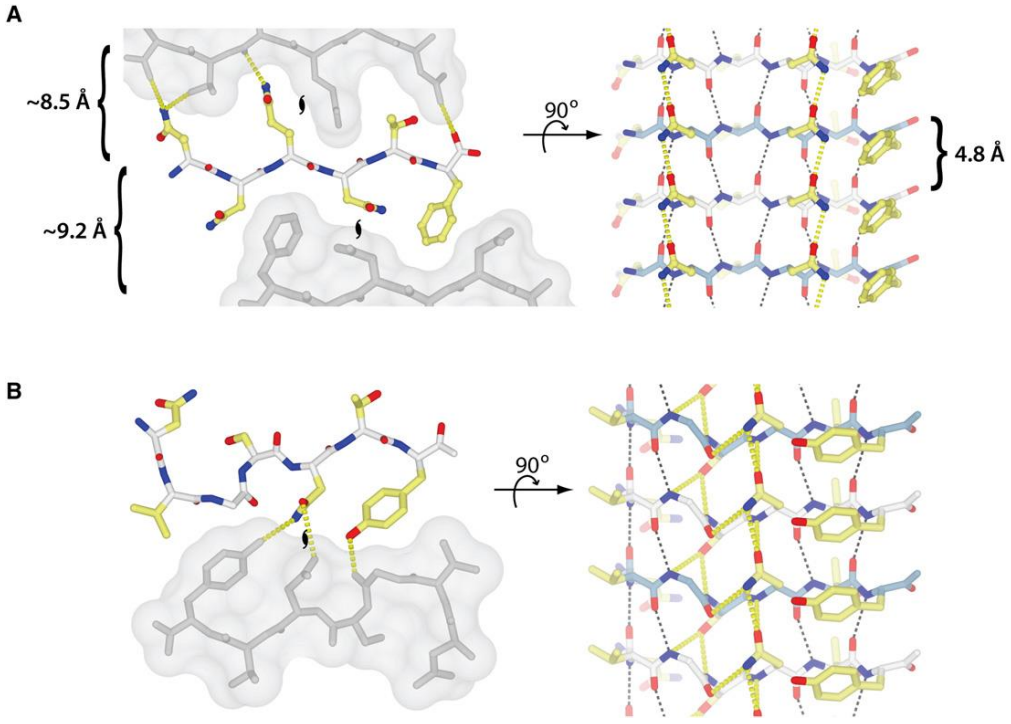


Figure 1. The wet and dry interfaces of a steric zipper interface.

Both images on the left are views looking down the fibril axis, and they show the interactions between β sheets. Image A is representive of the dry steric zipper model, showing the tight interdigitation between the side chains. The image B shows the wet interface, and hydrogen bonding between the side chains is represented by the yellow dotted lines. The images on the right are the views looking at the side of a single β sheet. Intrasheet hydrogen bonding is present in both models, and is represented by the black dotted lines. A shows the intra strand aromatic stacking of the Phe ladder, as well as the polar stacking of the two Asn ladders. Image B shows Asn interactions with an oxygen atom on residue 3. Adapted from (Greenwald & Riek, 2010)

The structural complexity of amyloid filaments can vary, and can be considerably more complex compared with the basic structure involving a pair of antiparallel in-register β sheets composed of a linear arrangement of short repeating peptide β strands (Jahn et al., 2010a; X. Wang & Chapman, 2008). The structure of amyloid fibrils of Alzheimer's Aβ has been extensively researched and is a good example of an amyloid fibril that generally conforms to the standard model of fibrils. Supported by X-ray diffraction patterns in sync with what would be expected (meridional diffraction of ~4.7Å and equatorial diffraction of 10-11Å) (Louise C Serpell, 2000). Further studies using quenched hydrogen/deuterium-exchange NMR derived complex 3D structures that

detail the intermolecular and intermolecular arrangement of $A\beta_{1-42}$ protofilaments (Ritter et al., 2005).

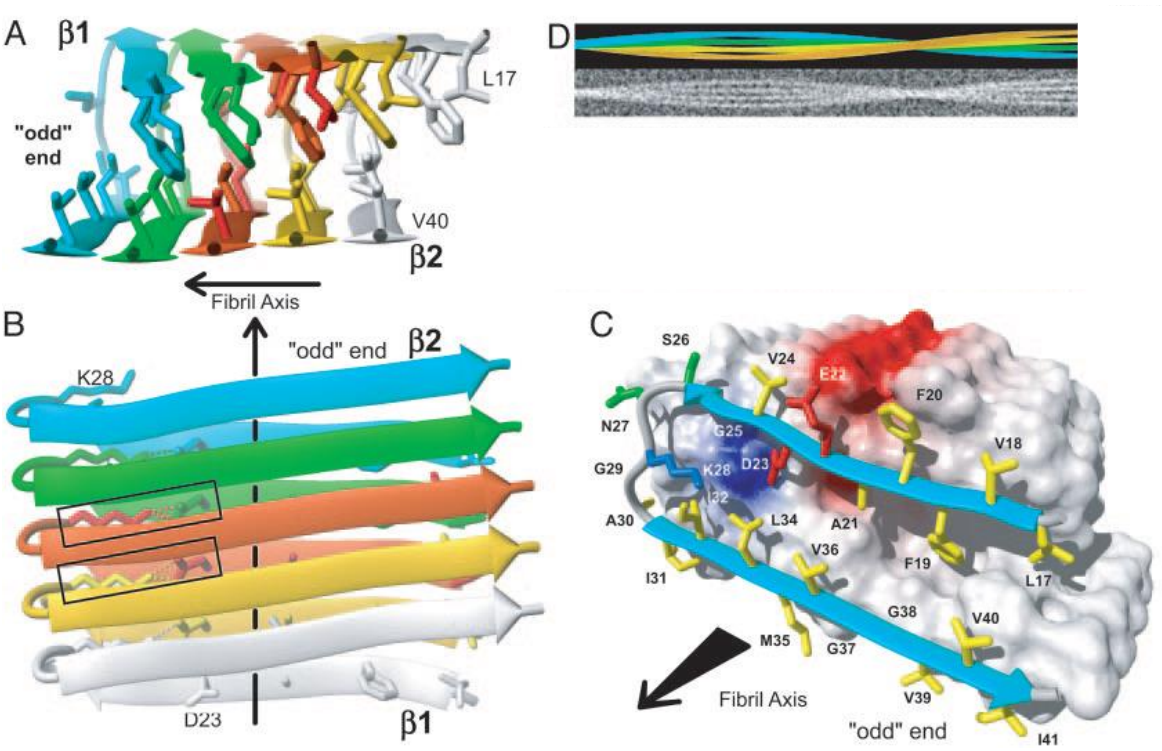


Figure 2. The molecular interactions at the core of Aβ

Images A, B and C show the intra and intermolecular interactions that stabilise and give Aβ protofilaments its sandwich β-sheet structure. Fibrils are composed of multiple protofilaments associating and is represented in D with each separate protofilaments represented by a different colour. Adapted from (Ritter et al., 2005)

Structural models of $A\beta_{42}$ protofilaments show two stacked, in register, parallel β sheets with intermolecular interactions. Each $A\beta_{42}$ monomer has a β strand – turn – β strand conformation, which gives rise to a structure with two parallel β sheets through β strand stacking (β1 and β2). B1 is composed of residues 18-26 and β2 is composed of residues 31-42 with the first 17 residues being highly disorders. The intermolecular interactions occur between the nth strand of β1 and the n-1 strand of β2, which results in partially unpaired fibril ends (Ritter et al., 2005). Figure 2B shows a key intermolecular interaction between the n and n-1 strands of sheets β1 and β2 in the form of a salt

bridge between residues D23 and K28. Figure 2C shows visual representation of the van der Waals interactions between side chains, of the end molecule of a protofilament. $A\beta_{42}$ fibrils are composed of multiple filaments associating and is represented in figure 2D with each separate filament represented by a different colour. The interactions between β sheets characterised here with Aβ are what define the core structure shared between amyloid fibrils of different proteins, however the number of laterally associated sheets and the orientation in which they are arranged within a protofilament can vary greatly (Jahn et al., 2010a). The intramolecular hydrogen bonded β sheet arrangement and intermolecular interactions of in register repeating β strands is the core structure shared among amyloid fibrils. In more complex arrangements, several β sheets may all twist around a central axis and the core structural interactions of amyloid are still conserved (Jahn et al., 2010a). An example model of an amyloid with more than two β sheets per protofilament is the prion protein HET-s, which is a left handed β solenoid with 4 β sheets, 3 in which form a distinctive hydrophobic triangular core, with the 4^{th} as part of the c-terminal region which is arranged as an extension to this triangular core(Van Melckebeke et al., 2010). Each HET-s monomer will complete 2 helical turns and include specific intra-molecular interactions and side chain packing that helps stabilise the overall structure. Additionally, each monomer will interact with its adjacent monomer through intermolecular side chain packing and hydrogen bonding. Pi – Pi interactions and Asn ladders assist in the overall stability and structure of the complete protofibril, which altogether adheres to the idea that all amyloid structures share the same types of stabilising interactions, in an arrangement of interacting cross β sheets, composed of in-register β strands. While 4 β sheets are present in the full-length fibril, in each monomer there are only 2 β strands per helical turn, with 4 strands overall but due to the helical turns 2 pairs are in register. All four of

these β strands are broken by a kink which consists of a 2 residue β arc (Wasmer et al., 2008a), which effectively separates them into 2 separate strands dubbed β1a β1b, β2a β2b, β3a β3b, β4a and β4b (Wasmer et al., 2009). The arrangement of the β strands and how the monomers polymerise into a protofibril can be seen in figure 3. Figure 3 also shows how a HET-s protofibril will alternate between inter and intra molecular interactions. Intermolecular side chain interactions occur between β strands 2 of one molecule and β strands 3 of the neighbouring molecule. Intra molecular side chain interactions occurs between β strands 1 and 4 of the same molecule.

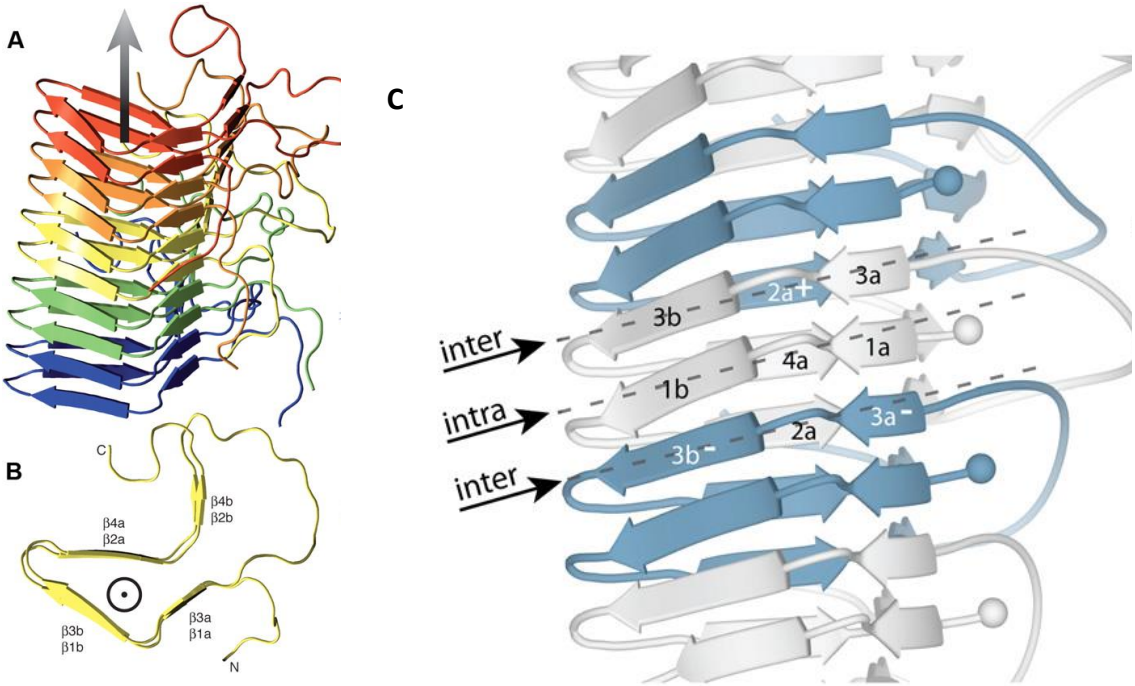


Figure 3. Inter and Intra-strand arrangement of HET-s filaments.

A and B (Wasmer et al., 2008a) *Shows the structural arrangement a polymerised protofibril and the separate β strands respectively. C shows the interactions between and within the separate molecules of a HET-s filaments l*(Greenwald & Riek, 2010).

In summary, the structural model for HET-s shows a left handed β solenoid core, with an extended structure that includes 6 β arcs, 2 salt bridges and 3 Asn ladders(Van Melckebeke et al., 2010). This structure varies from Aβ in its complexity, however the types of interactions are shared.

Nucleation dependent elongation defines amyloid formation

Amyloid fibril assembly is a nucleation dependent process. This is likely dependent on the formation of an oligomeric intermediate (nucleus) able to subsequently elongate through seeded growth at fibril ends (Andrews & Roberts, 2007; Linse et al., 2007; J C Rochet & Lansbury, 2000; W.-F. Xue, Homans, & Radford, 2008). A key feature of this process, regarding the kinetics, is the lag phase in which little or no fibril concentration can be detected. This phase can be circumvented through the addition of preformed fibril seeds, which are pre-formed fibrils. Therefore, it is likely that the lag phase is attributed to the formation of oligomeric structures that seed the elongation of fibrils (W.-F. Xue et al., 2008). The subsequent elongation phase is when fibrils begin to form through elongation at fibril ends. The lag phase is of interest, as it involves the conformational change of native proteins into the conformers that propagate Fibrillogenesis. The exact species present during the lag phase has been investigated extensively in the past, and it has been previously proposed that the nucleating species is the monomeric form of the protein, and this was determined through kinetic studies that looked at the protein concentration dependence of the rate of fibril formation (Ferrone, 1999). Analysis of such studies, showed a very small concentration dependence on the reaction progression, suggesting the size of the nucleus to be very small or monomeric(Pellarin & Caflisch, 2006). More recently however, the nucleating species has been further investigated due to increasing evidence of large amounts of oligomeric particles in the lag phase. These Oligomeric particles are intermediates between the monomeric form and the fibrillar form, and are well documented to be cytotoxic in their nature (Collins, Douglass, Vale, & Weissman, 2004; W. F. Xue, Hellewell, et al., 2009). Oligomers are considered an intermediate stage, with their pathway proposed as independent and competitive with the mechanism of monomer

seeded fibril growth (Collins et al., 2004). This suggests the possibility of more than one pathway of amyloid formation. In 2008 the development of a new strategy to analyse the kinetics of fibril formation has shed some light on the role of the oligomers(W.-F. Xue et al., 2008). The method uses a wide range of β2-microglobulin concentrations and involves the global analysis of a high number (>200) of fibrillation progress curves followed by systemic testing and ranking of many possible assembly mechanisms. The key measurements of the study being the time to reach 50% reaction completion and the length of the lag phase. Again, like the study in 1999 (Ferrone, 1999) the protein concentration dependence on the rate of fibril formation will help determine the size of the nucleus species. From this, it was determined that while monomer concentration is a contributor to the lag phase, these monomer species aggregate to form an oligomeric structural nucleus which is similar in size to a hexamer.

The multistep pathway from monomer form, to fibrillar form has been investigated using other qualative approaches also. The conformational changes and fibril formation of bovine β-lactoglobulin has been analysed using time resolved luminescence on molecules labelled with thioflavin T (ThT) and 1-anilino-8-naphthalenesulfonate (ANS) (Giurleo, He, & Talaga, 2008). ANS binds to hydrophobic regions of proteins, and its fluorescence is influenced by the specific interactions with amino acids and the polarity of its binding environment (Giurleo et al., 2008). Due to the contribution of hydrophobic driving forces during aggregation, the binding sites of ANS are structurally changed, providing clarification on the involvement and structural conformation of an amyloidgenic proteins hydrophobic core during fibril formation. When ThT binds to β sheet structures the absorption band changes to 450nm, therefore allowing the detection of a proteins conversion to amyloid (Giurleo et al., 2008). It seems that early lag phase aggregation displays reversible aggregation of the monomers into disordered

dimers and tetramers. Further aggregation is detected into an oligomeric state termed AggA. AggA shows altered location of ANS binding, suggesting conformational changes and rearrangement of hydrophobic regions of the protein. However, there is no ThT binding activity suggesting that this conformation is not rich in cross-β structure. Subsequent progression presents stabilising conformational changes in AggA, resulting in a second species of intermediate oligomer, termed AggB (Giurleo et al., 2008). AggB displays less dissociation of the aminosolanized surface so therefore seems more stable. This stable conformation displays high levels of ThT detected cross-β structure. Therefore, it seems that the lag phase involves two distinct oligomeric intermediates that are formed through the aggregation of monomeric species. Suggesting amyloid formation does not involve homogenous nucleation, but involves the sequential formation of multiple species. This proposed mechanism can be described through the process of colloid aggregation (Rohit V. Pappu, Xiaoling Wang, Andreas Vitalis, 2008). The driving force in colloidal aggregation is attributed to the energetically favourable relocation of amino acids unfavourable to protein-water interactions into regions of only protein-protein interactions i.e. the hydrophobic regions. While amino acids favourable to protein-water interactions are partitioned to the amyloid surface, suggesting this partitioning of amino acids is responsible for the conformational changes that results in a β-sheet rich nucleus. Some proteins which are below a certain size or have an unaccommodating sequence for an amyloidgenic conformational change; are unable to fold into stable conformations, which may be attributed to the fact that there are low numbers of thermodynamically favourable interactions to give the molecule stability. Through monomer association, the protein is able to effectively increase the molecular weight and increase the number of favourable interactions to the point where the β rich structure of amyloid is accessible (Giurleo et al., 2008).

Therefore, the lag phase is defined by the following stages; the initial instability of the monomer allows the favourable formation of an aggregated oligomer, thus giving the molecule a larger hydrophobic core, with favourable interactions. This therefore allows the molecule to rearrange into a second conformation that is sufficient in size and stable enough to contain rich β sheet structure. This then acts as the amyloid nucleus, which will polymerise to form amyloid seeds with the same β sheet rich structure, giving rise to the growth phase that involves the elongation of the seed through the addition of monomers which results in the formation of protofilaments. Association of more than one protofilaments with one another will form a filament which subsequently associate to form a mature fibril. A mature fibril's hierarchal structure and morphology is determined by the number of associated filaments.

An introduction to Prions

The mechanism for fibril formation and the core structure of amyloid is likely shared amongst all amyloidgenic proteins. However certain amyloid proteins present additional characteristics, creating a functional division for classification. Prions are a form of amyloid that can propagate mis-folding and adoption of the amyloid fold through their transmission between host cells. This characteristic describes them as infective proteins, with the ability to facilitate non-Mendelian transfer and induction of phenotype as epigenetic elements. Traditionally, the progeny inheritance of phenotype is a Mendelian process involving DNA encoded transfer of dominant and recessive genotype. Gregor Mendel proposed this theory in his description of the inheritance from an organism to its progeny, and attributes DNA as the molecule responsible. While natural reproduction is well known as a process involving the transfer of genotype, other infectious mechanisms also involve the exogenous modification of DNA. For example, viruses' method of reproduction involves the hijacking and reprograming of host cell

machinery through introduction of viral DNA and subsequent incorporation and modification of the hosts DNA. This is therefore a manipulation of an organisms' traits, through the application of DNA molecules. Therefore, when a protein only mechanism was proposed for the disease propagation of scrapie; a spongiform encephalopathy, it was met with much resistance in the scientific community (Prusiner, 1982). The process was suggested to involve the transfer of protein between cells, propagating the disease and resulting in the observed neurodegeneration. Termed Prion, it was derived from the words proteinaceous and infectious. In addition to the involvement of prions in scrapie, they have also been linked as causative factors in bovine spongiform encephalopathy and Creutzfeldt-Jakob disease. Following their suggested involvement in these disease processes, they were characterised and shown to conform to the core amyloid structure (Prusiner et al., 1983). Yeast Prions have been invaluable models for investigating the non-Mendelian propagation of prion phenotype, with one of the earliest phenotypes; [*PSI+*], reported in 1965 and observed in *Saccharomyces cerevisiae* (B. S. Cox, 1965). This phenotype is displayed in conjunction with the amyloid aggregate formation of the prion protein Sup35, and this model in yeast has been used extensively to understand the initiation and propagation of the prion state. The [*PSI+*] state only occurs upon the aggregation of the misfolded translation termination factor Sup35; the phenotype can be propagated from mother to daughter cell through the action of small amyloid particles known as propagons. These are fibrils of reduced length that have been fragmented through a chaperone (e.g., Hsp104) mediated mechanism (Byrne et al., 2009)

Functional amyloid: the other side of amyloid

In addition to the disease associated amyloid species, certain amyloidgenic proteins are deemed beneficial to the survival of the respective organism. Such amyloid is considered 'functional amyloid'. The term is ambiguous given that arguably all amyloid enacts a function, however the key principle for functional amyloid should be considered as the functional benefit conveyed by the protein. An example of a functional amyloid is the yeast protein Ure2p which propagates the [*URE3*] prion state. A protein of *Saccharomyces cerevisiae;* Ure2p regulates nitrogen catabolism through the interactive function of transcription factor Gln3p GATA (Wickner, 1994). Through the transcriptional inhibition of nitrogen permeases, monomeric Ure2p binds to Gln3p GATA and results in the cell unable to uptake and metabolise poor nitrogen sources. However, the amyloidgenic form of Ure2p and the induction of the [*URE3*] prion state results in a conformational inability of Ure2p to bind to Gln3p GATA (Wickner et al., 2004). This means that transcription of permeases is not inhibited, providing the ability of cells to utilise poor nitrogen sources; providing them an enhanced survival phenotype in environments lacking optimal nitrogen sources.

An additional and well known example of functional amyloid is a protein found in the spider silk; dragline silk. This example is definable as a biological nanomaterial, given its structural function. The fibrillar component of dragline silk, is spidroin and this protein aggregates into fibrils on a pH dependent basis aiding in the control of its amyloidgenic potential (Askarieh et al., 2011). Figure 4 displays dragline silk amongst other materials in regards to their tensile strength and elasticity (young's modulus). Included here are amyloid fibrils, which show relative variation in both properties; an inherent trait resulting from structural polymorphism. Dragline silk, and amyloid fibrils in general, can be observed as having a tensile strength comparable to steel and an elasticity

comparable to wood. The tensile strength and flexibility of amyloid therefore grants it high esteem as a structural nanomaterial; and enhances the ability of spider silk to effectively trap prey (Knowles & Buehler, 2011). Therefore, amyloid presents the ability to be applied in various functional material roles. Considering amyloid's efficiency of aggregation with tight levels of control, and the cost-effective production of starting material; amyloid has been the nanomaterial of choice in the development of nanomechanical projects such as the development of gold labelled conductive nanowires (Scheibel et al., 2003). This highlights an additional benefit of amyloidgenic nanomaterials; the ability to manipulate the machinery of organisms to engineer the construction of nanomaterials autonomously or with simple induction.

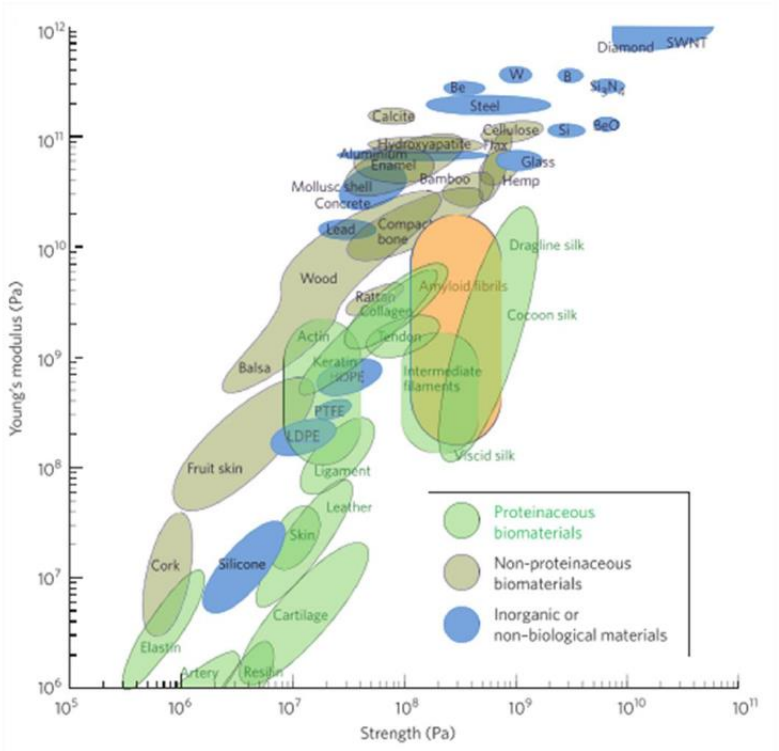


Figure 4. Mechanical comparison between organic and inorganic materials.

A graph providing a mechanical comparison between various organic and inorganic materials; including proteinaceous and biological material and non-biological materials. Amyloid is highlighted as having considerable mechanical properties (tensile strength and Young's modulus). Adapted from (Knowles & Buehler, 2011)

Fibril polymorphism, and hierarchal organisation of protofilaments.

Amyloid fibrils often display a structural polymorphism, meaning that while the core structural interactions are shared, there is variation in the overall morphology on both the molecular level, and the structural arrangement of interacting protofilaments (Fitzpatrick, Debelouchina, Bayro, et al., 2013). In the case of disease related amyloid, it is thought that since the transition of a polypeptide into its amyloid form is through un-controlled protein mis-folding events, the final structure may not represent the entirety of the accessible structures. Additionally, within that type of amyloid there may be variations (Chiti & Dobson, 2006). The structure adopted by many the amyloidgenic proteins may be representive of the local minima of the aggregation energy landscape. Further to this, the actual adopted structure may be influenced by the environment, and certain conditions may lead to different morphologies through changes in the energy landscape (L. Wang, Schubert, Sawaya, Eisenberg, & Riek, 2010). There are two key types of polymorphism in the molecular structure of protofilaments, and they both relate to the side chain packing of the β strands within a sheet. The first type is packing polymorphism, which arises when β sheets display either parallel or antiparallel arrangement of the β strands, however the residues involved in the side chain packing of the steric zipper model are identical. The second type is segmental polymorphism, and this arises when there are variations between the stretches of residues that are involved in the cross β core. Additionally, while fully assembled fibrils are made up of several associated protofilaments, Fibrils display polymorphisms regarding this ultrastructure. This type of polymorphism, is known as assembly polymorphism and is characterised by the number of associated protofibrils per mature fibril (Milanesi et al., 2012). EM and AFM images are able to differentiate between polymorphs of a single species of fibril, by identification of the number of filaments per fibril, degree of twisting

(via periodicity) and the length and diameter of fibrils (Jozef Adamcik & Mezzenga, 2012; Fändrich, Meinhardt, & Grigorieff, 2009; Fitzpatrick, Debelouchina, Bayro, et al., 2013; W. F. Xue, Homans, et al., 2009). Using NMR, cryo-EM and Atomic force microscopy, transthyretin (TTR) fibrils have been characterised extensively in regards to the interactions that result in the morphology and polymorphisms of amyloid fibrils (Fitzpatrick, Debelouchina, Dobson, & Et.al, 2013). From this it seems that the association of protofilaments with each other is dependent on the interactions ranging from β strands in adjacent protofilaments, as well as dipole-dipole interactions and van der Waals forces that while weak and non – specific, translate into potent interactions given the repeating nature of the fibrils, and the therefore repeat-induced cooperatively (Fitzpatrick, Debelouchina, Dobson, et al., 2013). The actual architecture of a mature fibril involves the association of multiple protofibrils that most commonly form left handed helical ribbons, which are double layered with a hollow core which is the axis in which the fibril will twist around. The two layers are identified by the electron density they express, with the high regions corresponding to the β sheets of each individual protofilament, and the central core being a region of low electron density. It seems, that fibrils tend to adopt doublet, triplet and quadruplet architecture. In each case, the number of protofilaments involved (due to the double layer) is 4, 6 and 8 respectively(Fitzpatrick, Debelouchina, Dobson, et al., 2013).

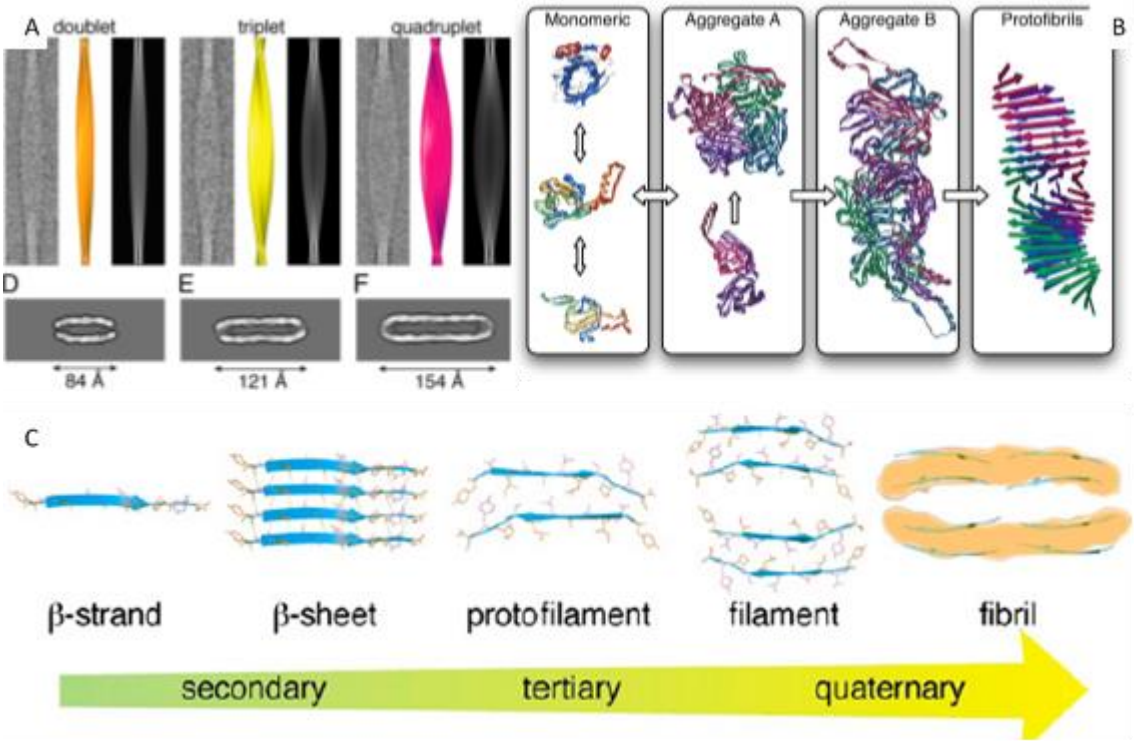


Figure 5. The hierarchical assembly of amyloid fibrils.

A) Representative Cryo-EM images of each class – (A) doublet, (B)triplet and (C)quadruplet showing; averaged fibrils left image, 3D reconstruction centre image, 2D projections of fibril reconstructions right image, and electron density cross-sections of the three types of fibril formed by TTR D, E and F for doublet triplet and quadruplet respectively(Fitzpatrick, Debelouchina, Dobson, et al., 2013) *. B) outlining the multistep pathway of nucleation dependent fibril formation*(Giurleo et al., 2008)*. C) Schematic displaying the hierarchical assembly of amyloid fibrils, from the primary constituent to a mature assembled fibril* (Fitzpatrick, Debelouchina, Dobson, et al., 2013)

It seems that the inter protofilament packing interface is made up of a number of repeating interactions, between adjacent β strands of separate protofilaments, in a head to tail fashion. The fibrils display an equatorial repeat of ~37Å which corresponds to the length of individual peptides of TTR, which corresponds with a lateral association of β strands and therefore protofilaments(Fitzpatrick, Debelouchina, Dobson, et al., 2013). Using isotopic labelling techniques with rotational-echo double-resonance, strong dipole interactions were identified between the N and C termini of adjacent molecules, suggesting the protofilaments do in fact interact in a head to tail packing interface between individual β strands. In addition to this Hydrogen bonding between

the terminal C=O and N-H groups and between the OH and O groups of tyrosine groups on adjacent molecules, meaning fibrils are in large stabilised by an array of repeating hydrophobic interactions, which makes sense given the large amount of exposed hydrophobic residues. Since these interactions are seen within the doublet triplet and quadruplet, it seems that the same interprotofilament interactions are conserved amongst assembly polymorphs. It also seems that given the presence of the two adjacent electron density regions, the electrostatic repulsive forces are overcome by the interprotofilament interactions. While the filaments of a fibril can overcome the repulsion, it is possible that the reason that the fibrils form a twisted ribbon is to minimise total free energy. If the distance between the axis of individual filaments is below a certain threshold which is maximised by the number of filaments in a fibril, and by an angle in which consecutive layers are tilted in respect to one another, then the filaments remain stable and will not shear off and break because of the repulsive forces. Therefore the most energetically stable conformation which minimises repulsion, is to have the layers of protein rotated in regards to one another, hence the twisted ribbon(Fitzpatrick, Debelouchina, Dobson, et al., 2013).

Predicting amyloidgenic sequences

The self-assembly of proteins into amyloid, is a phenomenon that results in the generation of both functional disease associated amyloid. Therefore, while it is important to study the structure of these amyloids, it is also crucial to identify potentially unknown amyloidgenic sequences. It has been suggested that all proteins can form amyloid under certain conditions and in physiological conditions, the balance of free energy between the two conformations (aggregated and native) is very small. Advances in algorithm-based prediction systems can help identify likely amyloidgenic sequences. Certain short sequences have the propensity to form amyloid when isolated,

despite residing within a longer non-amyloidgenic sequence. *In vivo*, certain proteins such as the amyloid precursor protein (APP) can undergo cleavage events that segments the protein into shorter peptides. Therefore, it is important to be able to predict such peptide sequences within any protein. The additional benefit to this type of prediction is the ability to expand the sequence space of structurally characterised amyloid systems, and generate model peptides that can be used to understand the relationship between sequence and self-assembly, stability and structure(Morris et al., 2013).

The WALTZ algorithm (Maurer-Stroh et al., 2010), represents an effective method for the prediction of amyloidgenic sequences. WALTZ uses a sequence position scoring matrix, and can identify short peptide sequences that form ordered amyloid aggregates as well as better distinguish between amyloid and amorphous aggregates (Maurer-Stroh et al., 2010). Previous studies using model peptides identified by WALTZ, explored the relationship between sequences and the core amyloid structure (Morris et al., 2013). The hexapeptides HYFNIF and RVFNIM, and the pentapeptide VIYKI are WALTZ identified amyloidgenic sequences buried within larger sequences for the bloom syndrome protein (human), Elongation factor 2 (human) and chlorion protein (Drosophila), respectively. When these sequences are isolated, CD and FTIR spectroscopy (Morris et al., 2013) confirm that the WALTZ identified peptides assemble and adopt cross-β sheet structures, however with differing lateral packing. In their native Protein sequences, the structure of the Waltz sequence is highly dependent on the flanking amino acid sequences, and do not aggregate into amyloid. Therefore, it seems that flanking amino acids, protect against aggregation and self-assembly. The WALTZ containing proteins are not predicted to form amyloid; for example, natively, the bloom syndrome protein adopts a predominantly α – helical structure and the proteins 3-D structure may protect against potential HYFNIF facilitated aggregation, in which the

flanking sequences possibly act as ‘gate keepers’ (Reumers, Maurer-Stroh, Schymkowitz, & Rousseau, 2009). The WALTZ sequences represent the maximum amyloid forming frame for themselves, while also representing a general minimum number of residues required to form amyloid fibrils. Therefore, the amyloid core is both small, and restrictive in that expansion can inhibit the amyloid forming potential (Reumers et al., 2009). The term ‘gatekeeper’ used here in regards to amyloidgenic sequences describes residues that flank the amyloid forming core sequence, that inhibit amyloid formation through charge repulsion and inaccessibility to a β-sheet conformation (Reumers et al., 2009). Arginine, lysine, aspartate, and glutamate are all charged residues, and the presence of these residues as gatekeepers disrupt the ability of the core sequence to form amyloid. In addition to this Proline, disrupts amyloid formation though its inability to conform to β-sheet structure. The WALTZ sequences themselves are characterised in their propensity to form amyloid, using the position sensitive scoring algorithm that identified them (Morris et al., 2013). However, in addition to WALTZ is the TANGO algorithm which is considered as more effective at detecting the window of amyloidgenicity. It does so by scoring the amyloid propensity from 0-100%; the window is the region in which the residues score >0% and the window itself scores above >50%. Long stretches of incompatible or detrimental residues will reduce an expanded windows score, therefore the window sequence >50% when isolated will form amyloid (Fernandez-Escamilla, Rousseau, Schymkowitz, & Serrano, 2004). WALTZ however uses a fixed 6 residue window; large sequences can subsequently be scanned and scored per the position sensitive algorithm on their amyloidgenic propensity. WALTZ is most effective at differentiating between amyloid forming and amorphous aggregation prone sequences than TANGO, however.

It has previously been shown by the removal of any phenylalanine residue in amyloid-like crystals, inhibits the ability of assembly (Wasmer et al, 2008). Therefore, amyloid self-assembly is likely driven by or aided by the aromatic content within a sequence. It has also been suggested that hydrophobicity is an important factor in assembly and elongation of β sheets (Maurer-Stroh et al., 2010). This has been backed up further using HYFNIF and VIYKI (Morris et al., 2013). In addition, it seems that charge and aromaticity influence the inter-sheet interactions and the lateral association of protofilaments, which is an important factor in the modulation of the fibril architecture. Figure 6 represents the residue properties of the 3 WALTZ peptides; each of the peptides present different net charges, hydrophobicity and aromaticity. This likely provides a difference in the core structural interactions as well as the inter-filament interactions, which in turn account for discrepancies in the per-population fibril polymorphism. Each 'mature' fibril presents an array of positively charged side chains along the fibril axis (Lysine – VIYKI, Histidine – HYFNIF, Arginine – RVFNIM; Figure 8). However, they differ in their overall charge; HYFNIF which contains a charged residue and overall presents a net charge of 0.1, likely due to the fact Histidine is not consistently protonated at physiological ph. VIYKI and RVFNIM which contain only one polar residue to their one charged residue have an overall charge of 1. Each sequence contains 4 hydrophobic residues (figure 8) however they differ slightly in their hydrophobic potential due to differences in the per-residue hydrophobic strengths present in each sequence. RVNIM potentially has the lowest hydrophobicity with methionine being moderately hydrophobic at pH7; HYFNIF and VIYKI have almost identical hydrophobicity with phenylalanine and isoleucine being the most hydrophobic residues at pH7. Finally, the WALTZ peptides differ in aromaticity with VIYKI and RVFNIM containing a single aromatic residue; Tyrosine and phenylalanine, respectively. HYFNIF however, contains

three aromatic residues; two phenylalanine's and a tyrosine (figure 8). There is a favourable alignment of aromatic compounds upon the conformational change into the core amyloid structure, then the subsequent polymerisation favours aromatic ladders through the β-sheet's (Amylin, Tracz, Abedini, Driscoll, & Raleigh, 2004). Therefore, the aromaticity also likely plays a role in the polymorphic landscape.

Figure 6. Residue properties of WALTZ sequences

Colour coded sequences for each of the WALTZ identified peptides; Orange represents hydrophobic residues, blue represents non-polar residues and green indicates a polar residue. Red residues are charged and a green hexagon surrounds aromatic residues.

The WALTZ peptides have also been analysed as models for the nano structure and morphology of amyloid fibrils. The hierarchal structure of associated protofilaments of HYFNIF, VIYKI and RVFNIM show different morphologies that are likely dependent on the differences in molecular packing (C S Goldsbury et al., 1997). Qualitative analysis has shown each of the WALTZ peptides to form fibril populations that are polymorphic. For example, HYFNIF fibrils have been observed as displaying a twisted morphology between a pair of protofilaments, with a maximum width of 17.04nm and a periodicity of 95.16nm (Morris et al., 2013). VIYKI on the other hand displayed the least variation in morphology, in which all the fibrils were twisted rope morphology with a width of 21.0nm and a periodicity of 58.1nm (Morris et al., 2013). Long term incubation displayed further association of fibrils into the formation of tube like fibrils which are essentially nanotubes. Which itself is interesting as it suggests a progressive and dynamic morphological landscape, while also granting the short sequence access to structurally

desirable polymer structures. This has implications in the biotechnology field as the groundwork towards new self-assembling bio nanomaterial's.

Atomic force microscopy: An effective method for imaging amyloid

Visualisation of the hierarchal arrangement of a number of filaments into a mature fibril can effectively be characterised using atomic force microscopy(Jozef Adamcik et al., 2010)(Jozef Adamcik & Mezzenga, 2012). AFM works by physically interacting with a sample via a nanoscale tip attached to a flexible cantilever. The probe scans the sample surface and interacts with the sample on a force basis in which intermolecular attraction and repulsion causes deflection of the cantilever; representative of the samples height. This information is detected using a feedback mechanism in which a laser is serially reflected from the cantilever to a mirror and finally into a position sensitive photodiode. The lasers movement from a pre-calibrated position relays the degree of cantilever deflection, which is used to generate 3-Dimensional images that represent a sample's topology.

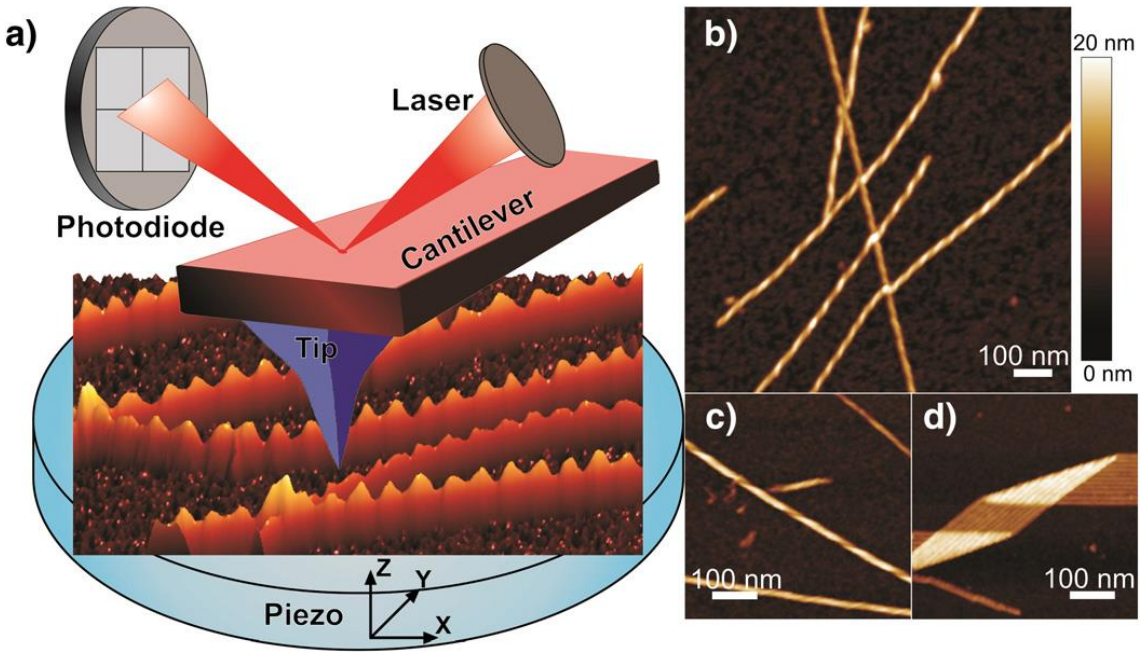


Figure 7. The operation of Atomic force microscopy.

A) Represents the mechanism in which AFM collects data from a sample. B), C) and D) display examples of images captured by atomic force microscopy (Jozef Adamcik & Mezzenga, 2012)

This information is acquired based on the principle that the cantilever under operation obeys Hooke's law, F= kx. In this application, F is the force experienced between the sample and the probe, K the spring constant of the cantilever and x is the deflection of the cantilever. The driving force applied to the probe can be controlled and determines the peak force in which it will interact with the sample, which in turn helps control and overcome both attractive and repulsive forces, respectively. Increases in peak force may also allow deformation as the probe indents the sample, providing information on the elastic properties of that sample. The scanning of the sample is facilitated by a piezo electric scanner which moves a sample mounted upon it, in 3-dimensional space and with nanometre precision. The AFM can operate in multiple software assisted scanning modes which are all subsidiary of the main operation modes; contact, non-contact and tapping mode. Tapping mode is arguably the most commonly used mode for imaging of samples based on their topology, and differs from contact mode in the fact it involves the mechanically induced oscillation of the cantilever. The amplitude and frequency of this oscillation can be controlled, and allows the acquisition of more accurate width measurements through a reduction in tip convolution. In addition, this image mode is superior in both sample and probe preservation as the interactions between sample and probe are intermittent and not continual such as with contact mode. An advancement in the tapping mode operation is peak-force tapping mode (PF-QNM). PF-QNM operates differently to tapping mode, including a tighter control of the oscillation frequency which is maintained at a level to actively reduce possible resonance. More importantly however, is the enhanced feedback system in which each sample measurement plots a force curve based on the force interactions on that contact. This repeated action provides a dynamic force curve through the scanning time course that allows for the readjustment of peak force in response to the sample; this differs to standard tapping

mode where the feedback system relies only on cantilever amplitude. Additionally, the incorporation of peak force measurements means that with pre-calibration on a known sample additional mechanical properties of an experimental sample can be obtained in real time; for example, the young's modulus and deformation derived elasticity(Kurland, Drira, & Yadavalli, 2012).

Through the height topology imaging technique of the AFM, many details about fibrils can be analysed using this method. For example, the height of single filaments and that of mature fibrils, can give insight into the packing mechanism and the morphology that incorporates filaments into fibrils. Therefore, the height of a fibril can give details in regards to the polymorphism of a fibril in regards to the number of filaments and the lateral assembly. The height profile of a fibril axis contains fluctuations in height that are indicative of various twisting morphologies. This also corresponds to the number and arrangement of filaments within fibrils in that the more constituent filaments, the longer the twisting pitch along the fibril. This is determined through the distance between peak heights along the profile and is known as the periodicity. This is represented in AFM images as the periodic variations in height along a given fibril, which when analysed on β lactoglobulin fibrils (Jozef Adamcik et al., 2010) was displayed on all fibrils over the height of 3nm. This implies that periodicity is displayed when 2 or more filaments are involved in a fibril. Through the visualisation of periodicity, it was shown that there are 4 common polymorphisms in regards to periodicity with β lactoglobulin fibrils, and these showed regular fluctuations of max height in intervals of 35, 75, 100 and 135 nm, with max heights of 4, 6, 8 and 10nm respectively. Therefore, it seems that periodicity increases with the height of the fibrils, and therefore since a higher fibril represents more filaments in a fibril, the higher the number of filaments means that the distance required for a fibril to complete a 360-degree twist increases.

The change in both fibril height and periodicity will conform to a linear incremental increase when assuming a twisted ribbon morphology (Jozef Adamcik & Mezzenga, 2012). For this work, fibril morphology will be broken down into 2 factors: archetype i.e. the twisting type such as helical ribbon. Then filament arrangement, which includes the number of filaments per fibril. Fibrils adopt morphologies with additional archetypes such as twisted ropes, helical ribbons and nanotubes (Usov, Adamcik, & Mezzenga, 2013). In such cases the height and periodicity relationship becomes more complex. Twisted ribbon polymorphism has been studied in regards to the heptapeptide CH3CONH-βAβAKLVFF-CONH2 (Jozef Adamcik, Castelletto, Bolisetty, Hamley, & Mezzenga, 2011). When determining the number of filaments involved in a fibril, the width and the twists of the fibril that is analysed. When scanning fibrils with AFM, images are liable to and artefact resulting from the tip geometry, known as sample convolution. This artefact results in images that are not accurately representative of the width of a sample, providing only an apparent width. The height of a sample however is not affected by convolution given the symmetry of a fibril, the height is therefore directly representive of the width. In addition, AFM can be used to determine the contour length of a fibril. This parameter is important when conducting length dependent fragmentation analysis and has previously formed the basis of studies determining the size of fragments and the relationship towards cytotoxicity(W. Xue, Hellewell, Hewitt, & Radford, 2010a). Additionally, the persistent length can be determined through AFM images via the bond correlation of fibrils for semi flexible polymers. Formally, persistent length is rate of decay of correlation between two tangent vectors across a polymer of x length. Informally it is the length in which when physically exceeded a polymer will display flexible behaviour, and when receded will behave rigid. All these properties obtained through AFM are related; the height is

indicative of polymer thickness and periodicity is representive of structural twisting, which both influence the mechanical properties of a polymer and the subsequent resistance to breakage. The mechanical properties such as flexibility can be defined by the persistence length, and is a characteristic that is dynamic in its behaviour; on a length-dependent basis (Relini et al., 2010).

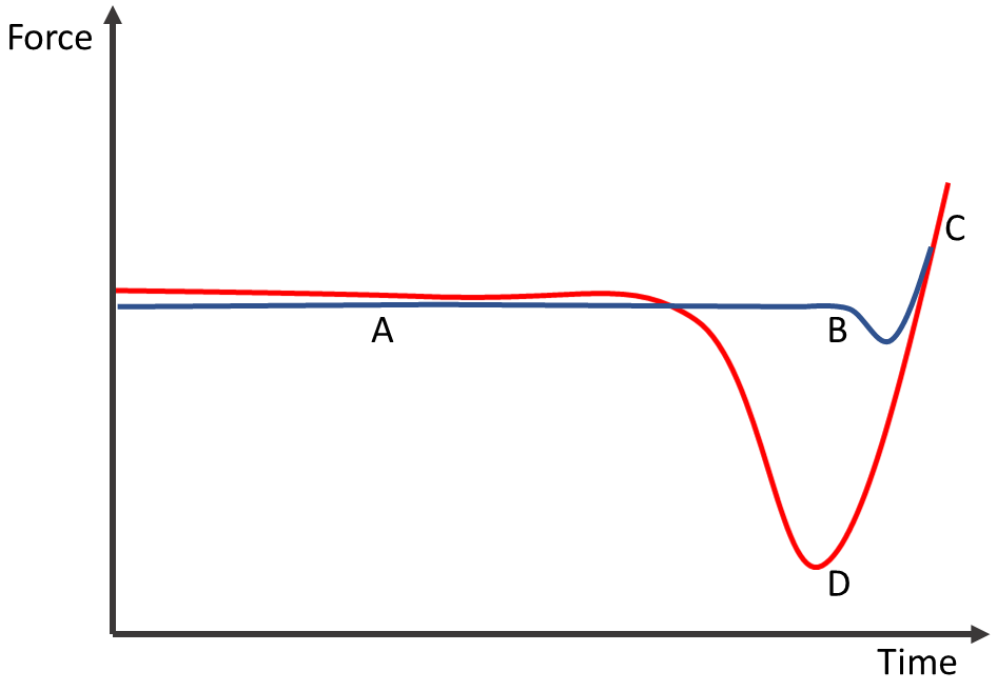


Figure 8. A dynamic force curve as seen in AFM operating software

A dynamic force curve as seen in AFM operating software. The graph allows for the monitoring of force interactions between the probe and the sample; A) represents the non-interactive trace(blue)/retrace(red) of the probe as it approaches a sample object. B) registers a negative force feedback as the probe indents the sample and experiences repulsion from the sample object. C) The peak force is reached, which is used for adjustment within the feedback system and triggers the retraction portion of the cycle. D) as the probe retracts from the sample, it experiences adhesive forces from the sample and returns to its original position.

The AFM collects image data on a per-pixel basis in which each pixel is a sample data point attributed a numerical figure representative of the sample height. During tapping mode operation, each contact made with the surface acquires a measurement and represents a sample pixel. The number of samples taken per image can be controlled be inputting predetermined resolutions via the software. The piezo-electric scanner moves

the sample in horizontal lines allowing for a sample trace and then a retrace on the return movement. A horizontal movement back and forth once represents a line in the image and can also be controlled and is affected by the set resolution. Therefore, the resolution and subsequent level of detail obtained by the AFM is directly related to the number of sample data points taken per image; which is controlled by the number of lines per image, and the samples per line. Using peak force tapping mode, the visualised force curve is representative of the movement and interaction between the sample and the probe. Figure 7 displays a typical force curve, and represents the oscillation of the probe at its baseline position, the repulsion experienced by the probe on approach, the peak force applied to the sample and then the adhesion on experienced by the probe on retraction and return to the original position. The magnitude of each part of the graph may change depending on the peak force applied and the interactive forces experienced.

Aims of this thesis

This thesis aims to expand on the understanding of amyloid fibril polymorphism by presenting what we have called the “amyloid morphology landscape” which presents the range of structures adopted within a population of fibrils, representing the global minimum in free energy for accessible morphologies. This thesis presents the methods used to quantitatively characterise fibril morphology distributions and class boundaries using fibrils formed from short peptide sequeunces. The application of hierarchical clustering to achieve this analysis represents a novel method to characterise fibril structures and its full development from the ground up is detailed in this thesis. In addition to this, the relationship between sequence, structure and mechanical stability is expanded upon in this thesis. We present a relationship between mechanical properties such as persistence length, with the propensity of a fibril to break; therefore aiding the development of the link between structure and cytotoxicity.

Overal, this thesis aims to enhance the understanding of fibril structure and factors that may affect the morphological landscape of amyloid, which has crucial implications in the relationship between fibril structure and disease.

Chapter 2

Developing the methodology for imaging and quantitative analyses of amyloid structure and morphology

Preparation and imaging of amyloid fibrils by AFM

Fibrils formed from amyloidgenic proteins and peptides have been optimised for imaging with AFM here. α-synuclein and Sup35NM have been used in coordination with fragmentation of the WALTZ peptides. However, other amyloid fibril samples have also been imaged. Di-tyrosine cross linked α-synuclein has been imaged in collaboration for the paper (Al-Hilaly et al., 2016) as well as lysozyme and Aβ40 fibrils for the paper(R. A. S. Smith et al., 2015). Finally, Sup35NM fibrils formed using the C-DAG system in *E. coli* have been imaged for the structural comparison. This chapter reports on the methods and the optimised conditions required to image these respective amyloid samples and provides a detailed account of the methodology recommended for imaging a range of different amyloid species by AFM.

Generally, amyloid fibrils can be imaged using AFM under similar conditions in air. This may be due to the common intrinsic properties of amyloid fibrils charge based interaction with surfaces. In addition to this the hydrophobic nature of amyloid means that fibrils are insoluble, allowing relatively easy deposition on an imaging surface. For AFM imaging, the imaging surface must be atomically flat to allow for a flattened background which enables the accurate height measurement of the desired sample. Mica is commonly used for this purpose. Here we have used mica discs (Agar scientific, F7013) attached to steel discs (Agar scientific) allowing for magnetic based 3-dimensional manipulation of the surface by the AFM's piezo electric scanner. Mica itself is negatively charged, therefore positively charged residues of a sample polymer will interact with the surface enhancing deposition. For a standard procedure to 'scout' deposition conditions for imaging, each of the amyloid species imaged were initially prepared in the same way. The samples were prepared to a concentration of 0.02mg/ml

in either filter sterilized milli-Q water or in the respective filter sterilized sample buffer. Then 20ul of the sample is deposited onto a freshly cleaved mica surface and incubated for 5 minutes. The surface is then washed with 1ml of filter sterilized milli-Q water and dried using a gentle stream of nitrogen gas. The samples are then imaged using a Bruker Multimode AFM with a Nanoscope V controller operating under peak force tapping mode. ScanAsyst probes (silicon nitride triangular tip with tip height = 2.5-8µm, nominal tip radius = 2nm, nominal spring constant 0.4N/m and nominal resonant frequency 70kHz) were used unless where stated otherwise. The imaging software used is Nanoscope 8.15. Low resolution images for each sample are 20µmx 20µmin size and 512 x 512 pixels and were usually collected first. Images were subsequently collected at ranges of surface sizes and resolutions (1-50^{2}um surface size and 512-4096^2 pixels; figures annotated accordingly). This standard sample preparation and imaging procedure provides preliminary information regarding the optimal conditions for a sample. Subsequent optimisation of the preparation method achieves the desired surface population density. The optimum scanning parameters may vary depending on the sample/tip interaction.

Adjustments made are based on observations seen following this standard procedure, and usually involve changes in the ionic strength and pH of the deposition solution, to enhance interaction between the mica and polymers with a net negative charge. In addition, pre-treatment of the mica surface with positively charged ionic species such as Mg^{2+} facilitates the same enhancement without effecting the fibril solution conditions. Overall, parameters include the deposition concentration, volume, and incubation time. Additionally, the volume and solvent used for post-incubation washes can be adjusted. Each preparative parameter has been optimised depending on the samples imaged.

All images were processed using Nanoscope software (Nanoscope Analysis version 1.4 or 1.5, Bruker) to remove sample tilt and scanner bow. Third order baseline flattening was carried out in conjunction with threshold correction of the baseline where particles of interest are masked based on their Z coordinates. Streaks are also removed with a built-in function, which works by the approximation of pixel values based on adjacent pixel data. The appropriate height scale is selected and represented by an off-image colormap.

WALTZ

WALTZ Peptide synthesis

The WALTZ peptides were synthesised with an N-terminal acetylation and a C-terminal amidation. Multistage solid phase synthesis using Fmoc protection chemistry generated lyophilised powder which had a purity >95% by HPLC (JPT peptide technologies. Or synthesised on site (Kevin Howland, University of Kent).

In vitro polymerisation

To prepare samples of WALTZ peptide fibrils 1mg of the respective lyophilised powder is suspended in 100ul of 0.2µm syringe filtered filter sterilized milli-Q water to a final concentration of 10mg/ml. Due to the potential instantaneous formation of fibrils, the solution is not vortexed to mix, to avoid potential fragmentation of fibrils. The solution was mixed by pipetting to completely dissolve the peptide powder. The solution was incubated at room temperature for 1 week prior to experiments and imaging. This time allows for the ‘maturation’ of fibrils providing time for morphological adoption across all 3 WALTZ samples prior to imaging. All manipulations of the WALTZ peptide solutions were with low bind tips (VWR 53503-294 (with SpidersilkTM surface)), and stored in Protein LoBind 1.5ml tubes (Eppendorf, 022431081).

To prevent potential contamination of samples by microorganisms, parallel fibril formation test samples were either treated with 0.02% sodium azide or snap frozen and stored at-80°C immediately upon resuspension. While the treatment with sodium azide should not affect fibril morphology, there are potential issues with snap freezing of the WALTZ. The freezing could result in fragmented and unrepresentative morphologies, formed at low temperature. Therefore, it was not adopted as routine procedure for fibril storage. In addition to this, through careful aseptic and sterile techniques during sample preparation the addition of sodium azide was not adopted as routine either. While it is unlikely it will affect fibril morphology, a fibril forming environment of sterile filtered water provides one with the least potential variables within different batches.

Specimen preparation and AFM sample preparation for structural characterisation

Imaging conditions for each of the WALTZ peptides, have been extensively optimised. Initial imaging of the WALTZ peptides, was conducted as per the aforementioned methodology. This provided the images seen in figure 9 which displayed desirable distribution of fibrils, which adhered to the surface efficiently. However, samples prepared in this way did not provide a consistent population density, and sample surfaces of HYFNIF and RVFNIM notably displayed randomly distributed background particles. This unwanted background may present issues with accurate height measurements for structural characterisation. Therefore, the conditions were further optimised to minimise the background. The deposition pH was reduced to pH2, providing a higher concentration of H^+ ions which may interfere with any potential charge based interactions of the background material with the mica surface, however due to the unknown nature of the image background, the exact mechanism of action cannot be confirmed. It is possible that the H^+ ions neutralise potential interactions between the negative mica surface and the likely positively charged low molecular

weight material. This reduction in pH could be assumed to disrupt fibril binding in a similar manner. However, it is possible that due to an increased array of positively charged side chains along a full fibril axis (Lysine – VIYKI, Histidine – HYFNIF, Arginine – RVFNIM; figure 8) the increase in H+ concentration is not sufficient to inhibit arrayed interactions with the mica surface. In addition to this, the non-polar residues (figure 8) contain COO- groups which can act as proton acceptors, thereby reducing any electrostatic repulsion experienced between the polymer and the negatively charged mica surface. HYFNIF which contains a charged residue, overall presents a net charge of only 0.1 while VIYKI and RVFNIM contain only one charged residue and display a charge of 1. The lower iso-electric point for HYFNIF means that at pH7 it is closer to having a net neutral charge, as Histidine is only partially protonated at physiological ph. Only the asparagine will have surface exposed side chains as the Tyrosine is hydrophobic and likely contains its side chains in the fibril core. However, following a reduction in pH the net charge of HYFNIF becomes 1 as the Histidine residue population likely becomes fully protonated. Both VIYKI and RVFNIM have higher iso-electric points and maintain a charge of 1 with reduction in ph. While residues such as Valine, which are non-polar and neutrally charged, will accept protons and become positively charged in an acidic environment. They are also highly hydrophobic and likely have their side chains buried within the amyloid hydrophobic core restricting their H^+ mediated charge increase (Van Gerven, Klein, Hultgren, & Remaut, 2015). Therefore, it seems that the change to pH2 of the three fibril populations inhibits the deposition of background material as seen in

figure 9, while maintaining or enhancing the charge interactions between WALTZ peptide fibrils and the mica surface.

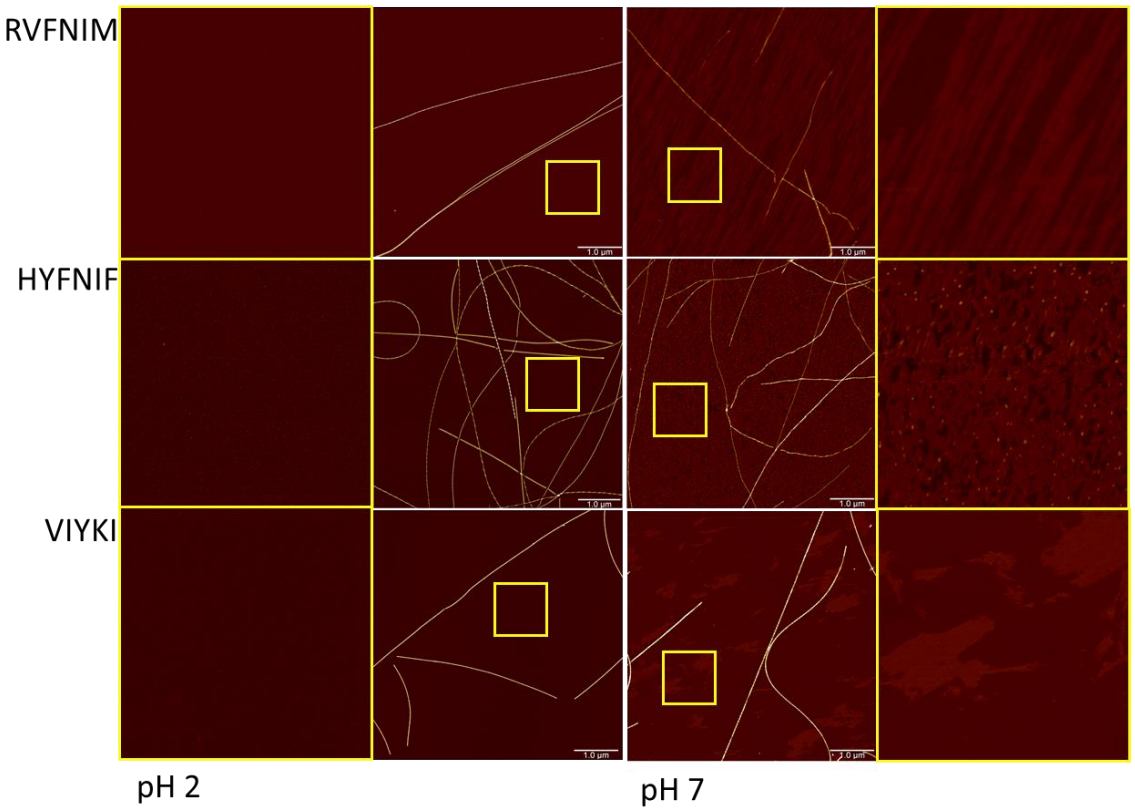


Figure 9. The effect of pH on background material of fibril samples

AFM images showing WALTZ fibril surfaces with a pH7 deposition (right) and a pH2 deposition (left). Flanking images framed in yellow represent zoomed sections of the respective central images (yellow squares). A pH 2 deposition provides a clear surface with no unknown background material which may reduce fibril height accuracy.

The pH change was achieved by making a 20ul deposition aliquot of each sample with 10% pH1 HCL (syringe filtered), to a final pH2.0. To ensure a desired fibril density the deposition concentration was 0.05mg/ml and the sample was loaded on to the mica surface and incubated for 10 minutes. The final deposition aliquot was made as a dilution from either 1mg/ml sample stock solutions or from 0.1mg/ml stock solutions. Fibril samples were stored at concentrations higher than 0.1mg/ml to minimise the fibrils binding to the plastic of Eppendorf tube. Higher concentrations stored in Protein

LoBind 1.5ml tubes (Eppendorf, 022431081) minimalizes the apparent reduction in a samples fibril density. Low bind tips (VWR 53503-294 (with SpidersilkTM surface)) were also used for all manipulations, to further reduce the likelihood of a loss in fibril particles through sample preparation. Finally, the sample is washed with 1ml of syringe filtered filter sterilized milli-Q water and dried using a gentle stream of nitrogen gas and edge absorption of residual water onto a piece of filter paper.

AFM imaging parameters

For initial imaging and optimisation of the WALTZ peptides, the standard imaging procedure was used as previously detailed. Using the nanoscope software, ScanAsyst automatic parameter adjustments were used. However, Optimised imaging parameters were used for the acquisition of images for structural characterisation which provided the required pixel density to resolve structural details within the fibrils. Within the nanoscope software, specific parameters were set manually to provide optimal image collection. Scan rate was set at 0.305Hz, noise threshold was reduced to 0.5nm while maintaining automatic gain adjustment. Additionally, Z limit was reduced to 1.5um and the peak force set point set to 0.05- 0.15mV (depending on sample), thereby allowing a steady peak force throughout the images; automatic adjustments in peak force set point when scanning promotes skipped lines or increased noise as the tips interaction with the sample changes and when at slow rates this provides an exaggerated representation within the image. However, if the samples within the scan area are relatively homogenous in terms of likely tip interaction (height and charge properties), this parameter can be left unchanged. For structural analysis of the images were captured at 6µm x 6µm at 2048 x 2048 pixels or at 4096 x 4096 pixels for an area of 12um x 12um and subsequently cropped. The images have a pixel density of 2.93 pixel/nm for the structural analysis. For fragmentation images, initially the scan size was 20µmx 20µm at

2048 x 2048 pixels; however, increasing this allows for larger per image particle number and maintains an adequate pixel density required for subsequent analysis. Therefore, fragmentation images were captured at 30µmx 30µm at 2048 x 2048 pixels. Alternatively, images were captured at 60µmx 60µm and a resolution of 4096 x 4096 pixels. For both image types, the scan speed of 0.305Hz or 0.203Hz for 4096 x 4096 images was used. This is relatively slow for images of this resolution and is not necessary when using ScanAsyst parameters, however it enables a more accurate re-trace providing more accurate apparent widths which is helpful for visual analysis of structural features. Additionally, the reduced scan rate accommodates for the reduced Z limit, which increases the chances of sample/tip interaction and therefore degradation of the tip quality. The reduced Z limit however, is beneficial as it allows for more accurate height measurements and apparent widths as the tip is maintained at a lower overall height from the surface. This parameter was reduced upon confirmation of the particle heights within the image. Finally, the noise threshold was set to a reduced value of 0.2nm which allows for cleaner images, and in turn maintains the automatic gain adjustments within a tighter range which provides a more uniform level of noise across the image. Images can be seen in figures 9 and 10. Image collection following this method was used for the analysis in chapter's 3 and 4.

Fibril Fragmentation

To analyse the reduction in fibril length over time of the WALTZ peptide fibrils under mechanical perturbation, we initially fragmented the fibrils using a stirring method. This was achieved using 500ul of sample at a concentration of 0.1mg/ml in filter sterilized milli-Q water (HYFNIF – 113uM, VIYKI – 148uM, RVFNIM – 120uM) inside clear 1.5ml HPLC vials (HPLS, Chromacol, V944/950). The sample was then mounted on a IKA color squid stirrer plate (IKA, 0003671000) and stirred at 1000 rpm for a 4-hour time course

using a 3x8mm polytetrafluoroethylene-coated magnetic stirring bar (VWR, 442-4520). Samples were taken at appropriate time-points and diluted with filter sterilized milli-Q water to a final concentration of 0.05mg/ml. The samples were then prepared for AFM imagine per the procedure listed previously, however this procedure resulted in a lack of fibrils on the mica surface. Therefore, optimisation was required to modify interactive behaviour between the fibril particles and the mica surface. This apparent reduction in interaction may be due to the reduction in length itself resulting in a per particle reduction in repeated charged residue side chains that may facilitate the charge based interaction between fibril and mica. Undiluted deposition with an overnight incubation resulted in overloaded mica surfaces, however this observation supported a reduction in surface interaction. Therefore, to enhance the deposition of fibrils the post-fragmentation incubation time was increased step-wise. However, as the fragmentation progressed and fibrils become shorter, the issue persisted. An incubation time of 45 minutes was adopted, which increased adherence of the fibril particles to the mica surface at both the first and last time point. The fragmentation was conducted in filter sterilized milli-Q water with no adjustment to pH as with unfragmented imaging due to the duration of fragmentation. It has been shown that prolonged incubation in a low pH environment of mature insulin and lysozyme fibrils can modify the morphology of the fibrils by adopting different filament twisting archetypes (Kurouski, Dukor, Lu, Nafie, & Lednev, 2012)(Kurouski et al., 2013).

Following this method results in the images seen in figure 10. The issue presented here is the clumping of the fibrils as the fragmentation progresses. This is more evident with RVFNIM than the other two peptides and considerably less evident with HYFNIF. This is potentially due to the exposure of hydrophobic surfaces to the amyloid core through fragmentation, which would promote clumping driven by hydrophobic interactions. In

addition, the hydrophobic residues which bury their side chains within the amyloid core, upon fragmentation may present interactions between particles following their exposure. Regardless, homogenous fragmentation conditions for each of the WALTZ peptides were optimised and implemented for each sample. Both the pH and the ionic strength of the deposition were adjusted. As with long term exposure to a low pH environment, ionic strength of a mature fibril solution has been shown to modify the morphology of b-lactoglobulin fibrils (J Adamcik & Mezzenga, 2011). Therefore, all adjustments to the solution conditions were only made at the time of deposition to minimise samples exposure to high ionic strength, low pH solution conditions. Post fragmentation pH adjustments were made by adjusting the 20ul deposition aliquot to 10% pH1 HCL. Adjustments to ionic strength were made by adjusting the deposition aliquot to contain 2.5mM sodium phosphate. Various combinations of conditions can be seen in table 4 and images of specified conditions seen in figure 10. Unfortunately, while HYFNIF presented little issue, RVFNIM was particularly unresponsive to the changes in deposition conditions when using the stirring method of fragmentation. This is possibly due the nature of the slow mechanical agitation which may encourage interaction between fibrils by maintaining an increased particle proximity. In addition to this, the adjustment of conditions post-fragmentation may not be sufficient to reverse any clumping behaviour.

Sonication	HYFNIF	VIYKI	RVFNIM	Stirring	HYFNIF	VIYKI	RVFNIM
Pre-pH2	+	~	~	Pre-pH2	N/a	N/a	N/a
Post-pH2	+	-	-	Post-pH2	+	~	-
Pre-NaP	-	-	-	Pre-NaP	N/a	N/a	N/a
Post-NaP	-	-	-	Post-NaP	-	-	-
PrepH+PreNaP	+	+	+	PrepH+PreNaP	N/a	N/a	N/a
PrepH+PostNaP	~	~	~	PrepH+PostNaP	N/a	N/a	N/a
PostpH+PreNaP	+	~	-	PostpH+PreNaP	N/a	N/a	N/a
PostpH+PostNaP	~	~	~	PostpH+PostNaP	+	~	-

+ Good result (no clumping, dispersed fibrils)
- Bad Result (clumping, unusable fragmentation data)
~ Sub-optimal result (varying degrees of clumping)

Table 1. Solution conditions affect fibril fragmentation behaviour

The various pre- and post-fragmentation conditions tested for samples of the WALTZ peptide fibrils; sonication and stirring. Concentrations and incubation times were as detailed in the text for the respective methods. Sample concentration of Sodium phosphate was 2.5mM.

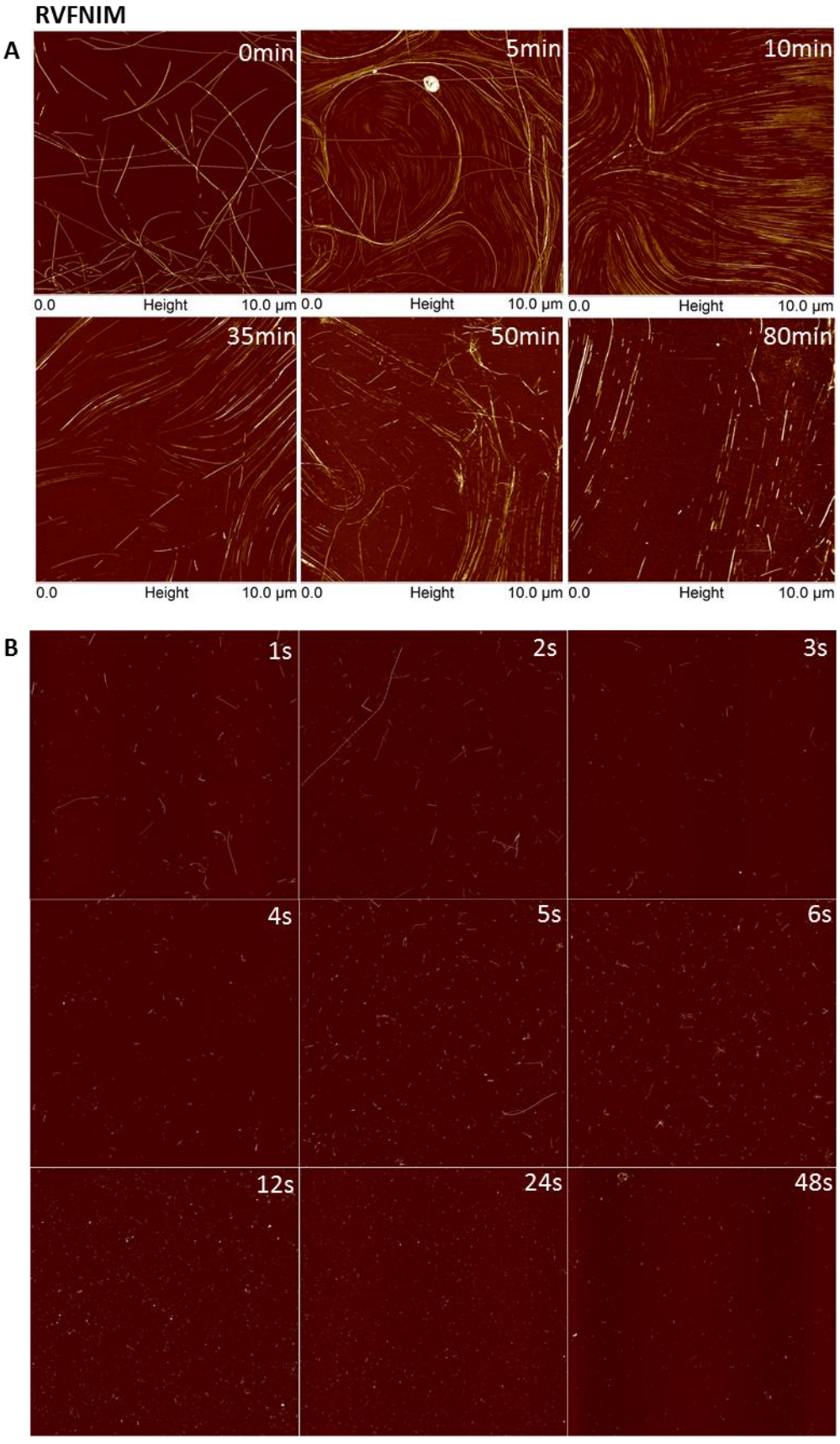
Interestingly, higher concentrations of sodium phosphate as well as the addition of NaCl (25mM and 50mM, respectively) had the adverse effect as a small concentration of sodium phosphate. An alternate fragmentation method is sonication, and this was adopted to acquire the data for the quantitative analysis of reduction in length over time. The potential benefit of sonication is the likely dispersal effect experienced by the fibrils through the release of sonic energy into the solution, which may reduce the likelihood of clumping. The fibrils fragment in a considerably shorter time with this method, meaning that the adjustments to pH and ionic strength can be made prior to the fragmentation while avoiding the issue of morphology change during the experiment. However, to determine the ideal conditions using sonication, combinations of both pre-and post-fragmentation adjustments to ionic strength and pH were tested. Each combination as well as the equivalent combination for stirring can be seen in table 4 with representive images in figure 10. By fragmenting with no adjustment to pH and

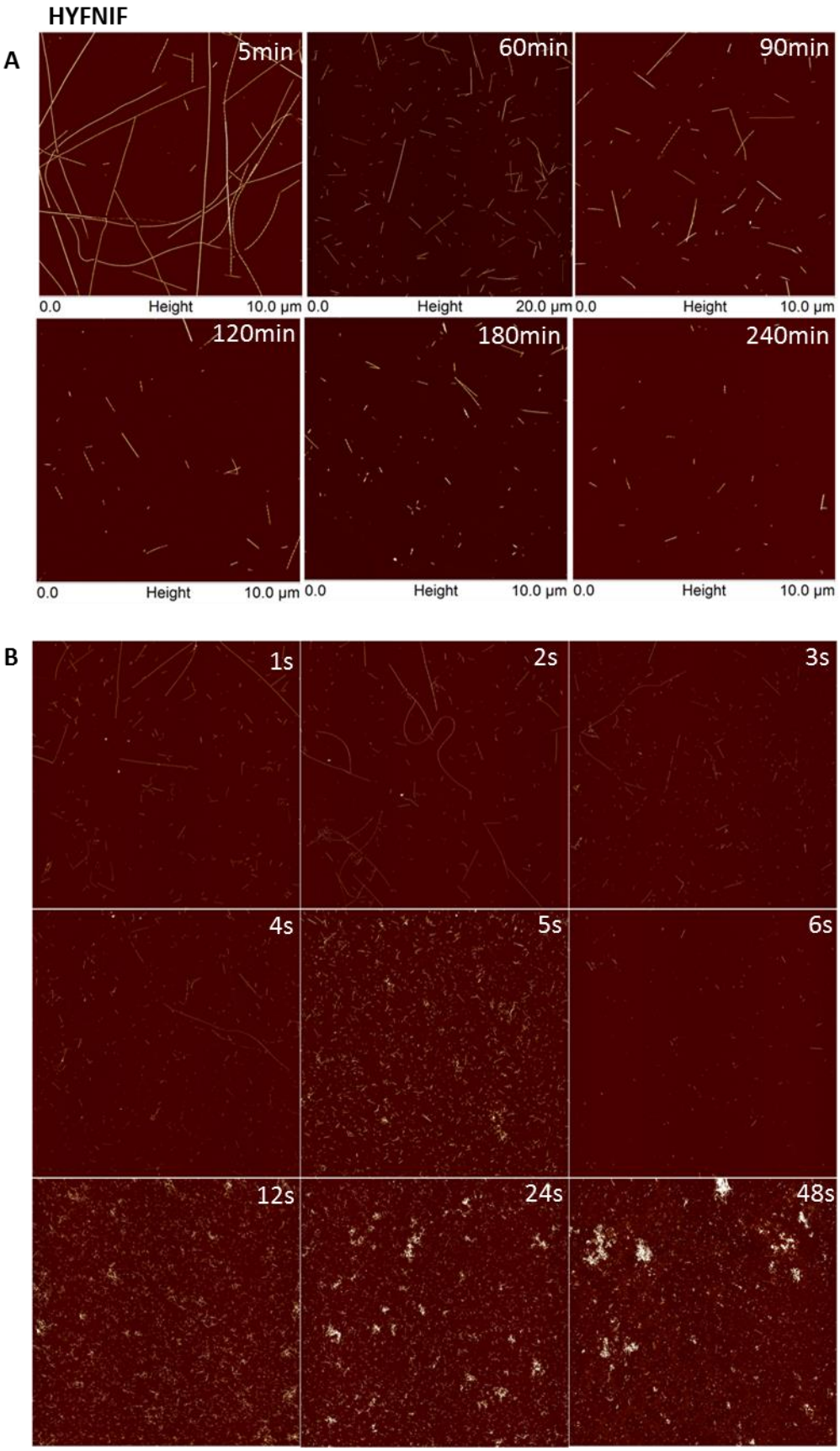
ionic strength, sonication resulted in clumped fibrils which were not adequate for analysis. By adjusting the pH post-fragmentation, only HYFNIF fibrils responded yielded desirable results while fibrils formed from the other two peptides remain an issue in regard to clumping. Sonicating the fibrils at pH2 marks some improvement but is still not ideal. Therefore, it seems that adjustments must be made prior to fragmentation, and pH alone is not sufficient. The conditions that granted the optimum fragmentation behaviour was the pre-fragmentation adjustment to pH 2 accompanied by an increase in ionic strength through the addition of sodium phosphate (table 4). The pH2 adjustment is beneficial in both inhibiting the presence of previously observed background material (figure 9). Additionally, an increased concentration of H^+ may neutralise exposed negatively charged carboxylate groups; reducing possible charged based interactions. Hydrophobic material has been shown to be reduced in its propensity to cluster in solution through the addition of ions with medium to weak charge density – known as salting-in. This restricts hydrophobic interactions through an increase in the number of ordered water molecules in solution as well as possibly forming near-hydrophobic interactions with the insoluble surfaces themselves, further restricting additional interaction (Zangi, Hagen, & Berne, 2007). Therefore, the addition of a relatively small concentration of sodium phosphate into the fragmentation solution may aid in the inhibition of hydrophobically driven fibril clumping.

Using the optimal conditions, the full sonication time course was 48 seconds and each of the appropriate time points was designated an individual sample as to not expose the fibrils to altered solution conditions more than necessary. Each time point sample was made to a final volume of 120ul and a fibril concentration of 0.1mg/ml within Protein LoBind 1.5ml tubes (Eppendorf, 022431081). Each sample contained 10% pH1 HCL and 10% 25mM sodium phosphate to a final concentration of 2.5mM at pH2. Each sample

was then sonicated at 20% amplitude for the desired per time-point duration on ice (1s, 2s, 3s, 4s, 5s, 6s, 12s, 24s, 48s). Following sonication, 20ul is taken and deposited on a freshly cleaved mica surface and left to incubate for 45 minutes. After this, the sample is washed with 700ul of filter sterilized milli-Q water, and gently dried using nitrogen gas and filter paper to help remove water from the edges of the mica. All manipulations were with low bind tips (VWR 53503-294 (with SpidersilkTM surface)). Images of fragmentations using these conditions can be seen in figure 10 and were also used for the analysis in chapter 4.

Figure 10. A comparison of fragmentation methods




HYFNIF
A
5min
60min
90min
0.0 Height 10.0 µm
0.0 Height 20.0 µm
0.0 Height 10.0 µm
120min
180min
240min
0.0 Height 10.0 µm
0.0 Height 10.0 µm
0.0 Height 10.0 µm
B
1s
2s
3s
4s
5s
6s
12s
24s
48s

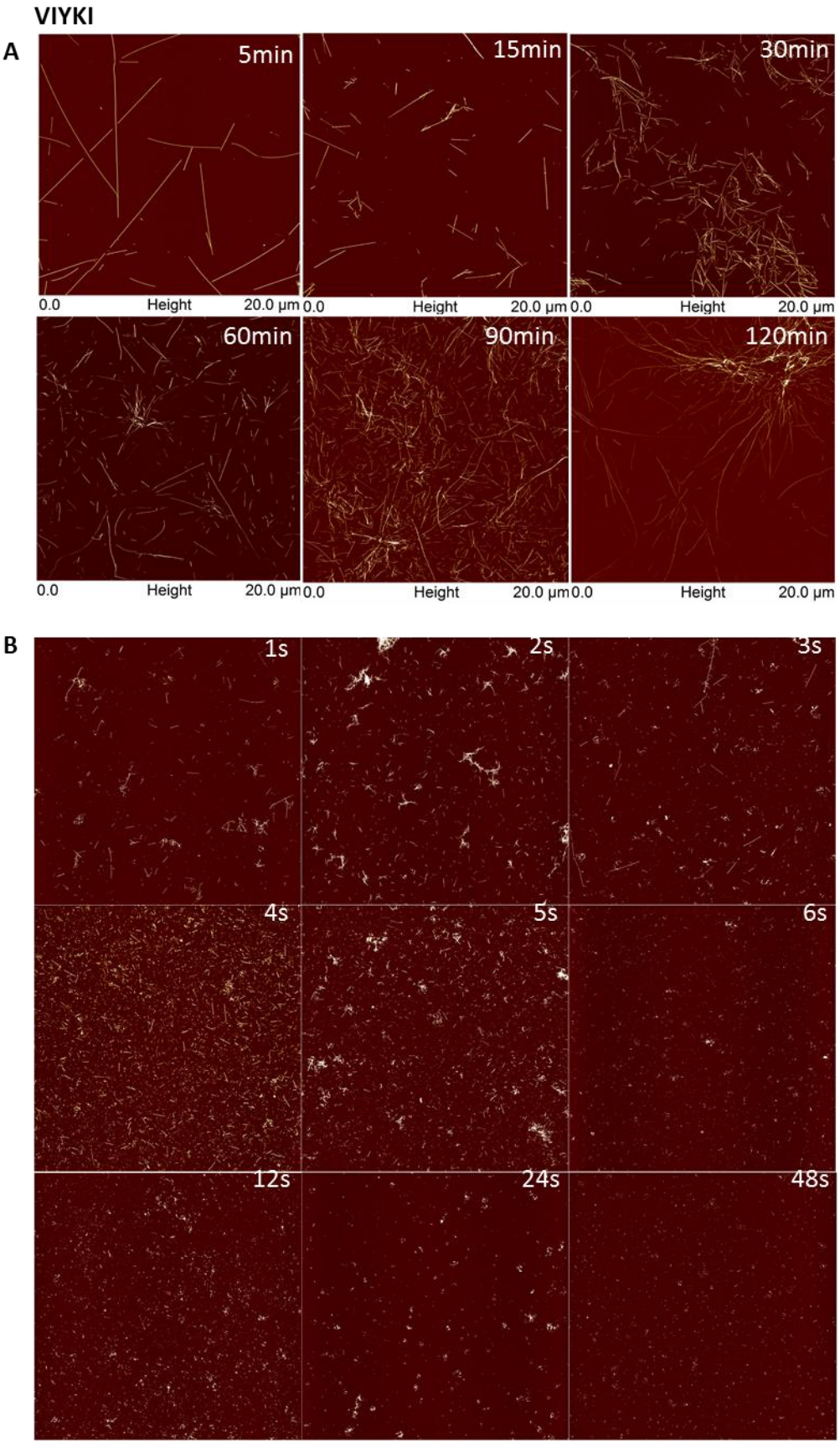

VIYKI
A
5min
15min
30min
0.0 Height 20.0 µm
60min
90min
120min
B
1s
2s
3s
4s
5s
6s
12s
24s
48s

Figure 10.

An AFM image comparison of fibrils formed from RVFNIM, HYFNIF and VIYKI for both fragmentation techniques used: A) stirring and B) sonication. Each image in A) is scaled per the bar at the bottom of each image and has a resolution of 2048 x 2048 pixels. Each image in B) is 30 x 30 μm and a resolution of 2048 x 2048 pixels. For both fragmentation techniques, the time point is presented in the top right corner. The stirring technique in A) used different time points for each of the peptides; VIYKI and RVFNIM displayed clumping behaviour. For the sonication method in B) each peptide sample was tailored the same optimal buffer conditions (table 10) and imaged at identical timepoints.

Amyloid associated buffers

Buffer	Component
WALTZ deposition buffer	10mM HCl pH2
WALTZ fragmentation buffer	2.5mM Sodium phosphate 10mM HCl pH2
α-Synuclein fibril forming buffer	25mM Sodium phosphate pH 7.5
Sup35 fibril forming buffer	20mM Sodium phosphate 50mM NaCl pH 7.4

Table 2.

AFM sample buffers used for deposition of fragmented and unfragmented fibrils

Buffer	Component
Sup35 Lysis Buffer (SLB)	20mM Tris-HCl 1M NaCl 20mM imidazole 100mM PMSF pH8
Sup35 Lysis Buffer Denaturing (SLBG)	20mM Tris-HCl 1M NaCl 20mM imidazole 6M guanidine HCl pH 8
Denaturing Sizing Buffer (DSB)	20mM Tris-HCl 0.5M NaCl 6M guanidine HCl pH 8

Table 3.

Buffers used for the purification of Sup35NM protein monomer

Buffer	Component
α -Synuclein Lysis Buffer (αLB)	20mM Tris 100mM NaCl pH 7.5
α-Synuclein Ion Exchange Entry Buffer (αEB)	20mM Tris pH 7.5
α -Synuclein Ion Exchange Elution Buffer (αElB)	20mM Tris 1M NaCl pH 7.5

Table 4.

Buffers used for the purification of α – synuclein protein monomer4

α-synuclein

Protein purification

α-synuclein material was prepared and purified in lab by Dr David Beal and Tracy Purton (University of Kent). *E. coli* expression cultures were sonicated in an α -Synuclein Lysis Buffer (αLB) with Protease Inhibitor suspension (Complete Protease Inhibitor tablets Roche, 11697498001). HCL based protein precipitation and subsequent NaOH neutralisation was performed prior to clarification via centrifugation at 13000rpm and treatment with (NH4)2SO4. The Precipitate was collected via centrifugation and resuspended in α-Synuclein Ion Exchange Entry Buffer (αEB). Purification was by ion exchange chromatography using an ÄktaPrime Plus system with a strong anion exchange HiTrapTM Q FF 5ml cartridge (GE Healthcare, 17-5156-01).

In vitro polymerisation

α-synuclein fibrils were formed by Dr David Beal (University of Kent). 2.5ml of protein in α -Synuclein Ion Exchange Elution Buffer (αElB) was buffer exchanged with α-synuclein fibril forming buffer (25mM sodium phosphate buffer, pH7.5) using a PD10 column per the manufacturer's instructions. In addition, 0.2µm syringe filtered sodium azide was added and the sample was incubated at a concentration >200µM and at 37°C with orbital agitation (160rpm) for 2 weeks.

AFM sample preparation

For unfragmented fibrils, 20ul of α-synuclein at a concentration of 1um was loaded on to a freshly cleaved mica surface (Agar scientific, F7013). Mica cleavage is achieved using tape to remove the top layer, and the mica disc is mounted on to a metal disc using double sided tape. The sample is incubated on the surface for 5 minutes, then washed with 1ml of filter sterilized milli-Q water. The surface is then dried with a gentle stream of nitrogen gas and blotted dry with filter paper to aid in the removal of residual liquid.

Fragmentation conditions

Fragmentations of α-synuclein were conducted in parallel with WALTZ peptides; therefore, the same time points were used (1s, 2s, 3s, 4s, 5s, 6s, 12s, 24s, 48s). Sonication was carried out with a 20% amplitude and samples were on ice for each time point's duration. Fibril solutions were diluted to match the mass concentration of the WALTZ samples, 0.1mg/ml (7uM) using 25mM sodium phosphate buffer (pH7.5). Deposition of this concentration on to the mica resulted in overloaded surfaces so post-sonication dilutions were made. This was done serially, first a dilution to 0.007mg/ml (0.5uM) with a subsequent dilution to 0.003mg/ml (0.2uM) - which resulted in heavy but clearly dispersed fibril distribution on the surface. Dilutions were made using fibril forming buffer.

The reason for a considerably lower required concentration for comparable surface fibril density may be due to the per fibril size difference between WALTZ formed fibrils and α-synuclein fibrils in that in general a WALTZ fibril is morphologically larger. Additionally, reduction in surface interaction between WALTZ fibrils and the mica affects the possible surface density. By matching the molar concentration of fragmentation, α-synuclein samples were considerably more overloaded; therefore, the mass concentration was

more reliably repeatable and potentially more indicative of an equivalent fibril population. Due to the relatively small molecular weight of the WALTZ peptides, an equivalent molar concentration presents a lower number of WALTZ formed fibrils due to the higher number of monomer incorporation per fibril.

Following a 5-minute incubation of 0.003mg.ml α-synuclein, the sample was washed with 1ml of filter sterilized milli-Q water and dried under a gentle stream of nitrogen gas with filter paper to aid in the removal of residual liquid from the mica surface. All manipulations were with low bind tips (VWR 53503-294 (with SpidersilkTM surface)), and fragmentations performed in Protein LoBind 1.5ml tubes (Eppendorf, 022431081).

AFM imaging parameters

Images were collected at various resolutions and scan sizes (annotated accordingly). Generally, α-synuclein was used for quantitative fragmentation analysis therefore images were 30µmx 30µm or 20µmx 20µmand at a resolution of 2048 x 2048 pixel. A scan rate of 0.305Hz was used with a noise threshold of 0.2nm (manually set within the nanoscope parameter selection workspace). Additionally, Z limit was reduced to 1.5µmand the peak force set point set to 0.05- 0.15mV (depending on sample). Images can be seen in figure 11 and Image collection following this method was used for the analysis in chapter 4.

Dityrosine cross-linked α-synuclein

In vitro polymerisation

Dityrosine cross-linked α-synuclein fibrils and comparative α-synuclein fibrils for fragmentation were obtained from Dr Youssra K. Al-Hilaly (Al-Hilaly et al., 2016). For cross-linked material, α-synuclein monomer was incubated with Cu^{2+} at a 1.1 molar ratio (50µM). Oxidation begun following the addition of H_2O_2 (1.25 mM). This was performed

in the presence of 20mM HEPES (pH 7.4) and a temperature of 37 °C including agitation of 400 rpm. For a separate α-synuclein sample run in parallel, oxidation was inhibited with EDTA at 24h using a solution concentration of 1.25 mM.

AFM sample preparation

500ul of both the Cu^{2+} oxidised and Cu^{2+} depleted fibril samples were diluted to a concentration of 2µm with filter sterilized milli-Q water inside clear 1.5ml HPLC vials (HPLS, Chromacol, V944/950). It was then stirred at 1000rpm for 1-hour using a 38mm polytetrafluoroethylene-coated magnetic stirring bar (VWR, 442-4520) to cause fibril distribution prior to deposition. 20 µl was then loaded on a freshly cleaved mica surface and incubated for 8 minutes. The sample was then washed with 1 ml filter sterilized milli-Q water and subsequently dried using a gentle stream of nitrogen gas and filter paper to absorb residual liquid.

AFM imaging parameters

Images were collected a 20x 20µm and at a resolution of 2048 x 2048 pixel. A scan rate of 0.4Hz was used with a noise threshold of 0.2nm (manually set within the nanoscope parameter selection workspace). Additionally, Z limit was reduced to 1.5µm and the peak force set point set to 0.05- 0.15mV (depending on sample).

Sup35NM

Protein purification

Sup35NM material was prepared and purified in lab by Dr Nadia Koloteva-Levine and Tracy Purton (University of Kent). E. coli expression cultures were sonicated in a Sup35NM Lysis Buffer (SLB) suspension (Table 2). Lysates were clarified via centrifugation prior to Ni-NTA resin incubation. Ni-NTA attached material was eluted in 1ml of Sup35NM Lysis Buffer Denaturing (SLBG) and subsequently purified by size exclusion chromatography using a GE Healthcare HiLoadTM 16/600 SuperdexTM 200pg,

from GE Healthcare (28-9893-35), using Denaturing Sizing Buffer (DSB) on an ÄktaPrime Plus system at a flow rate of 0.3ml/minute.

In vitro polymerisation

Sup35NM fibrils were formed by Dr Nadia Koloteva-Levine (University of Kent). 2.5ml of protein in DSB was buffer exchanged with 3.5ml of Sup35 fibril forming buffer (25mM sodium phosphate 50mM NaCl) using a PD10 column according to the manufacturer's instructions. Fibril forming solutions were made at 10µm and required incubation at 37°C without agitation. All manipulations of the Sup35 solutions were with low bind tips (VWR 53503-294 with SpidersilkTM surface), and Protein LoBind 1.5ml tubes (Eppendorf, 022431081).

AFM sample preparation

Unfragmented Sup35NM fibrils were imaged for the structural comparison with C-DAG formed fibrils. Sup35NM tends to form fibril clumps, due to strong interfibrillar interactions. Therefore, to produce an adequate surface for imaging in which fibrils can reproducibly be found on the mica surface, the concentration must be relatively high. Various dilutions were used with the original 10µM stock (dilutions with Sup35 fibril forming buffer) however, the final conditions for the images seen in chapter 5 used the undiluted deposition on to a freshly cleaved mica surface. In addition to the strong inter-fibril interaction, Sup35NM interacts strongly with the mica surface. A 5-minute incubation is sufficient. This is followed by a wash with 1 ml filter sterilized milli-Q water and dried using a stream of nitrogen gas (not gentle). All manipulations of the Sup35 solutions were carried out with low bind tips (VWR 53503-294 (with SpidersilkTM surface)), and Protein LoBind 1.5ml tubes (Eppendorf, 022431081).

Fibril Fragmentation conditions

Sup35nm fragmentations provided a comparison for the WALTZ peptides of a fibril population which is relatively resistant to fragmentation. Fragmentations were conducted in accordance to previous conditions optimised within the lab. 10 µm Sup35NM in fibril forming buffer (25mM sodium phosphate 50mM NaCl) was sonicated for 4minutes in 5 second bursts at 20% power. Fibril solution was contained using Protein LoBind 1.5ml tubes (Eppendorf, 022431081), and kept on ice for the duration of fragmentation. Post-fragmentation, 20ul of fibril solution was loaded on to a freshly cleaved mica surface and incubated for 5 minutes. This was followed by a wash with 1 ml filter sterilized milli-Q water, then dried using a stream of nitrogen gas.

AFM imaging parameters

Images were 30x 30µm or 6 x 6 µm at a resolution of 2048 x 2048 pixel, for fragmentation analysis and structural analysis respectively. A scan rate of 0.305Hz was used with a noise threshold of 0.2nm (manually set within the nanoscope parameter selection workspace). Additionally, Z limit was reduced to 1.5µmand the peak force set point set to 0.05- 0.15mV (depending on sample). Images can be seen in figure 11 and Image collection following this method was used for the fragmentation analysis in chapter 4 and the C-DAG/*in vitro* Sup35NM comparison in chapter 5.

Aβ40

In vitro polymerisation

1 mg of lyophilised Aβ40 peptide (Bachem, Bubendorf, Switzerland) was dissolved in 200 µL hexafluoroisopropanol (HFIP) and subsequently vacuum dried for 24 hours for HFIP removal. The peptide film was re-suspended in 1-part 10x PBS, 2 parts 20 mM sodium hydroxide, and 7 parts milliQ water for a final concentration of 1 mg/mL which

was followed by a 7-day incubation at 37°C. As described in publication (R. A. S. Smith et al., 2015).

AFM sample preparation

The original stock of Aβ40 was diluted to 0.5mg/ml with filter sterilized de-ionized water and then 20µl deposited on a freshly cleaved mica surface. The fibrils were incubated on the surface for 1 hour due to their relatively weak interaction with the mica surface. The samples were then washed with 1ml of filter sterilized milli-Q water and dried with a gentle stream of nitrogen gas in conjunction with filter paper absorption.

Fragmentation conditions

Fibrils were fragmented at a concentration of 1mg/ml; samples were freeze-thawed (-20°C) and sonicated for 30 minutes (Branson 1210 sonicator). Fibril solutions were diluted to 0.5mg/ml and prepared on mica surfaces the same as unfragmented fibrils (R. A. S. Smith et al., 2015).

AFM imaging parameters

Both fragmented and unfragmented fibril samples were imaged at 6 x 6 µm and a resolution of 2048 x 2048 pixel or 12 x 12 µm and a resolution of 4096 x 4096 pixels and cropped. Images were collected at a scan rate of 0.488Hz or 0.244Hz depending on resolution, with a noise threshold of 0.5nm and a z limit of 1.5 µm; manually set in the nanoscope analysis workspace.

Lysozyme

In vitro polymerisation

Lysozyme fibrils were prepared as per the publication (R. A. S. Smith et al., 2015) by dissolving 1 mg of protein monomer (Sigma Aldrich) in 1 mL of pH1.6, 0.1 M HCl . A subsequent overnight incubation was performed at 65°C with agitation at 550 rpm. (R. A. S. Smith et al., 2015).

AFM sample preparation

The original stock of lysozyme was diluted to 0.1mg/ml with filter sterilized de-ionized water and then 20µl deposited on a freshly cleaved mica surface. The fibrils were incubated on the surface for 5 minutes. The samples were then washed with 1ml of filter sterilized milli-Q water and dried with a gentle stream of nitrogen gas in conjunction with filter paper absorption.

Fragmentation conditions

Fibrils were fragmented at a concentration of 1mg/ml; samples were freeze-thawed (-20°C) and sonicated for 30 minutes (Branson 1210 sonicator). Fibril solutions were diluted to 0.5mg/ml and prepared on mica surfaces the same as unfragmented fibrils (R. A. S. Smith et al., 2015).

AFM imaging parameters

Both fragmented and unfragmented fibril samples were imaged at 20 x 20 µm and a resolution of 2048 x 2048 pixel and cropped. Images were collected at a scan rate of 0.488Hz, with a noise threshold of 0.5nm and a z limit of 1.5 µm; manually set in the nanoscope analysis workspace.

Sample	Deposition volume	Deposition concentration	Incubation time	Wash volume	Deposition solution conditions
WALTZ	20µl	0.05mg/ml	10 minutes	1000µl	pH2 HCL
WALTZ-CDAG	10µl	-	40 minutes	500µl	pH7 PBS
Lysozyme	20µl	0.1mg/ml	5 minutes	1000µl	pH2 HCL
Aβ40	20µl	0.5mg/ml	60 minutes	1000µl	10mM NaOH, 5X PBS
α-synuclein	20µl	0.1mg/ml	5 minutes	1000µl	25mM sodium phosphate buffer
Dityrosine α-synuclein	20µl	0.03mg/ml	8 minutes	1000µl	2µM Cu2+, 60µM H_2O_2, 1mM HEPES
Sup35NM	20µl	0.3mg/ml	5 minutes	1000µl	25mM sodium phosphate 50mM NaCl

Table 5.

AFM Sample preparation and deposition conditions for various amyloid species

Sample	Fragmentation volume	Deposition volume	Fragmentation concentration	Deposition concentration	Incubation time	Wash volume	Deposition solution conditions
WALTZ	120µl	20µl	0.1mg/ml	0.1mg/ml	45 minutes	1000µl	pH2 HCL, 2.5mM sodium phosphate buffer
Lysozyme	200µl	20µl	1mg/ml	0.1mg/ml	5 minutes	1000µl	pH2 HCL
Aβ40	200µl	20µl	1mg/ml	0.5mg/ml	60 minutes	1000µl	10mM NaOH, 5X PBS
α-synuclein	120µl	20µl	0.1mg/ml	0.003mg/ml	5 minutes	1000µl	25mM sodium phosphate buffer
Dityrosine α-synuclein	500µl	20µl	0.03mg/ml	0.03mg/ml	8 minutes	1000µl	2µM Cu2+, 60µM H_2O_2, 1mM HEPES
Sup35NM	120µl	20µl	0.3mg/ml	0.3mg/ml	5 minutes	1000µl	25mM sodium phosphate 50mM NaCl

Table 6.

Fragmentation and subsequent AFM sample deposition conditions for various amyloid species

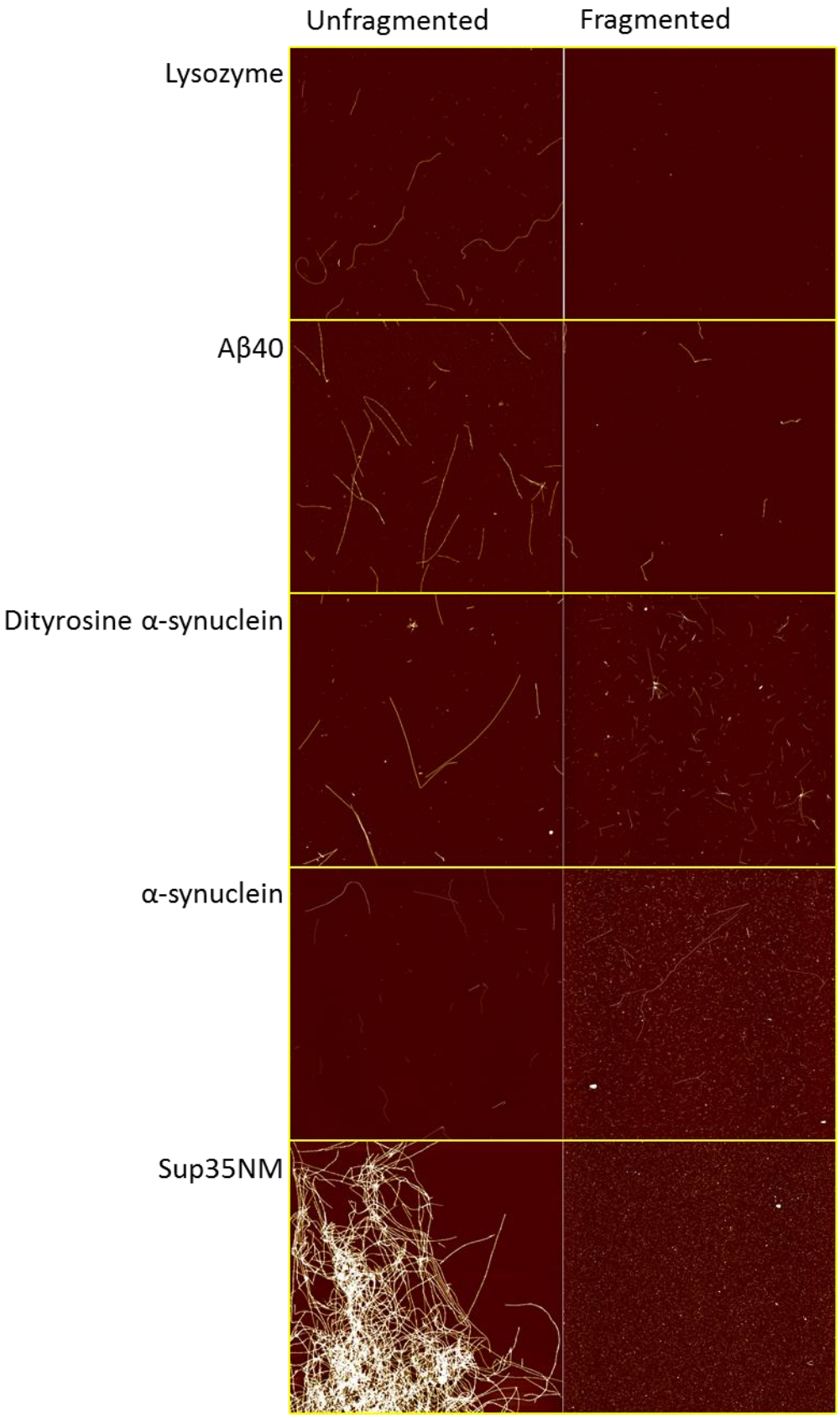


Figure 11. Imaging various types of amyloid

Example AFM images of respective amyloid species fragmented and unfragmented. All images are 2048 x 2048 pixels with scan sizes ranging from 6 x 6 µm – 20 x 20 um.

Generation of fibrils using the Curli-dependent amyloid generator (C-DAG)

Media	Components	Amount
Lysogeny broth (LB)	Tryptone Yeast extract NaCl Agar (for solid media)	1g/100ml 0.5g/100ml 1g/100ml 2g/100ml
Super Optimal broth with Catabolite repression (SOC)	Tryptone Yeast extract NaCl KCL MgCl2 MgSO4 Glucose	2g/100ml 0.5g/100ml 0.05g/100ml 0.01g/100ml 0.09g/100ml 0.2g/100ml 0.3g/100ml
Induction LB	LB (as detailed above) Ampicillin Chloramphenicol 0.2% Arabinose 1mM IPTG	 100 µg/ml 25µg/ml 1 ml/100ml (20% stock) 1ml/100ml (100mM stock)
Congo red induction media	Induction LB (as detailed above) Congo red	 10µg/ml

Table 7.

Table containing media components used for chapter 5

Preparation of competent cells

The three strains of *E. coli* used for the C-DAG generation of amyloid fibrils were Top10, DH5α and VS45. The VS45 strain contains C-DAG machinery, part of C-DAG Amyloidogenicity Kit purchased from Kerafast, EF2001. Competent versions of each (subsequent generation for VS45) were prepared with the same protocol. 40µl of the respective strain is grown overnight in 2ml of LB media at 37°C. The overnight culture is inoculated into 100ml of LB media and grown at 37°C until the OD600 is ~0.4. *E. coli* culture is then harvested by centrifugation at 3000rpm for 8 minutes at 4°C in prechilled centrifuged bottles. The pellet is resuspended in 25ml of chilled 100mM $CaCl_2$ and incubated on ice for 10minutes. The culture is again harvested in the same manner and

subsequently resuspended in 2ml of 100mM CaCl2/30%(v/v) glycerol and divided into 100 µl aliquots. Aliquots are snap frozen using liquid nitrogen and stored at -80°C.

Assembly of WALTZ constructs

Restriction digest reactions

Component	Amount
DNA sample	0.5-2µg DNA (X µl)
Restriction Enzyme 1	1µl
Restriction Enzyme 2	1µl
Buffer	3µl
mQ water	Up to 30µl

Table 8.

A typical setup for a restriction enzyme reaction

For the expression and production of the WALTZ peptides using the C-DAG system, VS45 *E. coli* needed transforming with the respective WALTZ sequence containing pVS72 construct. A restriction digest of PVS72 (C-DAG Amyloidogenicity Kit purchased from Kerafast, EF2001) with Xba1 and Not1 was set up and either incubated at room temperature overnight in a 37°C water bath for 1 hour. A typical restriction digest can be seen in table 8 and this is the template reaction set-up for all digests conducted. The digest of PVS72 was using 10x 'buffer H' (Promega) which was selected for optimum activity using the Promega restriction enzyme tool. Following the assembly of pVS-WALTZ constructs, restriction digest reactions were set up to confirm the successful ligation of WALTZ sequences into the PVS backbone. These digests were set up using the template reaction described in table 8 and used EcoRV and Xba1 with '10x multicore buffer' (Promega) which was selected for optimum activity using the Promega restriction enzyme tool.

Gibson assembly

5x isothermal buffer		Gibson master mix	
Component	***amount***	***Component***	***amount***
PEG-8000	500µl	5x isothermal buffer	100µl
2M Tris	250µl	Taq ligase	50µl
500mM MgCl2	100µl	T5 exonuclease	2µl
DTT	50µl	Phusion polymerase	6.3µl
NAD	50µl	mQ water (sterile)	217µl
Nucleotides (ATCG)	10µl (each)		
mQ water (sterile)	10µl		

Table 9.

Component assembly for the lab made Gibson master mix

Component	Amount
Vector DNA fragment	100ng DNA (X µl)*
Insert DNA fragment	Xng of DNA**
Gibson assembly master mix	10µl
mQ water	Up to 20µl

*Calculate pmols of vector DNA

**1:2-3 molar ratio (vector: insert) with 2 fragments; equimolar ratios for 3+

Fragments less than 200bp, use up to 1:5 molar ratio

Table 10.

Details for a typical Gibson assembly reaction setup

Assembly of the PVS-WALTZ constructs (figures 40-42) was with Gibson assembly, which followed a 2-fragment procedure. Fragment 1 of the reaction was the pVS backbone and fragment 2 was the respective WALTZ sequence (RVFNIM, HYFNIF, VIYKI; sequences seen in figures 40-42 chapter 5). The WALTZ encoding DNA sequences were synthesised by IDT (gBlocks® Gene Fragments) and arrived as 250ng of freeze dried powder which was resuspended in 25 µl of TE buffer to a final concentration of 10ng/ µl. The Gibson assembly master mix was either purchased from New England biolabs (NEB UK, E2611S) or made in the lab in accordance to table 9. Both master mixes were effective however had slight variations in the respective protocol.

For the reaction mixture, 100ng (0.032 pmol) of vector (pVS-backbone) was used in a 1:5 molar ratio (second fragment less than 200bp) with the WALTZ insert. Therefore, 13ng (0.16 pmol) of WALTZ insert was used. Fragment mixture is then added to a 10 µl (NEB) or 15 µl (lab made) aliquot of Gibson assembly master mix and adjusted to a total volume of 20 µl with filter sterilized de-ionized water. The reaction is incubated for 15 minutes (NEB) or 1 hour (lab made) at 50°C and subsequently transformed into DH5α competent cells.

Transformation of Gibson product into DH5α

One aliquot of DH5α cells are thawed on ice, then 2 µl of the Gibson reaction mixture added and mixed by gentle flicking of the tube. The cells are then stored on ice for 30 minutes prior to a 30 second heat shock at 42°C. The cells are again stored on ice for a further 2 minutes then SOC media is added (up to a total volume of 1ml). The transformation mixture is incubated at 37°C for 1 hour with shaking at ~200rpm. Pre-warmed LB plates supplemented with 0.1% ampicillin (100µg/ml) were then streaked with 100µl of transformation mixture and incubated overnight at 37°C

Cloning required plasmid constructs

All cloning of required plasmid; PVS72, PVS105 (negative control, C-DAG Amyloidogenicity Kit purchased from Kerafast, EF2001), WALTZ-PVS constructs – was conducted using either Top10 or DH5α competent cells. Following the previously explained transformation protocol, Plasmid DNA purification was achieved using Qiagen's QIAprep® Spin MiniPrep kit or Thermofisher's GeneJET Plasmid MiniPrep kit, following the manufacturer's instructions. DNA quantification was measured by absorbance at 260nm using an Eppendorf Biophotometer.

Confirmation of constructs by gel electrophoresis

50 x TAE Buffer		**1% Agarose gel**	
Component	***amount***	***Component***	***amount***
Tris base	242g	1 x TAE buffer	50ml
0.5M EDTA	100ml	Agarose	0.5g
Glacial acetic acid	57.1ml	Ethidium bromide	1µl
mQ water	Up to 1 L		

Table 11.

Components required for TAE buffer and a 1% agarose gel used for DNA gel electrophoresis

Plasmid DNA constructs were validated using a standard gel electrophoresis procedure. PVS72 and PVS-WALTZ plasmids were digested using restriction enzymes as described previously, then run on an agarose gel to confirm the respective band sizes. To make the 1% agarose gels generally used in this study; 0.5g of agarose is first dissolved in 1x TAE buffer and microwaved on full power, taking care to monitor the dissolution with careful mixing of the flask. After sufficient cooling, 1µl of 10µg/ml ethidium bromide

(Sigma Aldrich, E1510) was added to allow the UV dependent visualisation of DNA bands. The agarose solution was then decanted into a gel tank defined with appropriate dam's prior to inserting a lane comb (ensuring the lanes are at the anode of the tank). Once the gel had set, the tank was filled with 1 x TAE buffer ensuring the gel itself was sufficiently covered. Each sample, is diluted with 6X DNA Loading Dye (Thermofisher scientific, R0611) prior to loading into individual lanes. Additionally, 6µl of Thermo Scientific GeneRuler 1 kb DNA Ladder (Thermofisher scientific, SM0311) for the size-based identification of each sample band. Gels were run at 80-120V until the dye front surpassed the mid-point of the gel. The agarose gels were subsequently imaged using a Fujifilm FLA-5100 scanner and Image Reader FLA-5000.

Induction of the C-DAG system

Transformation of VS45

VS45 cells (50 µl aliquots) were thawed on ice per construct (PVS72, PVS105, PVS-RVFNIM, PVS-HYFNIF, PVS-VIYKI). 50ng of plasmid DNA is added to the competent cells and is mixed by gentle flicking. The cells are then stored on ice for 30 minutes prior to a 2-minute heat shock at 42°C. The cells are again stored on ice for a further 2 minutes then SOC media is added (up to a total volume of 1ml). The transformation mixture is incubated at 37°C for 1 hour with shaking at ~200rpm. Pre-warmed LB plates supplemented with 0.1% ampicillin (100µg/ml) and chloramphenicol (25 µg/ml) were then streaked with 100µl of transformation mixture and incubated overnight at 37°C. The chloramphenicol is required for the maintenance of the c-DAG machinery encoded in pre-transformed plasmid PVS76.

Induction of the C-DAG system

Colonies were picked from each of the respective plasmid transformation plates, and inoculated into 2 separate 15ml falcon tubes, each containing 15ml of LB liquid media supplemented with 0.1% ampicillin (100µg/ml) and chloramphenicol (25 µg/ml) and incubated overnight at 37°C with shaking at ~180rpm. The cultures are then diluted to an OD600 of 0.01 using the 3ml of appropriate media; subsequent cultures are incubated with shaking at ~180rpm, for 30 minutes at 37°C. For each sample to be induced, 5 µl spots are loaded from one of the falcon tubes on to induction plates (LB agar with 0.1% ampicillin (100µg/ml) and chloramphenicol (25 µg/ml), 0.2% arabinose and 1mM IPTG). L-arabinose was prepared as a 20% (v/v) solution in water, and filter sterilised (0.2µm). Spotted plates are then incubated for 5 days at 22°C.

Congo red plate assay

For each sample to be induced and visualised, 5 µl spots are loaded from the second falcon tube on to Congo red induction plates (LB agar with 0.1% ampicillin (100µg/ml) and chloramphenicol (25 µg/ml), 0.2% arabinose, 1mM IPTG and 10ug/ml Congo red.). L-arabinose was prepared as a 20% (v/v) solution in water, and filter sterilised (0.2µm). Spotted plates are then incubated for 5 days at 22°C.

Imaging with AFM

Sample preparation

To image the respective VS45 strains, AFM was used. Mica surfaces (Agar scientific, F7013) were prepared for each sample to be imaged by mounting on a steel disc (Agar scientific) and cleaving the top layer with tape. 25 µl of PBS is then transferred to each colony that needs imaging and is gently resuspended using the pipette action. A maximum of 20 µl of the resuspension is then transferred to an Eppendorf tube and 10 µl is subsequently transferred to the mica surface. The sample is incubated on the

surface for 40minutes before a wash with 500 µl of filter sterilized milli-Q water. Finally, the sample is dried with a gentle stream of nitrogen gas.

AFM imaging parameters

Samples were imaged using a Bruker Multimode AFM with a Nanoscope V controller operating under peak force tapping mode. Images ranged from 6 x 6 µm to 30 x 30µm at a resolution of 2048 x 2048 pixel. Equipped with a ScanAsyst probe (silicon nitride triangular tip with tip height = 2.5-8µm, nominal tip radius = 2nm, nominal spring constant 0.4N/m and nominal resonant frequency 70kHz). A scan rate of 0.203Hz was always used with a noise threshold of 0.1nm (manually set within the nanoscope parameter selection workspace). Depending on orientation of cells on the surface, the scanning angle was set to either 0°, 45°, or 90°. Due to the size of the cells, the Z limit was automatically adjusted by the ScanAsyst scanning mode. Due to the high level of interaction between the bacterial cell surface and the AFM probe, the peak force set point was maintained at a relatively low value of 0.05mV.

Matlab analysis of data

Analysing AFM images using Trace-y

Data collected by AFM was analysed using an in house Matlab (mathworks, 2015b) application (Xue, unpublished). The application automatically searches for particles within AFM images based on input parameters including a z distance (nm) threshold, apparent width of the particles (pixels) and search angle (°). These values are set on a per image and per sample basis. The apparent width in which each image may have variation within the same sample type due to tip convolution. The remaining two parameters are usually maintained throughout multiple images of the same sample (i.e. height range is expected to be maintained throughout a sample). Once particles have been selected, they are traced and the application extracts information from the data

file on a per pixel basis. Additionally, particles can be manually selected for tracing giving a semi-automatic function to the application. Particles that are cut off by the image boundaries are saved as particle segments, while particles in which both natural ends are included within the trace are saved as full particles. Here the key properties that are collected using this application is the height profile and length of each particle. Additionally, the correlation of bending angles along a polymer axis is extracted for persistence length measurements. Optimisation of the experimental data collection was important for the subsequent analysis to provide accurate and reproducible data analysis. Samples that contain fibrils that are either contained within highly associative clumps or are overloaded on the mica surface; provide unreliable height and length information. Optimisation of the analysis using Trace-y required refinement of the input parameters in conjunction with experimental parameters. E.g. if traces cross between individual fibril particles then a reduction in the either the apparent width of the particles (pixels) or search angle (°) is required. Alternatively, if particle traces surpass the apparent boundary then an increase in the Z threshold (nm) is required.

Structural data extraction

For the initial structural analysis included in chapter 3 the extracted property of focus was the height profile. The height profile of each particle traced contains the per pixel z value. Data extracted from the Trace-y output file was used to calculate the mean particle height (Appendix 1). This was run on both individual image data files as well as collated datasets of each fibril traced per sample. In addition to this, the periodicity of the particles was determined which also utilised the height profile of each particle. Fast-Fourier transform was performed on both single particle and on batch datasets per sample to provide aperiodic measure of a fibril height profile. The batch script used (Appendix 1) was adapted from the single particle application AFMHeightFFT (Xue,

unpublished) which performs FFT on individual particle traces and provides graphical representation of the height profile, the frequency, and the predominant periodicity. Appendix 1 identifies the predominant frequency from the height profile using FFT and provided the peak periodic value (nm). To supplement the mean height information, both the maximum and minimum height of each particle was extracted from the Trace-y files (Appendix 1). The extracted datasets were analysed within both Excel and Matlab to determine relationships between particle height and periodicity (Analysis used for chapter 3). The Final structural datasets consist of a varying multitude of images comprising >100 individually traced fibrils per WALTZ sequence. The same pixel density is maintained for all images within the datasets.

Developing a hierarchical cluster analysis

The previous method of defining fibril morphology looked at the correlation between height and periodicity, however this was not sufficient to provide clear groupings of the fibrils into classes. Therefore, Hierarchal clustering was introduced to resolve the data by finding clusters of data based on similarity. Linking these clusters to one another in a hierarchal nature forms new clusters based on the similarity or dissimilarity. This was done using agglomerative clustering or 'bottom-up' clustering which assigns each individual data point an initial cluster or based on Euclidean distance derived similarity. Euclidean distance is the straight line numerical distance between two objects within the two-dimensional Euclidean plane. Distance is determined using the root of square pairwise differences. Once the pairwise distances are defined this plane becomes a metric space which can be linked to form clusters. This is a novel approach to characterisation of fibril morphologies, therefore its implementation into the project required development and optimisation. The development of the method was

performed primarily using the VIYKI dataset and then applied to HYFNIF and RVFNIM for validation.

The initial step of the clustering process is to measure the Euclidean distance between each possible pair within a dataset. For the fibril morphology dataset, this was two dimensional with the x and y coordinates represented by average height and periodicity, respectively.

$$d_{st}^2 = (x_s - x_t)(x_s - x_t)'$$

Equation 1

Determines the Euclidean distance between vector x_s and x_t (Mathworks, pdist documentation.

Equation 1 defines the Matlab input 'Y=pdist(X)'. This command measures the pairwise distance of a matrix 'X' and returns the vector Y in which each element presents the distance between a fibril pair. Y=pdist(X) can be modified to define the type of measurement, however by default is Euclidean distance.

$$d(r,s)=min(dist(x_{ri},x_{sj})),i\in(i,\ldots,n_r),j\in(1,\ldots,n_s)$$

Equation 2

Single linkage uses the smallest distance between objects within 2 clusters. Complete linkage replaces 'min' with 'max' and uses the maximum distance. n_r is the number of objects in cluster r. *xri is the ith object in cluster r (Mathworks, linkage documentation)*

Next the vector Y is targeted with a linkage function. Equation 2 defines the Matlab input 'Z = linkage(Y)'. This function determines the dataset clustering based on the proximity between measured objects. Each subsequent cluster contained within the matrix 'Z' contains 2 objects and are therefore binary clusters. The ascension through

the matrix Z is with increasing pairwise distance, and the relationship between the closest linked pair is an object, or cluster which is in turn linked to the next proximally closest object forming a new cluster. This process is repeated forming increasingly large clusters until the entire original dataset is contained within the same final cluster. Again, the linkage function (equation 2) can be modified, and at default it performs a 'single' linkage function that forms new clusters based on the minimum distance between objects within separate pre-cursor clusters.

$$c = \frac{\sum_{i<j} (Y_{ij} - y)(Z_{ij} - z)}{\sqrt{\sum_{i<j} (Y_{ij} - y)^2 \sum_{i<j} (Z_{ij} - z)^2}}$$

Equation 3

This equation defines the cophenetic correlation between Z from the linkage function and Y from the Euclidean distance. Y_{ij} *is the distance between objects* i *and* j *in Y.* Z_{ij} *is the cophenetic distance between objects* i *and* j, *from Z.* y *and* z *are the average of Y and Z, respectively (Mathworks, cophenet documentation)*

The linkage function can determine the fit of the method and can be adjusted if needed. This is achieved by checking the cophenetic correlation coefficient (Equation 3) which relates the linkage of objects in Z to the distances between objects in Y. This is achieved using the Matlab command 'c = cophenet(Z,Y)'. A strong correlation is required to represent valid clustering, and therefore best represent the original dataset. To determine correct clustering, the correlation coefficient should be as close to 1 as possible. For VIYKI fibrils, the single linkage function provided a coefficient of 0.98 and therefore was deemed an accurate method for continuation.

The cluster tree formed by the linkage function can be visualised as a dendrogram using the command 'dendrogram(Z,P)' in Matlab. In this command, Z is the linkage data set and P is a defined number of leaf nodes in the dendrogram. While leaves define, the distance-based relationship on the dendrogram, visualised by the linkage lines; nodes in general are points in which a leaf is linked to another leaf. Axial Leaf nodes are the minimum defined clusters which are displayed on the x-axis. On the resulting dendrogram, the x axis displays data nodes that correspond to clusters containing the linked objects (in this case fibrils) and the y axis displays the similarity as a measure of relative distance. However, unless specified, P=0 (which delimits the number of leaf nodes) the lower branches of the tree will be collapsed when the number of original objects is >30. The result is the incorporation of multiple objects into the same node or group. This results in an inability to accurately determine where clusters have been made between data points, from the dendrogram alone. Therefore, for when the number of data points in the original data set exceeds the number of nodes it is beneficial to modify the dendrogram argument to return the leaf node number. This is achieved using the Matlab command '[~,T] = dendrogram(Z,P)'. T is a column vector equal in size to the number of data points in the original data set. Each original data point will be clustered into groups denoted by a value of T. The number of clusters in T can be defined, and is defined by the maximum number of leaf nodes, P. The subsequent X axis leaf nodes will be denoted by a value of T and you can observe the original fibrils assigned to each value of T which represents separate clusters. Therefore, the number of clusters represented within the cluster tree is = P which is in turn = T. Translation of the number of nodes on the dendrogram, P, to a column vector T is required to identify each original object assigned to respective clusters.

Appendix 3 is an example of a delimited dendrogram (VIYKI dataset), which displays each data point of the original set on the X axis, and the subsequent similarity-based linkage and hierarchal clustering. This displays the Raw data; therefore, you can see all the linkages between individual fibrils, and then the formation of small clusters which are then linked into larger clusters. It is interesting to see how all the fibrils are morphologically related to one another in terms of Euclidean distance; for example, fibrils 55 and 85 on the far right of the dendrogram, are weakly clustered however they are considerably dissimilar to the rest of the data set with a similarity distance of 60 separating their cluster from the rest of the data set. However, it is difficult to define class clusters from a delimited dendrogram, therefore a dendrogram with a lower number of clusters are required.

The dendrogram in Appendix 4 presents clustered VIYKI fibrils with a P value of 10, meaning that each x-axis node from 1-10 is a cluster, containing multiple data points from the original data set. Therefore, the data points in these clusters display the highest similarity in respect to one another. The dendrogram then displays 9 leaf nodes (relationship of leaf nodes to leaves is P-1). The leaves stemming from the axial nodes represent similarities between those original clusters and is defined by the Y distance of the links joining them. Cluster 5 and cluster 8 are most similar, these two clusters have been linked together forming the first non-axial leaf node which is identified as cluster 11. while all the original data points have been grouped into 10 clusters, to form additional linkages new clusters are made beginning with T+1 for every additional leaf node. You can observe which original data points are contained in each cluster using the function ‘find(T==n)’.

The dendrogram in Appendix 5 has a P value of 5, meaning that there are 5 clusters, and 4 leaves. This dendrogram displays the result of a tighter clustering, with all fibrils

assigned to 1 of 5 groups. From here you can see that the height of each leaf is higher than previously, meaning that these 5 clusters are more dissimilar than the individual clusters of the P10 dendrogram, therefore it would be unrepresentative of the natural correlation to further cluster this dataset. From this dendrogram (P5) there could be 4 clusters; 2, 4 and 5 as separate clusters as the height of their linking leaves is relatively high suggesting dissimilarity to the other clusters. Then clusters 1 and 3 have been grouped into a new cluster (cluster 6) with a similarity that is relatively low on this dendrogram, suggesting a potential merge between clusters.

Regardless, to better visualise how this clustering looks on the actual data then a scatter graph is required that has been colour coded based on the T value. To do this, we plot the original data set of height against periodicity however each data point is assigned a colour based on the T value for that point. As previously mentioned, the T value is derived from the Linkage data set Z that detects areas of correlation based on a predetermined clustering method (single or complete), which in turn was derived from the Pdist dataset which measures the Euclidean distance between each pair within the dataset. Therefore, the scatter graph is potentially more representive of what is detected by the Hierarchal clustering algorithm.

The Scatter graph (A) in Appendix 6 represents the clustering of the VIYKI dataset in to 10 groups, using agglomerative hierarchal clustering. This was achieved by measuring the Euclidean distance between every pair which determines the similarity between points and groups them. Next, the linkage function attempts to merge groups of data points based on the similarity. This example has been done using a single linkage method. This method however, despite a C value relatively close to 1 provided clustering that was seemingly un-representive of the data, extrapolating clusters from a wide range of binary clusters resulting in groups containing fibrils that are likely different

morphologies all together. Therefore, a ‘complete’ linkage function was adopted which per equation 3, C=0.997; suggesting a tighter representation of the actual data. The Matlab command can be modified to ‘Z = linkage(Y,'complete')’ which is defined by equation 2 (modified with max).

Scatter Graph (B) Appendix 6, the data has been grouped into 10 separate clusters using ‘complete’ linkage. This method separates large inclusive clusters into separate ones, however, the linkages that have been formed are also not representative of likely fibril classes. This is due to the standard method of Euclidean distance measurement with this type of data. The difference in scale of the X coordinates versus the Y coordinates means that all the height coordinates are close relative to the distance between the periodicity coordinates; therefore, the data is only being clustered in regard to the dissimilarity of periodicity as all the height coordinates are relatively similar. To confirm that this was the issue, all the height values were multiplied by 10 to put them in the same range as the periodicity coordinates. When multiplied by 10, clustering becomes more indicative of likely classes where clusters are defined per both height and periodicity. However, this identified an issue with clustering based on Euclidean distance. The straight-line distance between two points in the Euclidean plane is based on 2 dimensions (in this case the Height and Periodicity coordinates), however the measurements are taken based on the geometric locations of the data points, and therefore not considering the individual coordinates. Therefore, the Height coordinates (which only differ by a few nm) put the points in the same relative area of the Euclidean plane in comparison to the distance defined by the periodicity coordinates. Multiplying each of the X coordinates brought this issue to attention, however it is not a suitably accurate way to standardise the data points. Therefore, the Pdist function, which measures the distance between each pair of points must be modified to measure the

points using a standardised Euclidean distance; which scales the difference between each X and Y coordinate by dividing each element by the corresponding standard deviation. Equation 4 defines the Matlab command ‘Y=Pdist(X, 'seuclidean')’ which is a modified input to measure the standardised Euclidean distance.

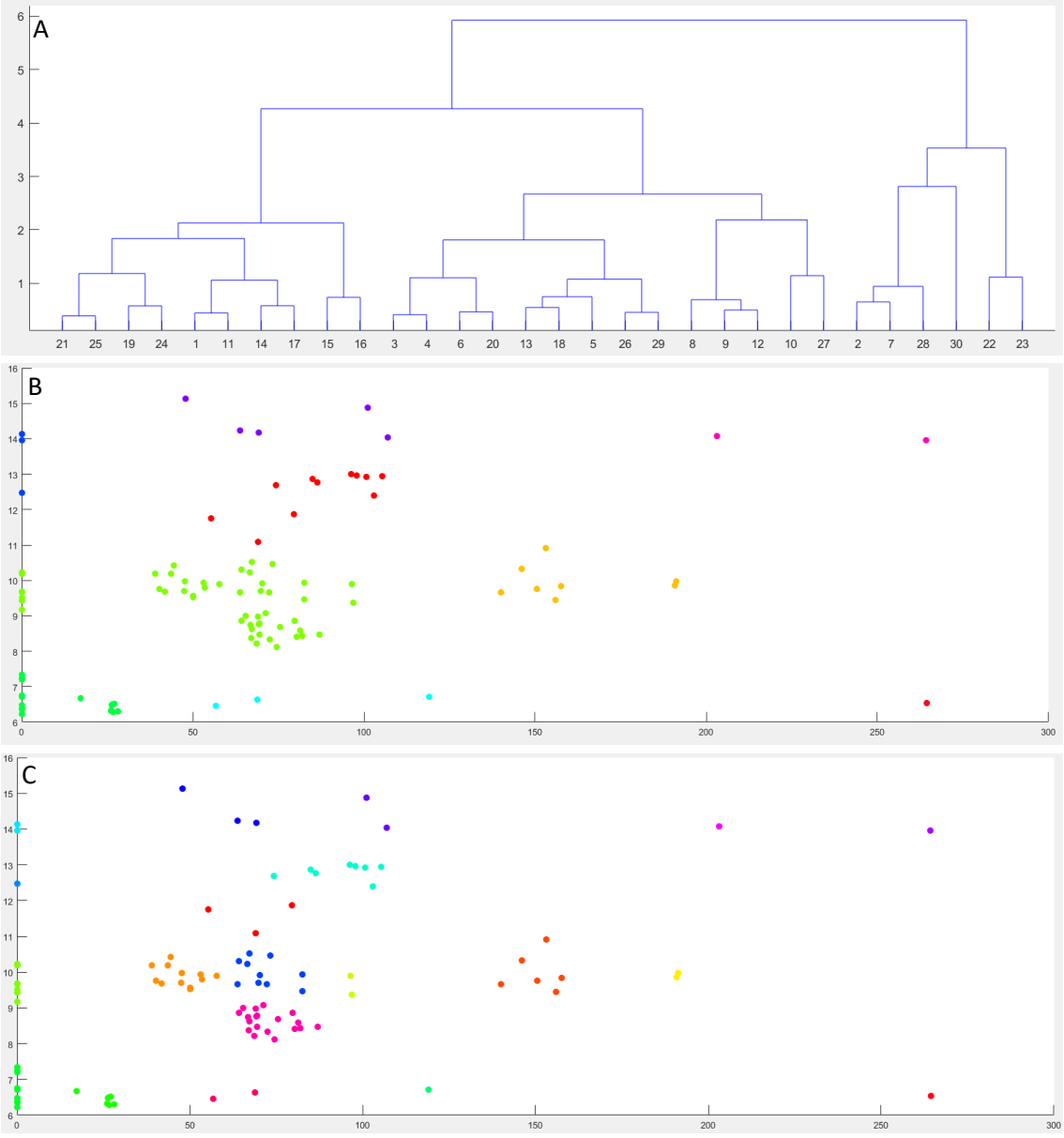


Figure 12. Graphical representation of hierarchical clustering

Demonstrating standardised Euclidean distance with complete linkage. The dendrogram in A) was constructed in Matlab and represents the 30-node linkage of clusters measured using standardised Euclidean distance (equation 7). The subsequent scatter graphs in B) and C) represent corresponding scatter graphs with complete linkage using a P value of 9 and 20, respectively.

$$d_{st}^2 = (x_s - x_t)V^{-1}(x_s - x_t)'$$

Equation 4

Determines the standardised Euclidean distance between vector x_s and x_t . V is the n-by-n diagonal matrix whose jth diagonal element is $S(j)^2$, where S is the vector of standard deviations (Mathworks, pdist documentation)

Using this method of standardised Euclidean distance, the resulting dendrogram (figure 12, A) has a Y scale that is much more representive of a data set that displays areas of high similarity, albeit the units on this axis are relative. However, when constructing a dendrogram with the default (30) number of nodes, you can see that the X axis groups are linked into subsequent groups by the leaf nodes; with 9 clusters that have a height of around 1 or less, but relatively large dissimilarity to the other clusters within the data set. These clusters likely represent separate classes of fibrils. Using the dendrogram identified potential classes, you can limit the scatter graph to represent 9 groups which have been linked based on similarity (figure 12, B). The correlation presented by this method appears to be representive of separate classes. To a certain extent, it is considering the separation of the 'major' clusters into different groups. However due the nature of fibril morphology, the addition of a single extra filament defines a new morphology. In addition to this, a change in the periodicity brought on by a change in the arrangement of filaments can also define a new fibril morphological class. Within a data set with such large differences in X and Y values, such as when comparing a 6nm fibril to a 15nm fibril; it is difficult to determine definitive class boundaries when taking in to account the slight changes in height and periodicity. The large cluster in figure 12, (B) coloured in green contains data points ranging from 10.5nm in height to 8nm in height, with periodicity values ranging from 40nm to 100nm. This variation is relatively similar in comparison to other data points in the set, however in regard to fibrils this is

clear indication of separate fibril classes. Therefore, to represent the polymorphism displayed by the Waltz peptide fibrils using this method, these large clusters need to be separated in to individual groups based on an even closer measure of similarity.

This is a simple step, and the tree just needs to be expanded by increasing the value of P. By standardising the X and Y coordinates and using a P value of 20; the resulting scatter graph shows clusters which both represent the correlation between data points and segregates the large clusters into separate groups which likely represent different fibril classes. However, the issue with this method of similarity linkage and cluster assignment is the involvement of a user defined value for the maximum number of clusters. While this does represent the data that in a way that shows the natural correlation and the likely class divides, it is a biased method that assigns a predetermined number of clusters. An alternative method to determine the number of clusters that are naturally linked in this data set, is to incorporate the inconsistency coefficient into the linkage function. The inconsistency of hierarchal clustering is a measure of the change in link height between nodes, and whether it is a change representative of a natural division in the data. This is done by determining the consistency of the links within the hierarchal tree, which involves a comparison of the height of each link to the heights of the links below in which it is attached to. When the links of sequential leaves within the data tree display heights that are similar, then these links are said to be consistent because the distance between the nodes being joined is like the distance between the sub nodes these clusters contain. However, when the height of a link considerably changes compared to the links between previous cluster sub nodes; then the data is said to be inconsistent and potentially marks a natural division in the data. Therefore, when two consistent regions of the tree are linked into a new inconsistent cluster; those two original clusters are unrelated and therefore

define clusters with high level of dissimilarity and in this application; separate fibril classes. The Matlab function used to analyse the inconsistency of a data set is I= inconsistent (Z). This function returns a matrix (I) that contains the number of links involved in the calculation, the standard deviation of the links included, the mean height of the links involved in the calculation and the inconsistency coefficient (IC) which is defined by the difference between the height of that node and the mean of the constituent links divided by the standard deviation. The inconsistency coefficient can then be used as a threshold mechanism for the number of clusters in the data set. By assigning a cut-off point, each of the linked clusters with a coefficient value equal to or lower than the threshold value, will be expanded into a single cluster consisting of all data points contained by the constituent clusters. The application of the inconsistency coefficient is to assign clusters of data into larger linked clusters until the point in the hierarchal tree where additional linkages are inconsistent; their coefficient exceeds the defined value for the cut-off. Unfortunately, it is unclear as to which threshold value is most accurately representative of the natural divisions within the data. The set value for the cut off can be adjusted to incorporate a varying number of final clusters based on how many binary clusters from the linkage matrix have a coefficient which exceeds the set value. For an unbiased approach, a final step is required to incorporate the inconsistency coefficient to determine the natural clusters within the 3 WALTZ peptide data sets. The mean IC is determined, from all values minus those equal to 0. The Leaf nodes, which are the binary clusters on the X axis, do not have a coefficient as they are the extremity of the cluster tree. The mean value is then taken as a distance cut-off instead of an IC cut-off. This requires the implementation of the 'distance' criterion in the function for clustering the data. The height of a node is measured by the distance between the two sub nodes which are connected to form that node. By defining a cut-

off value for distance, all clusters equal to or less than a node with the set value will be grouped into a single larger cluster. The mean IC values (Ii) for each of the WALTZ peptides are: VIYKI Ii = 0.89, HYFNIF Ii = 0.92, RVFNIM Ii = 0.91. However, only VIYKI is used as an example in this methodology.

To implement this threshold for natural division determination, the Matlab command 'T=cluster(Z,'cutoff',Ii,'criterion','distance')' can be used; this performs a cluster function on the linked data Z, adhering to input criteria and returning a vector, T to identify the original objects assigned to each cluster. Subsequent scatter graphs are constructed using the T value to assign class specific colours to the original objects.

Using the mean IC for the distance cut-off for VIYKI, shows un-biased clustering of the data which is representative of the natural divisions of the data and therefore the separation of classes within the fibril population. Providing a valid method for the hierarchical clustering of fibril morphologies within a quantitative dataset. However, alternate approaches are possible and further development resulted in the implementation of another method for the final cluster analysis seen in chapter 3.

Similarity distance thresholding for determining natural division

Another way to find the natural clusters within the data, is to look at the rate of change in distance or similarity across the binary clusters. This method does not incorporate the IC, and sets a threshold cut off for the hierarchal tree at a point in which the similarity within the data begins to rapidly reduce; defined by the rate of change in distance of the binary clusters formed by the linkage function. Plotting the distance between the binary clusters returns a curve which shows the change in distance across the clusters. From here you can see the point in which the relatively shallow gradient changes abruptly into a steep gradient; which is an indicator of a sharp reduction in similarity.

This point is where the clusters that have been formed, start to show relative dissimilarity and therefore additional expansion of the clusters would not be representative of the natural divisions in the data. The graph seen in (Appendix 8) shows the increase in similarity distance across the binary clusters of VIYKI. The binary clusters are formed per increasing distance in similarity from the pdist function (equation 4). You can see on the line (Appendix 8), the regions in which the gradient increases rapidly and this is the region which will be the cut-off for cluster formation.

To get the exact point on the line in which the gradient does increase relatively steeply, you can plot the difference in similarity distance between each binary cluster plotted on the distance graph (Appendix 9). First however you need to return a matrix that contains the difference values, and this is achieved with the Matlab command 'differences = diff(Z(:,3));'. This is where Z is the linkage matrix, with column 3 containing the similarity distance information. The differences can be plotted against the binary clusters formed by the linkage function and can be seen in Appendix 9.

The point on the graph in which the difference between cluster similarity rapidly increases, is the point that will define the cut-off. This is identified by the point in which the difference doubles from the previous largest value. This point located on Appendix 9 provides a binary cluster number, which can be applied to the graph in Appendix 8 to determine the similarity distance value for that binary cluster.

Once this value is determined for each WALTZ dataset, the cluster function is modified accordingly, by specifying a distance based cut-off value (N). The command 'T=cluster(Z,'cutoff',N,'criterion','distance')' returns a T vector, which when incorporated into a scatter graph results in a hierarchal clustering that accurately represents of the data, due to the fact the clusters have been identified and assigned

using the natural similarity measurements within the data. This provides an unbiased approach to identifying the correlation and therefore the class separation within the respective fibril populations. Resulting scatter graphs can be seen in chapter 3. Additionally, by plotting the number of clusters against the similarity distance you can see how the clustering changes with similarity which highlights plateau regions where changes in similarity distance does not affect the number of clusters. This depicts areas in the data with high inconsistency and therefore large natural divisions in the data. This can also be seen in chapter 3, figure 18. A combination of these methods provided a means to observe how the similarity distance defines the natural divisions in the data and subsequent number of clusters. Using this method, the WALTZ datasets used in this study were hierarchically clustered to represent the natural divisions in the data while also maintaining the ability to tighten or loosen the cluster number based on the similarity distance patterns produced by A final issue of bias with this methodology remained. Periodicity is representive of each height profiles predominant frequency. The problem here is that certain fibrils which do not display any clear periodicity have a height profile with an erratic pattern, therefore the value for the period of the most common frequency is a value that often exceeds 1000nm. Up until now such values have been zeroed to represent zero periodicity; however, the cut off for when a fibrils height profile displays no periodicity is unclear. Meaning that the cut off value for periodicity is user defined and therefore represents some bias in the clustering. To resolve this, the reciprocal value of a height profiles periodicity will instead be represented for the cluster analysis. The Frequency for a height profile for a fibril is the number of peaks in height per nm. This type of value allows the periodicity data to be represented on a plot with each point in a tighter relative proximity to each other, meaning that a threshold is not required to allow accurate clustering.

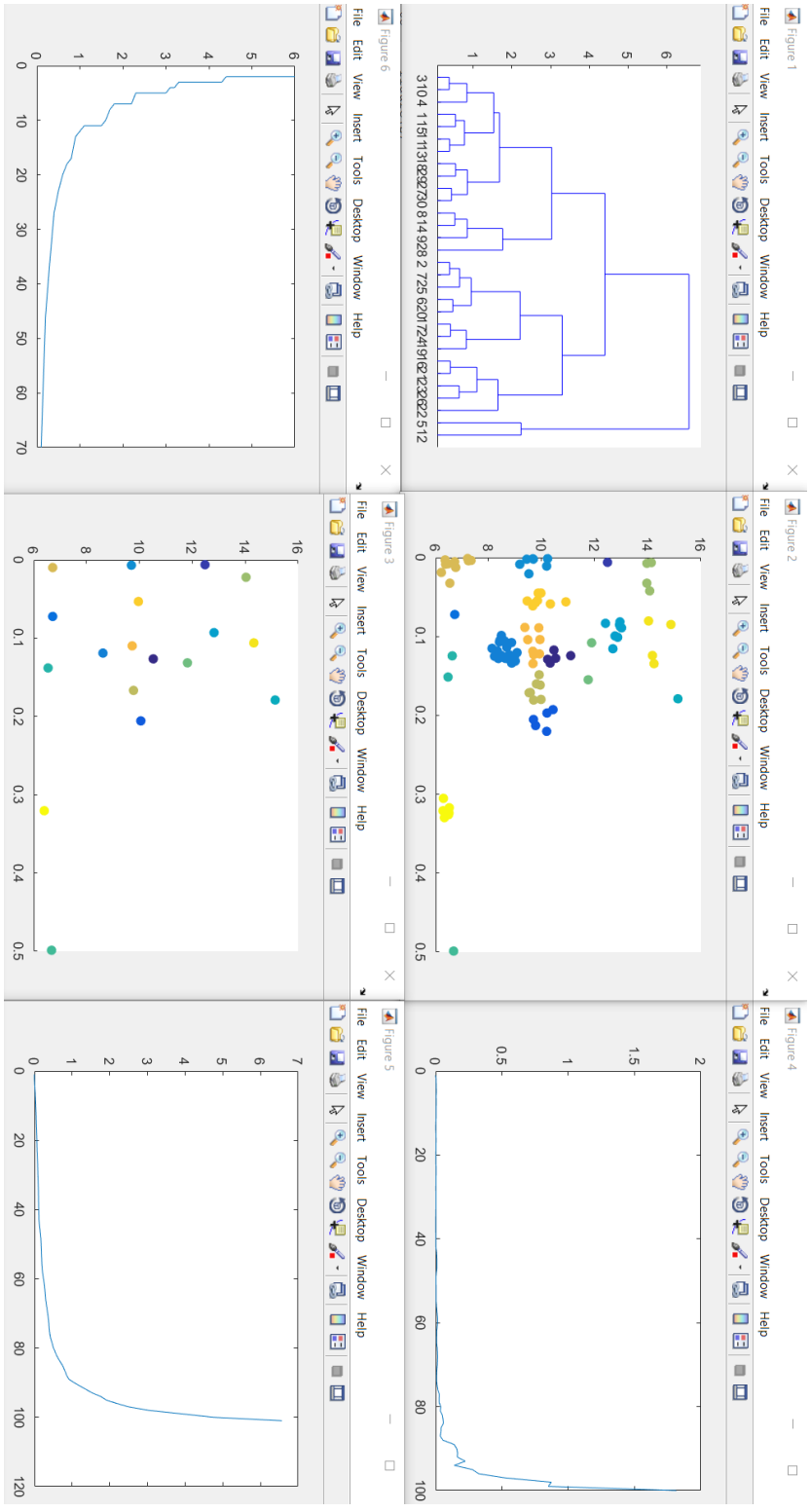


Figure 13. Output Graphs of the hierarchical cluster application

Output figures generated in Matlab using the Hierarchical clustering script developed for chapter 5. From left to right top row includes: Standardised dendrogram with 30 nodes, scatter graph using complete linkage of height vs profile frequency and then a graph displaying difference in similarity distance between binary clusters. Bottom row includes: a graph showing the change in number of clusters generated with changes in similarity distance cut off, scatter graph showing the average height and frequency of each cluster and finally a graph for the similarity distance change across binary clusters; all graphs produced in the figure are for VIYKI using a distance cut off value of 0.9 which corresponds to 18 separate clusters.

All the steps required to produce a hierarchical cluster analysis using this methodology were written as a script within Matlab (Appendix 1). This can be applied to a dataset to automatically perform the steps necessary to reproduce the analysis detailed.

The input data must be acquired using Trace-y (Xue, unpublished) then the traced polymers isolated into a structure containing all the polymers or polymer segments that are to be analysed. The structure can contain polymers traced from multiple images. For the WALTZ datasets, each input structure contained the >100 traced polymers from multiple images collected with AFM. Prior to running the script, a similarity distance threshold must be specified, Y. To acquire the value, the script is run with a value of 1 and then the graphs produced used to determine an appropriate value. Using the determined value of Y, the script can be run again. The script extracts the maximum height, minimum height, mean height, periodicity, and height profile frequency – per particle in the dataset. Agglomerative hierarchical clustering is then performed as detailed in this section, returning 6 graphs (figure 13). Each fibril from the original dataset is assigned a cluster (the number of clusters is defined by the Y value threshold) and the script returns a table 'cluster averages' which details the cluster average for the aforementioned properties. These values are determined by the average per-property value of each particle within a cluster (figure 25, chapter3). Therefore, this script automatically calculates and extracts all the required information to perform a structural characterisation and subsequent cluster analysis, on a dataset of Trace-y analysed polymers.

Fragmentation data extraction

Data collection from AFM images for the quantitative analysis comprising chapter 4 was the same as for the structural analysis. However, subsequent data extraction with the Matlab workspace varied. For the fragmentation analysis, in addition to the per-particle

height information; the population height average was required. To support this, both the per-particle and per-population average contour length was required to determine the reduction in length over time. To extract this information from the Trace-y output files, the data extraction script PolProp (Beal, code in appendix) was used with specific adjustments such as pixel density (nm/pixel). α-synuclein and WALTZ fragmentation datasets are comprised of 2 time-course repeats and 3 images per time point (6 images per time point total). Each image ranges in particle number depending on the time point, however each sample final dataset consists of >30,000 particles. Following the initial Height and length data extraction, it was evident that bias was present in the form of extremely small possibly non-fibrillar particles reducing the average fibril length; as well as the presence of fibril clumps in certain images that increases the average height. Therefore, prior to Polprop analysis each image was filtered for particles below a certain threshold length and above a threshold height; correcting for bias (Appendix 2, later incorporated into PolProp for efficiency). This was done on an individual image basis (for our database), as well as on a per time point basis (6 images per timepoint). Therefore, each timepoint seen in the data is representive of 6 images across 2 experimental repeats and filtered for bias as collated timepoints.

The information extracted for our database was:

- number of particles,
- mean length,
- weighted mean length,
- mean height,
- mean height standard deviation.

Fragmentation analysis

Once extracted and corrected using respective Matlab scripts, the data was analysed within both the Matlab and Excel workspace. The data was recorded within an Excel database allowing a quantitative comparison between WALTZ samples and α-synuclein and was the reduction in weighted mean length over time was graphically represented using Excel (Figure 31). Additional graphical representations (chapter 4) were constructed within Matlab and refined within Adobe illustrator. Height and length distributions were constructed simultaneously using a Matlab script (Appendix 10), which also provided the mean length and a probability distribution function for the height.

Persistence length

To determine the L_p of each sample (fibrils formed from RVFNIM, HYFNIF, VIYKI and α-synuclein) unfragmented fibrils were traced using the Matlab application Trace-y (Xue, unpublished). Images used for analysis were collected using AFM at a resolution of 2048 x 2048 pixel and a scan size of 30 x 30 µm, with a pixel density of 14.65nm/pixel. Each image was post-processed using to remove sample tilt and bow, in addition to the removal of image streaks which may identify as polymers. Each dataset consisted of 2000-10000 individual fibrils collected from multiple images and was analysed in Matlab using a script (Xue, unpublished) to determine the population $L_{p.}$ The script utilises equation 5, which is a probability-based method for determining persistence length.

$$p\left(\theta_{2D}(L)\right) = \sqrt{\frac{L_P}{2\pi L}} e^{-\frac{L_p \theta_{2D}^2}{2L}}$$

Equation 5.

Equation based on the Probability distribution of bending angles against length; determining persistence length (L_p) of a flexible polymer. Here L_p is defined by the probability of bending angles at a segment length (L) up to the full contour length] This allows for determination of the rate of decay of tangent vectors with increasing segment length (L).

Equation To determine a final L_p for each sample, datasets were resampled 500 times using a bootstrap method in Matlab. Each resampled dataset was analysed using the L_p Matlab script, then the Mean value and standard deviation was calculated from all resampled datasets. This provided an unbiased L_p that is not dependent on the order of polymers within the dataset whilst providing a means to determine the standard error of the final values.

Representing fibril structures with models

Fibril morphological representations were created using Autodesk Maya (Autodesk, Inc, version 6.3.16.0). To support the accurate representation of the fibril morphologies, the WALTZ peptide protofilament dimensions (Morris et al., 2013) were incorporated into the creation of the models duplicated unit; representive of individual filaments. For each WALTZ peptide, Multiple possible arrangements of the constituent filament unit were coordinated in the creation of a range of morphologies; based on either lateral or end to end interfilamentous interactions between β-sheets. From the hierarchically clustered AFM data; the per cluster maximum, minimum and average heights dimensionally informed the multistep creation process, identifying model structures which are representive of potential fibril classes on the nm level. This was achieved by

plotting the dimensions of the experimental fibril classes with that of the model structures with minimum and maximum height on the Y and X axis, respectively. While the dimensionally matched classes were assigned a structural representation, outlying clusters where fed back in to the design process. Experimental classes that were not identified by the original model-set were qualitatively analysed on a per-fibril basis and the maximum, minimum and average height were used in parallel with observations to create additional graphical representations. Observations and per-fibril measurements were made within Nanoscope analysis (Bruker, v1.5), the quantitative class information was provided by the hierarchical cluster analysis, performed in Matlab. The individual filaments are hierarchically assembled, and the resulting polymer is twisted to 80% of the degree which causes deformation within the structure. Measurements are made to characterise the resulting structure prior to cluster matching (0.1 (arbitrary units) = 1nm). Experimental periodic information was not incorporated into the creation of the model structures due the intra-class variability seen within the data.

Chapter 3

Characterising polymorphism of amyloid fibrils assembled from WALTZ peptides

Introduction

Amyloid fibrils are proteinaceous self-assembled polymers, that have adopted energetically favourable conformations compared to soluble proteins in their native state (Christopher M. Dobson, 1999). Such conformations are accessible to a whole host of proteins and peptides that are amyloidgenic either due to destabilizing events of the native protein. Alternatively, because of a sequence dependent propensity to from amyloid, as is the case for many short peptide sequences(Maurer-Stroh et al., 2010). It is thought that all proteins have the potential to form amyloid, given conditions that deem this the most energetically stable conformation (C M Dobson, 2001). There are many diseases linked to amyloid, for example Parkinson's which involves Alpha-synuclein and Diabetes mellitus type 2 which involves Amylin (IAPP). Alzheimer's disease (AD) is the most common of the fatal neurodegenerative diseases, and is an associated pathology of Aβ 40-42 and Tau cytotoxicity.

Amyloid fibrils share a common core structure in the form of a cross-β architecture (Sunde et al., 1997). The cross-β architecture is a repeating substructure that is composed of β strands that are orientated perpendicular to the long fibril axis, the β strands stack together to form a β sheet of indefinite length. Therefore, amyloid fibrils are composed of a highly-ordered arrangement of many repeating units of a protein or peptide. These β sheets will then associate laterally with one another to form protofilaments, which themselves associate to form filaments with a hydrophobic core. The individual filaments will subsequently group in varying numbers, resulting in morphologies defined by the number of filaments and the orientation in which they interact. This architecture has been supported by structural data from experiments using x-ray diffraction. Amyloid fibrils give rise to a characteristic x-ray fibre diffraction pattern with a meridional diffraction pattern at ~4.7Å which displays the interstrand

spacing between the stacked repeating β strands. An equatorial diffraction pattern at ~11Å corresponds to the lateral spacing between associated β sheets (Jahn et al., 2010b; Makin et al., 2005; Sunde et al., 1997). The interaction between individual β sheets and strands, known as the steric zipper interface defines the molecular architecture in terms of the inter and intra β sheet bonding. The types of interactions present, are therefore determined by the specific amino acid sequence, with different residues and the subsequent steric zipper pattern granting different levels of thermodynamic stability(Berryman, Radford, & Harris, 2009; J. Park, Kahng, & Hwang, 2009a).

Amyloid fibrils have been shown to have impressive mechanical properties, with tensile strength like steel and a bending rigidity comparable to spider silk (J. F. Smith, Knowles, Dobson, Macphee, & Welland, 2006). The interactions attributed to β sheet is thought to provide amyloid with such mechanical properties; observed when the level of β sheet structure is reduced within amyloid fibrils by the replacement of β strands with α-helices (Vandenakker, Engel, Velikov, Bonn, & Koenderink, 2011). This resulted in a reduction in persistence length, and therefore a reduction in rigidity. Additionally, the directionality of interacting β sheets has been shown to affect the rigidity of amyloid fibrils, with anti-parallel β sheets granting an increased rigidity. It has also been suggested a mechanical dependence on the steric zipper interface, meaning the amino acid sequence is also responsible for the mechanical stability of amyloid fibrils (Yoon, Lee, Kim, Na, & Eom, 2014). Fibril formation is thought to be a nucleation dependent process (W.-F. Xue et al., 2008) where there is a rate limiting formation of a β sheet rich oligomeric nucleus, before elongation via monomer addition. Structurally different nuclei may form, resulting in the extension of fibrils with different assembly polymorphism (Fändrich et al., 2009). Assembly polymorphism is the result of differences in the hierarchal arrangement of filaments within a mature fibril. EM and

AFM images are able to differentiate between polymorphs of a single species of fibril, by identification of the number of filaments per fibril, degree of twisting (via periodicity) and the length and diameter of fibrils (Jozef Adamcik & Mezzenga, 2012; Fändrich et al., 2009; Fitzpatrick, Debelouchina, Bayro, et al., 2013; W. F. Xue, Homans, et al., 2009). Using NMR, cryo-EM and Atomic force microscopy, transthyretin (TTR) fibrils have been characterized extensively in regards to the interactions that result in the morphology and polymorphisms of amyloid fibrils (Jozef Adamcik & Mezzenga, 2012; Jean Christophe Rochet & Lansbury, 2000). From this it seems that the association of filaments with each other is dependent on the interactions ranging from β strands in adjacent protofilaments, as well as dipole-dipole interactions and van der Waals forces that while weak and non – specific, translate into potent interactions given the repeating nature of the fibrils(Fitzpatrick, Debelouchina, Bayro, et al., 2013).

HYFNIF VIYKI and RVFNIM are amyloidgenic short peptide sequences, identified via the WALTZ algorithm, which is a residue position scoring system with the ability to isolate amyloidgenic sequences within larger proteins (Oliveberg, 2010). HYFNIF, VIYKI and RVNIM are sequences found within the Human Bloom syndrome protein, *Drosophila* Chorion protein and Human eIF-2, respectively (Morris et al., 2013). They are ideal models for exploring the sequence – structure relationship as they each display a wide range of fibril morphologies, potentially attributed to their specific amino acid sequence. using Atomic force microscopy (AFM) and agglomerative hierarchical clustering we present a novel approach to morphology characterisation of amyloid fibrils. Additionally, we provide models designed in conjunction with AFM data to predict and visualise the arrangement and twisting patterns adopted by the interacting filaments of a mature fibril.

We show that these samples, display considerable differences in both distribution and presence of certain morphologies. Suggesting the amino acid sequence is a contributing factor in the formation of different fibril morphologies, shedding light on the sequence-structure relationship of amyloid. We found that the level of polymorphism displayed by the 3 peptide sequences expanded of the classes described in previous characterisations of various amyloid species (Morris et al., 2013). In fact, the fibrils formed from these three WALTZ peptides have been previously described using a purely qualitative approach which identified twisted rope and twisted ribbon morphologies. Additionally, some quantitative approaches have been taken to determine the polymorphism of fibril populations, for example the distribution of different morphologies within populations of fibrils formed from the hexapeptide CapFF has been characterised showing the formation of additional morphologies over time(Jozef Adamcik et al., 2011). Previous characterisations have shown a linear and stepwise increase in height with the addition of new filaments, which most commonly conform to a twisted ribbon structure(Fitzpatrick, Debelouchina, Dobson, et al., 2013)(Jozef Adamcik et al., 2010). Further characterisation highlights the presence of additional structures such as helical ribbons, twisted ropes and nanotube like morphologies (Jozef Adamcik et al., 2011; Jozef Adamcik & Mezzenga, 2012; Usov et al., 2013). Our approach links aspects of previous characterisation techniques into an extensive quantitative characterisation of the 3 fibril populations. We expand on previous morphologies, suggesting a broader range, determined by both the arrangement and the number of filaments within a fibril. Our Hierarchical clustering emphasises variety in the fibril characteristics within a population, presenting a way to tailor the complexity of classification.

Recently the clinical relevance of this has been highlighted by showing that amyloid fibrils maintain their heterogeneity *in vivo* by analysing the polymorphism displayed by fibrils extracted from a patient suffering from light chain amyloidosis (Annamalai et al., 2016). Differences in morphology likely result in changes to the mechanical properties of fibrils, with the number of filaments and type of twisting giving rise to structures that differ in rigidity. This adds an extra narrative to our results, as the morphology and subsequent mechanical properties of amyloid may have implications in a length dependent cytotoxicity mechanism. Finally, the structure adopted by most the fibrils within a population is thought to be representive of the local minima in the aggregation energy landscape (L. Wang et al., 2010). Our data shows that the most energetically favourable morphology differs per the amino acid sequence. Which has implications in a biotechnology role where short peptide sequences may be designed to form amyloid with a desired set of structure dependent mechanical properties.

Materials and methods

WALTZ Peptide synthesis

Each of the WALTZ peptides were synthesised with an N-terminal acetylation and a C-terminal amidation. Multistage synthesis using an Fmoc protection mechanism created lyophilised powder which had a HPLC achieved purity >95% (JPT peptide technologies. Separate batches were synthesised on site using the same method (Kevin Howland, University of Kent).

In vitro polymerisation

To prepare populations of WALTZ peptide fibrils 1mg of the respective lyophilised powder is suspended in 100ul of 0.2µm syringe filtered filter sterilized milli-Q water to a final concentration of 10mg/ml. Due to the potential instantaneous formation of fibrils, the solution is not vortexed to mix; avoiding potential fragmentation. Pipette mix the solution to completely dissolve the peptide powder. The solution is incubated at room temperature for 1 week prior to experiments and imaging. This time allows for 'maturation' of fibrils providing equal time for morphological adoption across all 3 WALTZ samples prior to imaging. All manipulations of the WALTZ peptide solutions were with low bind tips (VWR 53503-294 (with SpidersilkTM surface)), and Protein LoBind 1.5ml tubes (Eppendorf, 022431081).

AFM sample preparation and imaging

Each WALTZ peptide sample is diluted to 0.05mg/ml using filter sterilized milli-Q water and 10% pH1 HCL (final solution pH2). 20µl is subsequently deposited on a freshly cleaved mica surface (Agar scientific, F7013) attached to steel discs (Agar scientific) surface incubated on the mica surface for 10 minutes. Following incubation, the sample is washed with 1ml of filter sterilized milli-Q water, then dried using a stream of nitrogen gas with filter paper to aid in the removal of residual liquid from the mica surface. The samples were then imaged using a Bruker Multimode AFM with a Nanoscope V controller operating under peak force tapping mode. Equipped with a ScanAsyst probe (silicon nitride triangular tip with tip height = 2.5-8µm, nominal tip radius = 2nm, nominal spring constant 0.4N/m and nominal resonant frequency 70kHz. The imaging software used is Nanoscope 8.15. Images for structural database were collected with a scan size of 6 x 6µm and at a resolution of 2048 x 2048 pixels. A scan rate of 0.305Hz was used with a noise threshold of 0.5nm (manually set within the nanoscope

parameter selection workspace). Additionally, Z limit was reduced to 1.5µm and the peak force set point set to 0.05- 0.15mV (depending on sample).

Hierarchical clustering

AFM images of WALTZ peptide fibrils were analysed using the in-house Matlab application Trace-y (Xue, unpublished). The application automatically searches for particles within AFM images based on input parameters including a z distance (nm) threshold, apparent width of the particles (pixels) and search Phi (°). Required height profiles of traced fibrils was subsequently extracted within Matlab, constructing per sample datasets consisting of 100 fibril particles; the data set consisting of multiple AFM images. Each fibril dataset was then analysed using the Hierarchical clustering script developed within Matlab (Appendix 1). The script extracts the average height, maximum height, minimum height, and profile periodicity per-particle. Periodic information is determined using Fast-Fourier transform which is implemented into the script. For each WALTZ peptide sample, each particle is then hierarchically clustered in regard to the displayed average height and periodicity. Used within the method, is the pair-wise 2D measurement of standardised Euclidean distance which is subsequently linked using a 'complete' linkage function (Equation 2). This determines the similarity between clusters I.e. particles within a cluster have high similarity and vice-versa. Each fibril population is then designated into clusters based on a similarity distance cut – off, in which particles residing within preliminary clusters with similarity distance below the threshold value are collapsed into final clusters. Each cluster represents a possible morphological class, and through the hierarchical tree are linked within a narrowing number of clusters with increasing inconsistency (inconsistency describes regions within the cluster tree with expanding dissimilarity). Dendrograms and scatter graphs were

constructed within Matlab, with data points colour coded depending on the assigned cluster.

Representing fibril morphologies

Fibril morphological representations were created using Autodesk Maya (Autodesk, Inc, version 6.3.16.0). To support the accurate representation of the fibril morphologies, the WALTZ peptide protofilament dimensions (Morris et al. 2013) were incorporated into the creation of the models duplicated unit; representive of individual filaments. From the hierarchically clustered AFM data; the per cluster maximum, minimum and average heights dimensionally informed the multistep creation process, identifying model structures which are representive of potential fibril classes on the nm level. This was achieved by plotting the dimensions of the experimental fibril classes with that of the model structures with minimum and maximum height on the Y and X axis, respectively.

Results and discussion

High-resolution atomic force microscopy

AFM is a high-resolution imaging technique that allows for the acquisition of 3D data of the fibril samples to high detail. The size of amyloid fibrils vary, however generally they are <20nm including the fibrils formed from HYFNIF, VYKI and RVFNIM (Morris et al., 2013). The images collected for this quantitative analysis were identical in their imaging parameters. Resulting in an equal pixel density for each image and therefore an equal level of both visual fidelity for qualitative observations, and height profile resolution for accurate and consistent quantitative analysis. Images were collected from randomly selected 6µmx6µmareas at a resolution of 2048 x 2048 pixels, resulting in a pixel density of 2.93 pixels per nm.

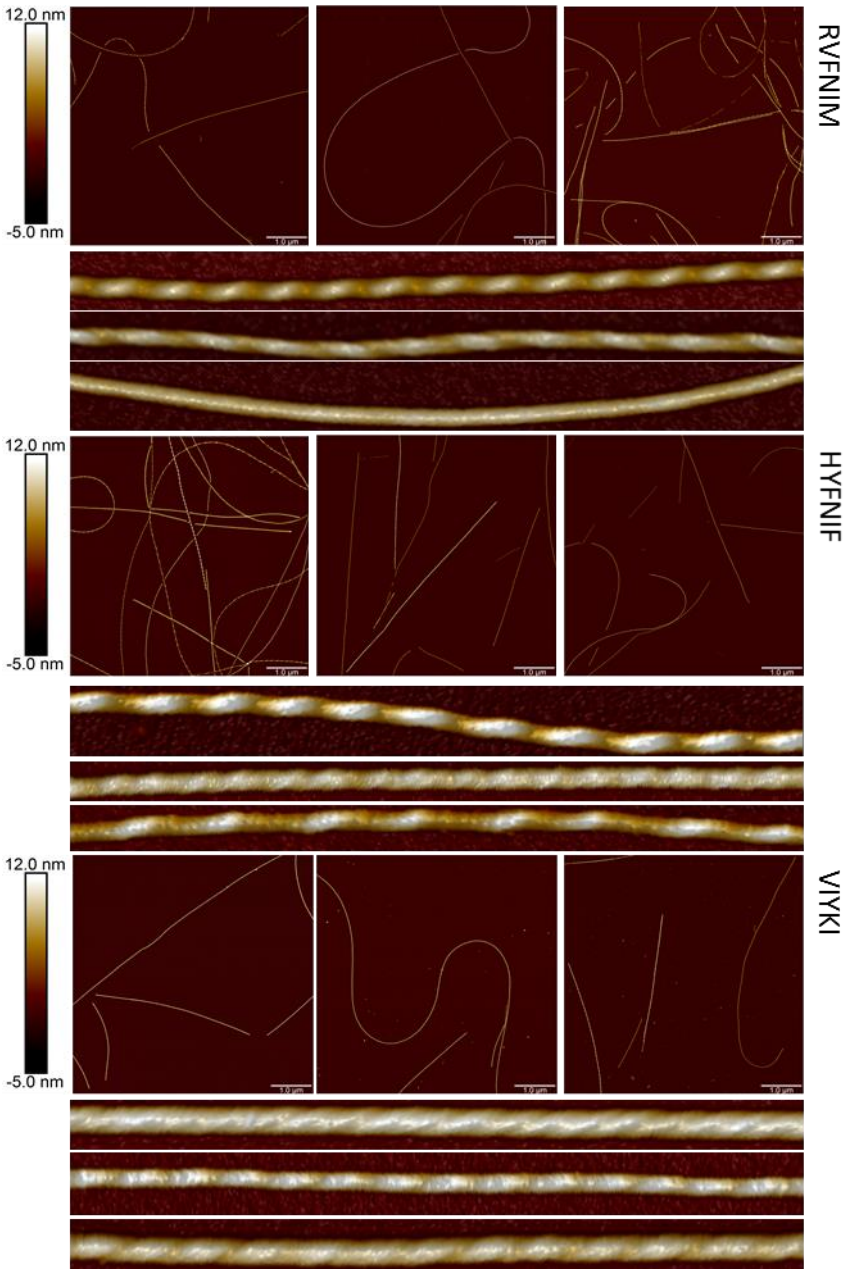


Figure 14. Using AFM to observe fibril polymorphism

AFM images of fibrils formed from RVFNIM, HYFNIF and VIYKI. Each of the larger images represents a 6 x 6 µm surface area and were captured at a resolution of 2048 x 2048 pixels. For each peptide sample, the figure displays 3 sample images included in the characterisation data sets, and 3 examples of the respective morphologies seen for the fibril populations. The colour bar represents the height contrast, ranging from 12nm to -5nm and the scale bar is equal to 1 µm.

Figure 14 shows AFM height images containing fibrils formed from RVFNIM, HYFNIF and VIYKI. The images show long unbranched polymers measuring >6µm in length. In addition to the displayed height of ~10nm (from colour scale bar), these polymers display the expected height characteristics of amyloid fibrils. Images of this resolution

allow for clear visual discrepancy between fibril morphologies. These are defined by differences in height, twisting patterns and periodicity. These images allow for direct observation of morphologies. The height contrast of each image is depicted by a height specific colour change, therefore the height profile along a fibril axis can be predicted by the colour profile along the fibril. The fibrils displayed conform to previously identified morphologies within populations of fibrils formed from other peptides or proteins (J Adamcik & Mezzenga, 2011; Jozef Adamcik & Mezzenga, 2012; Usov et al., 2013). The peptides give rise to morphologies beyond the commonly identified twisted ribbon morphology, with examples indicative of twisted ropes, helical ribbons and a potentially tubular morphology.

Amyloid fibrils predominantly form left handed twists (Lara et al., 2014; Meinhardt, Sachse, Hortschansky, Grigorieff, & Fändrich, 2009) with the WALTZ peptides in this study shown to form left handed twists in their morphology (Morris et al., 2013). This attribute is potentially determined by the specific residue interactions between filaments within a fibril, and protofilaments within a filament. Within the protofilament itself, the residue interface; steric zipper or otherwise results in twisting of beta the sheets around one another (Wasmer et al., 2008c). Therefore, it Is generally considered there is a residue encoded propensity to form twists of a handedness. The twisting observed within amyloid fibrils is a favourable confirmation that increase the stability towards shearing, by reducing the repulsive forces experienced between interacting protein molecules (Jozef Adamcik et al., 2010). Figure 14 shows the presence of right handed RVFNIM fibrils and there are multiple morphologies which display a right-handed twist (Figure 19). Most the fibrils seen within the RVFNIM population are left handed. VIYKI and HYFNIF do not display any clear examples of right handed twists. However, there are some possible examples amongst fibril classes that do not show

clear periodicity (figure 19). Generally, you would not expect to observe both left and right handed fibrils from the same monomeric precursor; however, it has been shown that under certain conditions you can switch the handedness of fibrils during fibril formation for example by changing the pH of growth conditions for insulin fibrils (Kurouski et al., 2012). Additionally, it has been shown that upon extended incubation of bovine serum albumin fibrils, the formation of rigid fibrils in which the persistence length exceeds the polymer length results in an expansion of the morphology landscape resulting in additional structures, including right handed twisted ribbons in the presence of left handed structures (Usov et al., 2013). Such a transition would be considered as the next stage in the hierarchical tree of amyloid structures, in which two mature fibrils interact to form a larger subsequent fibril Therefore, the data collected here show both left and right handed fibrils present in a single fibril population.

The issue however with relying on visual observations from AFM images, is the tip convolution. The intricacy of individual filaments twisting around one another is difficult to detect due to the exaggeration of the apparent width seen within the images. Therefore, the height profile must by independently analysed on a per pixel basis. The tip of the probe used for imaging here has a nominal tip radius of 2nm and a maximum of 12nm. This similarity in size of sample and probe results in tip-sample convolution, an artefact of AFM imaging negating the value of apparent width to a structural analysis such as this. Therefore, the height profile of the fibril is used to determine the dimensions of a fibril. If a fibril is cylindrical then the height is equal to the circuital diameter. If the fibril is dimensionally unequal then given the twisting nature of amyloid fibrils the height profile will allow the acquisition of the widest point and the narrowest point of the fibril (Volpatti, Vendruscolo, Dobson, & Knowles, 2013).

Height of fibril morphologies

The AFM produced height profile of an amyloid fibril is indicative of the morphology. The periodic fluctuations in height and the periodicity can suggest the number of filaments and how they are arranged within the mature fibril. To extract the required information from AFM images, we used an in house Matlab application that can trace individual fibrils within an image and collect the height information of each fibril ridge pixel. The application can detect the direction of the fibril and its boundaries. The key data that is extracted using this method, is the height profile of each traced fibril. This information can be used to characterise a fibril in regards to its super-structural morphology. Additionally, the contour length and bending angles of each fibril can provide the information required to characterise mechanical properties such as the persistence length; therefore, providing an effective tool in the morphological characterisation of amyloid fibrils.

To determine the polymorphism present within each fibril population (RVFNIM, HYFNIF and VIYKI) Characterisation was conducted based on a relationship between average height along a fibril axis and the periodicity displayed. In addition to this, fibrils which display a twisted rope morphology for example with have an average height closer to the maximum height as the change in height along the fibril axis will be comparatively small. Ultimately, each morphology would likely possess a characteristic height and periodicity profile given the previously defined and observed structures (Morris et al., 2013).

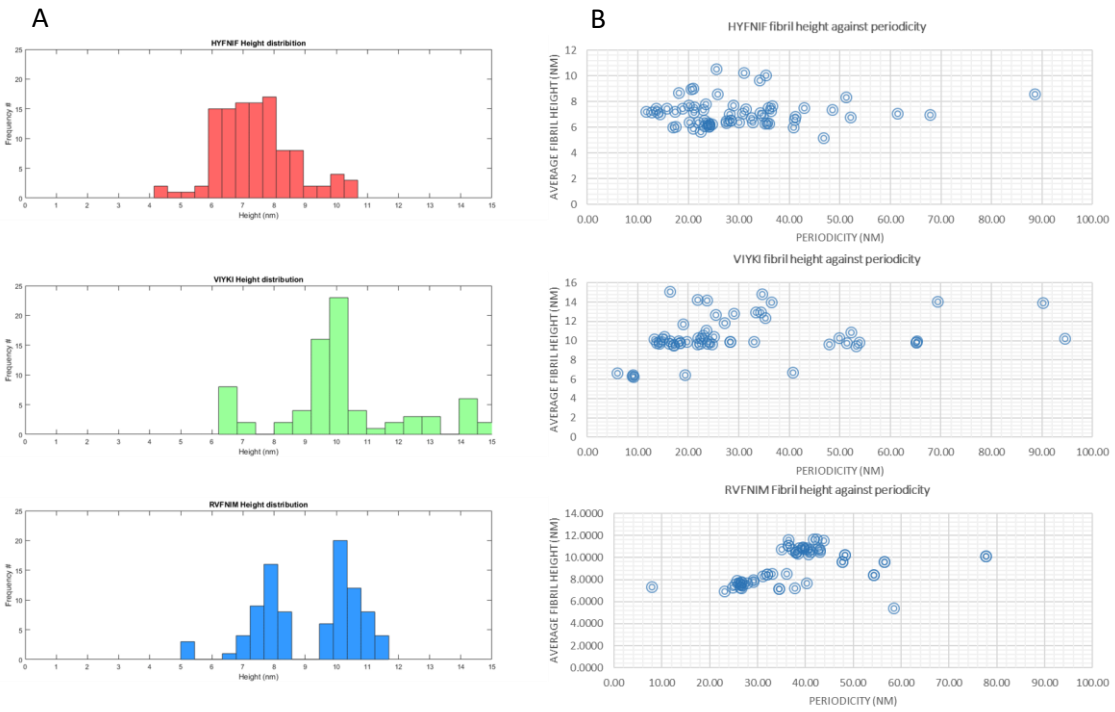


Figure 15. Detecting fibril classes based on height and periodicity

Histograms for each of the WALTZ peptides, showing the frequency of average fibril height for each of the data sets (A). Isolated, this information suggests fibril classes however when plotted in addition to the particle periodicity the classing becomes convoluted with no clear separation (B).

Figure 15 shows amongst the ~100 analysed fibrils for each population, there is a wide range of observed heights, with height ranges more frequent than others. However, when paired with the periodic behaviour of the respective fibrils, there is varying to little correlation. The correlation presented provides an ambiguous prediction of class profiles at best, however it does suggest that the respective peptide sequences can form an expanded range of morphologies than what has previously been observed(Morris et al., 2013). The height profile of each fibril is a wave with contour length (nm) on the x-axis and height (nm) on the y-axis. This is then Fast-Fourier transformed (FFT) using an in house Matlab application to provide both a frequency/nm and a peak to peak distance known as the periodicity (nm) (figure2X). The application was further modified to provide the dominant value for both frequency and periodicity, without visualisation. This allowed for the batch analysis of fibril data sets (~100 fibrils). Both Periodicity and

frequency provide information regarding the twisting pattern of a fibril by defining the distance between peaks and the twist frequency, respectively.

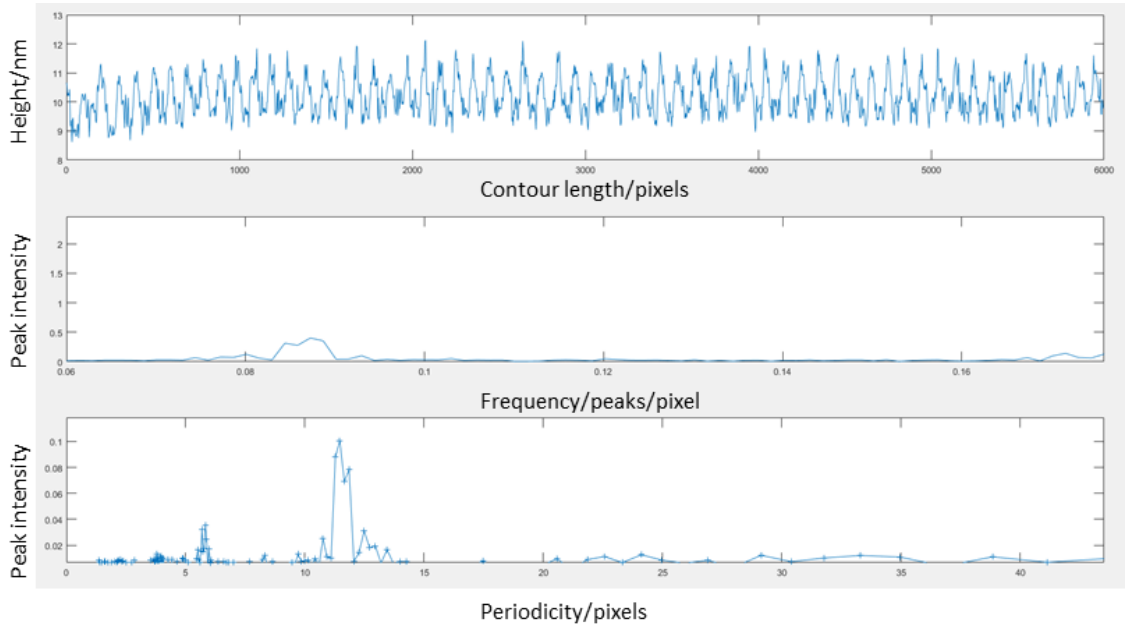


Figure 16. Periodicity of a fibril determined by FFT

This shows a fibril formed from RVFNIM, which has been FFT to determine the periodicity of the height profile. The Matlab application analyses the height profile of the traced fibril (top) and determines the peak frequency of the profile (middle) and the predominant periodic repeat (bottom). Output values are in pixels requiring conversion to nm.

The morphology landscape of amyloid means that multiple morphologies are accessible to a pool of amyloidgenic protein or peptide monomers, resulting in characteristically heterogenous fibril populations. However, the individual classes e.g. a 3-filament twisted ribbon or a 4-filament twisted rope; should be relatively homogenous given the predictable and consistent nature of protein – protein interactions within an individual morphology. Such as the beta sheet side chain interdigitation and the inter-protofilament hydrogen bonds between terminal residues of separate beta strands, which themselves are homogenous in sequence and therefore dimensions(Fitzpatrick, Debelouchina, Bayro, et al., 2013). The properties of the monomer material and the side chains presented along a single beta strand may be directly related. Elongation is

potentially driven by hydrophobicity and structural arrangement may be dictated by aromaticity and charge of residues (K E Marshall, Morris, Charlton, & O'Reilly, 2011; Tu & Raleigh, 2013). Therefore, given the homogeneity of the monomer precursor, the energetically favourable interactions and conformations adopted for each morphology, would be expected to show little intra-morphological variation.

Therefore, it is possible that the variation observed in our data (Figure 15) is representative of an expansion of the morphology landscape, with an increase in the number of different fibrillar structures that are accessible. A fibril consisting of 3 filaments (6 protofilaments) may have multiple possible arrangements, in which the likelihood of each is dependent on the sequence specific properties of the monomer unit. However more than one arrangement may be similarly favourable in terms of mechanical and thermodynamic stability. This adds variation in the type of morphology that an 'n filament fibril' can adopt; as n increases so does the number of possible structures. This is seen with the incubation of fibrils over time, where larger alternate morphologies are formed through the interaction of multiple smaller fibrils (Usov et al., 2013)(Jozef Adamcik et al., 2011). This may be a mechanically favourable transition for the fibrils given their extended length over time, an increase in rigidity through the culmination of multiple fibrils may increase their stability towards breakage. Such a transition may also become favourable due to the extended length of a fibril with time; an increased array of individually weak interactions provides a possible net interaction that is overcoming of the electrostatic repulsion between fibrils. This can be viewed as the next step in the complex hierarchal arrangement of amyloid structure's that encompasses the transition of beta strand to mature fibril.

A Hierarchical cluster analysis to define fibril classes

Our previous method of defining fibril morphology looked at the correlation between height and periodicity, however this failed to provide clear differentiation of the fibril classes. Hierarchal clustering has therefore been used to resolve the data and define the natural divisions within the dataset, which may mark the boundaries between fibril classes. This was achieved using agglomerative clustering or 'bottom-up' clustering which assigns each individual data point an initial ‘binary cluster’ based on Euclidean distance derived 2-D similarity between every possible pair of data points (pairwise distance between pairs of objects- pdist function). These clusters are then hierarchically linked using a defined linkage function (complete linkage) throughout a multibranch tree in which cumulative clusters are formed with increasing dissimilarity until all data points are linked within the same final cluster.

The Hierarchical clustering was performed using a purpose built Matlab script to provide the required sequential analysis required for the figures seen in figure 18. The input data was the respective fibril databases for RVFNIM, HYFNIF and VIKYKI each with more than 100 individually traced and characterised fibrils (using in house application). The 2 attributes that were measured and linked, were the average height of each fibril and its profile frequency. Certain fibrils do not display any clear periodicity and have a height profile with an erratic pattern, therefore the value for the period of the most common frequency is a value that often exceeds 1000nm. To resolve this, the periodicity of the fibrils will be represented by the reciprocal value of a height profiles period; or the frequency. The Frequency for a height profile for a fibril is the number of peaks in height per nm. This type of value allows the periodicity data to be represented on a plot with a better scale.

The cluster tree can be visualised as a dendrogram, with the horizontal axis displaying data nodes that correspond to the original data objects or data points (in this case each fibril) and the vertical axis displaying the similarity as a measure of relative distance. However, unless specified in the input argument the lower branches of the tree will be collapsed when the number of original objects is >30. The result is the incorporation of multiple objects into the same node or group. However, there is in an inability to accurately determine which original data points are contained in each cluster (leaf nodes on the x-axis figure 18). Therefore, for when the number of data points in the original data set exceeds the number of nodes it is beneficial to modify the dendrogram argument to return the leaf node numbers T. T is a column vector equal in size to the number of data points in the original data set. Each original data point will be clustered into groups denoted by the value of T. The number of clusters in T can be defined, and in this example, it is defined by the maximum number of leaf nodes P. This will determine the number of clusters in the column vector T and additionally the dendrogram will display P-1 leaves. The inconsistency of hierarchal clustering is a measure of the change in link height between nodes, and whether it is a change representative of a natural division in the data. When the links of sequential leaves within the data tree display heights that are similar, then these links are said to be consistent because the distance between the nodes being joined is similar to the distance between the sub nodes these clusters contain. However, when the height of a link considerably changes compared to the links between previous cluster sub nodes; then the data is said to be inconsistent and potentially marks a natural division in the data. Therefore, when two consistent regions of the tree are linked into a new inconsistent cluster; those two original clusters are unrelated and therefore define clusters with high level of dissimilarity and in this application; separate fibril classes. The

change in similarity as you ascend the hierarchical tree allows the differentiation between likely fibril classes through detection of the natural divisions. Therefore, to determine how many fibril classes are displayed by each of the respective fibril populations (RVFNIM, HYFNIF and VIYKI) we used the change in similarity across the clusters, to detect the point in which the tree becomes inconsistent, and dissimilarity increases.

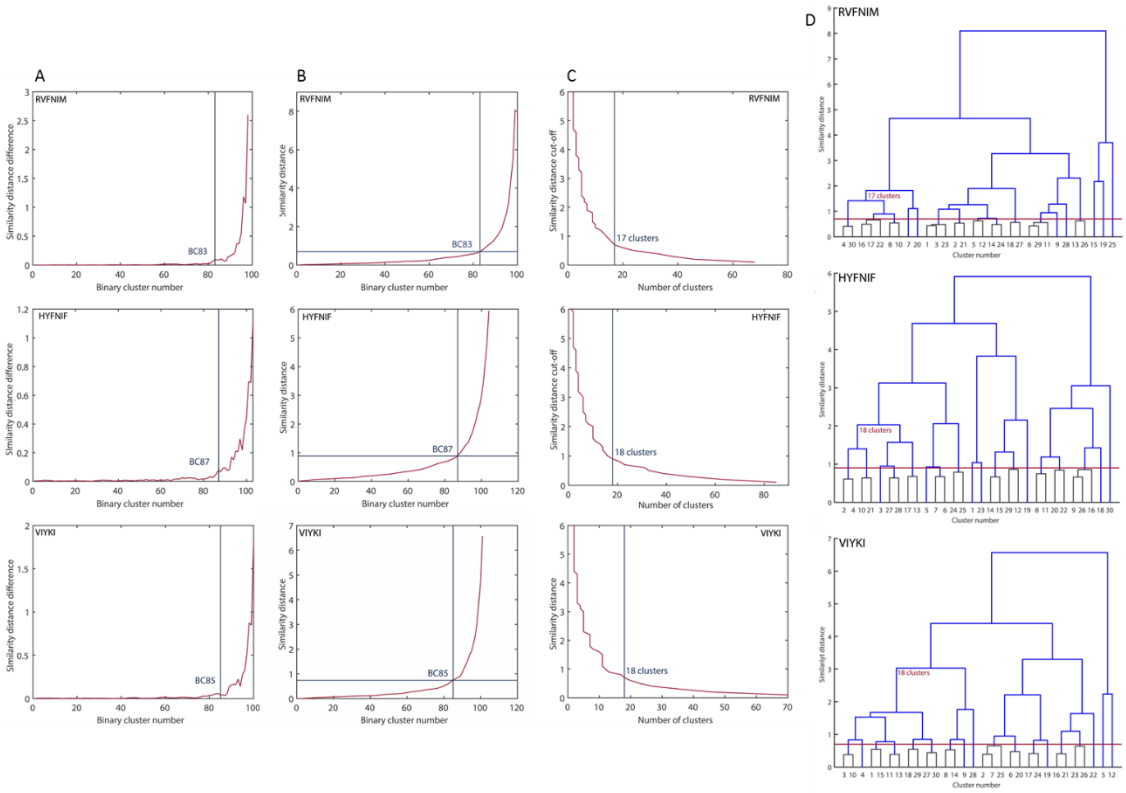


Figure 17. Method for determining natural division within clustered data

The difference in Similarity distance between binary clusters is plotted in (A) which is used in conjunction with the overall change in distance across the binary clusters (B) to determine the point in which the data becomes inconsistent and represents the maximum cluster number for natural divisions in the data (marked with blue lines; binary cluster number noted). The point in both graphs where similarity constantly decreases does not necessarily represent the cluster number which best fits the fibril data; rather provides a maximum likely value that can be reduced depending on the change in similarity distance cut off. This is visualised in (C) where the number of clusters produced is plotted against the similarity distance threshold. The blue line is the threshold derived from (A) and (B) and is at a point where equal reductions in similarity distance cut-off cause a consistently increased rate of cluster expansion. Increases in cut-off from this point results in smaller changes to the number of clusters representing the data, as the data becomes inconsistent. This distance value was used to determine a cut-off value for each dendrogram (D) which consists of 30 leaf nodes. The red line represents the similarity distance threshold, wherein linked nodes below this threshold (black leaves); are grouped into larger clusters that represent fibril classes. The threshold formed clusters, and the original unaffected nodes represent the natural divisions in the data and therefore the number of fibril classes. HYFNIF has 18 clusters, VIYKI has 18 and RVFNIM has 17; representing likely fibril classes. All figures are produced with the Matlab script seen in figure 21.

To find the natural clusters within the data, we looked at the change in similarity distance across the binary clusters. This method sets a threshold for the hierarchal tree at a point in which the similarity within the data begins to rapidly reduce; defined by the change in distance of the original binary clusters formed by the linkage function. Plotting the distance between the binary clusters returns a curve which shows the change in distance across the clusters (figure 17B). From here the point in which the relatively shallow gradient changes abruptly into a steep gradient, is an indicator of a sharp reduction in similarity. This point is where the clusters that have been formed, start to show relative dissimilarity and therefore additional expansion of the clusters would not be representative of the natural divisions in the data. Additionally, plotting the number of clusters against the similarity distance (figure 17A) shows how the clustering changes with similarity which highlights plateau regions where changes in similarity distance does not affect the number of clusters. This depicts areas in the data with high inconsistency and therefore large natural divisions in the data. The dynamic nature of this clustering technique means that a cut-off value can be defined to tailor the complexity of the hierarchical tree, assigning individual fibrils into classes and then linking these classes into new larger clusters. Of course, the wider the data is linked the more variation you will observe within an individual cluster; however, this may be of use in certain applications of polymorphic quantification. For example, to quantify the frequency of a common fibril morphology for a certain amyloid against all fibrils that do not adopt the same consensually agreed common morphology regardless of diversity within that second cluster. Additionally, in conjunction with qualitative data this technique allows the allocation of possible ‘sub-classes’ into parent classes. Such as classes with the same apparent filament arrangement and height, however they differ

in periodicity. These can be clustered into the same parent class, or into 2 different classes based on the defined cut-off value.

Here we wanted to detect and classify all possible fibril classes that adhered to the primary natural divisions of the data. Therefore, the overall change in distance across the binary clusters was used to find the point in which similarity begins to rapidly reduce (Grey line figure 17B). This point should also correspond to the point in figure 17A in which the difference in similarity distance between 2 clusters begins to increase. When the similarity distance at these points is used as a cut-off in figure 17C, the number of clusters begin to rapidly expand with a further reduction in similarity distance cut-off. Prior to this point also marks a reduction in the rate that clusters are linked. Vertical plateaus represent regions in which relatively large changes in similarity distance cut-off does not affect the number of clusters formed. Determining this value allowed the identification of the minimum number of clusters that are formed prior to a rapid increase in dissimilarity or where the data begins to become increasingly inconsistent. This represents the maximum natural division in the data, where a higher number of clusters is likely not representative of the data. The initial point for RVFNIM on figure 17B was a similarity distance of 0.7, attributed to binary cluster 83. This marked a region on figure 17C that defined 17 clusters. For VIYKI, a value of 0.9 marked the most significant change in similarity distance, however this did not correspond to the initial region in which cluster formation desensitises to changes in distance cut-off (figure 17C). Therefore, a value of 0.7 (binary cluster 85) was used, which marks both a considerable change in similarity distance rate, and a region in which cluster formation becomes less sensitive to changes in distance cut-off, providing the consensus for finding natural divisions. The initial point for HYFNIF on figure 17B was a similarity distance of 0.9 attributed to binary cluster 87 and is indicative of a subsequent rapid

change in similarity distance. This marked a region on figure 17C that defined 18 clusters prior to which the cluster-cut-off sensitivity only decreases. Clusters were then further characterised regarding the morphologies present in each and any additional linkage of clusters that are on the border of the cut-off value were also determined. in other words, the initial cut-off is not an absolute representation of class division, it is a guideline. These originally selected clusters were to be manually linked into a smaller number of clusters using a qualitative modelling approach.

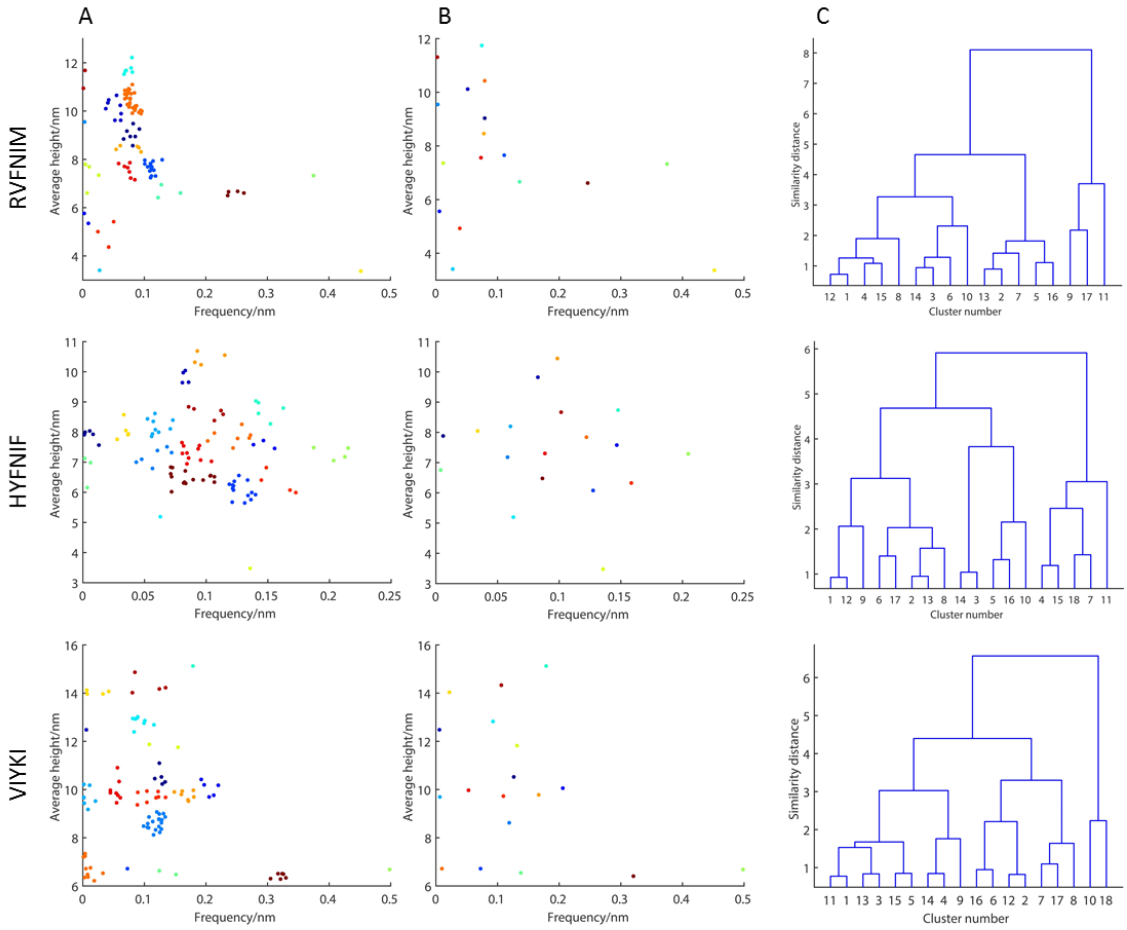


Figure 18. Graphical representation of the maximum number of likely clusters

Using the similarity distance cut-off determined in figure 17, the cluster analysis can be presented as a scatter graph; with each fibril particle colour coded to the respective clusters identified (A). The graphs seen in (B) represent the average cluster values; particle height (nm) and height profile frequency/nm. Both graphs are produced using the Matlab script seen in figure 21. Dendrogram seen in (C) for each of the WALTZ peptides represents the similarity relationship between each of the identified clusters, which comprise the x-axial nodes.

The rationale for selecting these cut-off values can be further supported by Figure 17 A which displays the difference between similarity distance of binary cluster n and n+1. The point on the graph (marked) in which distance difference increases and continues to only increase from that point on. This helps differentiate from areas in the cluster tree where consistency can decrease and then increase again. Therefore, further supporting the identification of the region in the cluster tree that defines natural boundaries.

These cut-off values were then used to filter the 30 cluster dendrogram seen in figure 17 D. By adding a cut-off value based on similarity distance, the result is a collapse of the tree in which all linked nodes below this threshold are assigned to the same cluster; thereby forming new cumulative nodes. These new nodes represent the maximum number of likely fibril classes within the data, and the additional similarity between these new classes can be seen on figure 18 C. To visualise how this clustering looks within the actual data then a scatter graph is required that has been colour coded based on the cluster assignment value (T). To do this, we plot the original data set of heights against periodic frequency/nm; however, each data point is assigned a colour based on the T value for that point (figure 17). As previously mentioned, the T value is derived from the Linkage data set (Z) that detects areas of correlation based on a predetermined clustering method (complete), which in turn was derived from the Pdist dataset which measured the Euclidean distance between each pair within the dataset. Therefore, the scatter graphs seen in figure 18 are representive of the fibril classes detected for the 3 WALTZ peptides. Each identified cluster contains variation between the individual data fibrils, however the similarity between fibrils must be considered in relation to the entire dataset. The fibrils within each cluster display a similarity that is below the defined cut-off, while the 2 furthest points between 2 clusters measure a similarity

distance above that threshold. Therefore, both the relative intra-cluster similarity and inter-cluster dissimilarity is high. There are 17 clusters for RVFNIM and 18 clusters identified for HYFNIF and VIYKI. Each cluster has been averaged and displayed as a single point on figure 18 B. On these plots, you can see the discrepancy that arises regarding a direct filament to height relationship. The basis of this relationship is that with each additional filament, you will net an increased height and periodicity (Jozef Adamcik et al., 2010; Fitzpatrick, Debelouchina, Bayro, et al., 2013). Of course, this is only true for fibrils that conform to a twisted ribbon morphology. In Figure 25 fibrils with similar heights are observed to display highly varied periodicity; for example, VIYKI displays fibrils with an average height of ~10nm and a frequency (nm^{-1}) of ~0.02. However, there are also fibrils present with a height of 10nm and a frequency (nm^{-1}) of ~0.2. This eludes to the presence of multiple filament arrangement's forming structures that result in a change or lack in periodicity such as a twisted rope or nanotube morphology, respectively (Jozef Adamcik et al., 2011).

The correlation presented by this method is representive of separate classes. To a certain extent and considering the large number of clusters compared to previously identified fibril morphologies for these peptides, this may represent the separation of the 'major' classes into different sub-classes (Morris et al., 2013). However due the nature of fibril morphology, the addition of a single extra filament defines a new morphology. In addition to this, a change in the periodicity brought on by a change in the arrangement of filaments can also define a new fibril morphological class (Jozef Adamcik et al., 2010). Within a data set with such large differences in X and Y values, such as when comparing a 6nm fibril to a 15nm fibril; it is difficult to determine definitive class boundaries when taking in to account the slight changes in height and periodicity.

Therefore, this method provides a quantitative way to identify regions of natural correlation and therefore classes, within a dataset of highly polymorphic fibrils.

Representing fibril class morphology

Due to the large range of structural profiles seen by fibrils formed from RVFNIM, HYFNIF and VIYKI (figure 19) it is likely that there are a range of morphologies determined by both the number and arrangement of filaments within a fibril. Using monomeric dimensions previously characterised for the 3 WALTZ peptides (Morris et al., 2013) we created a series of models that encompass a wide range of possible filament conformations. This series of models allowed the prediction and visualisation of potential morphologies seen within the fibril population regarding both the twisting archetype displayed and the number of filaments assigned to that structure. For this work, fibril morphology will be broken down into 2 factors: archetype i.e. the twisting type such as helical ribbon. Then filament arrangement, which includes the number of filaments per fibril.

Figures 27-28 shows the range of models constructed for RVFNIM, HYFNIF and VIYKI, each displaying a different twisting archetype and/or morphology in regard to filament number and arrangment. The models themselves were originally constructed in a step-wise manner in which each additional filament granted access to a wider range of filament arrangement's. The number of unique arrangements was dependent on the dimensions of the individual protofilaments; which are in turn defined by beta strand size. Using X-ray fibre diffraction, the individual beta strand size has been determined for each of the WALTZ peptides(Morris et al., 2013). In addition to this, the orientation and arrangement of the protofilament forming interaction of 2 beta sheets has been characterised. An individual cell unit is described here as the size of 2 interacting beta

strands. For HYFNIF the beta strands interact with a parallel, staggered non-steric zipper interface; resulting in a cell unit size of approximately 2 nm x 1.5 nm.

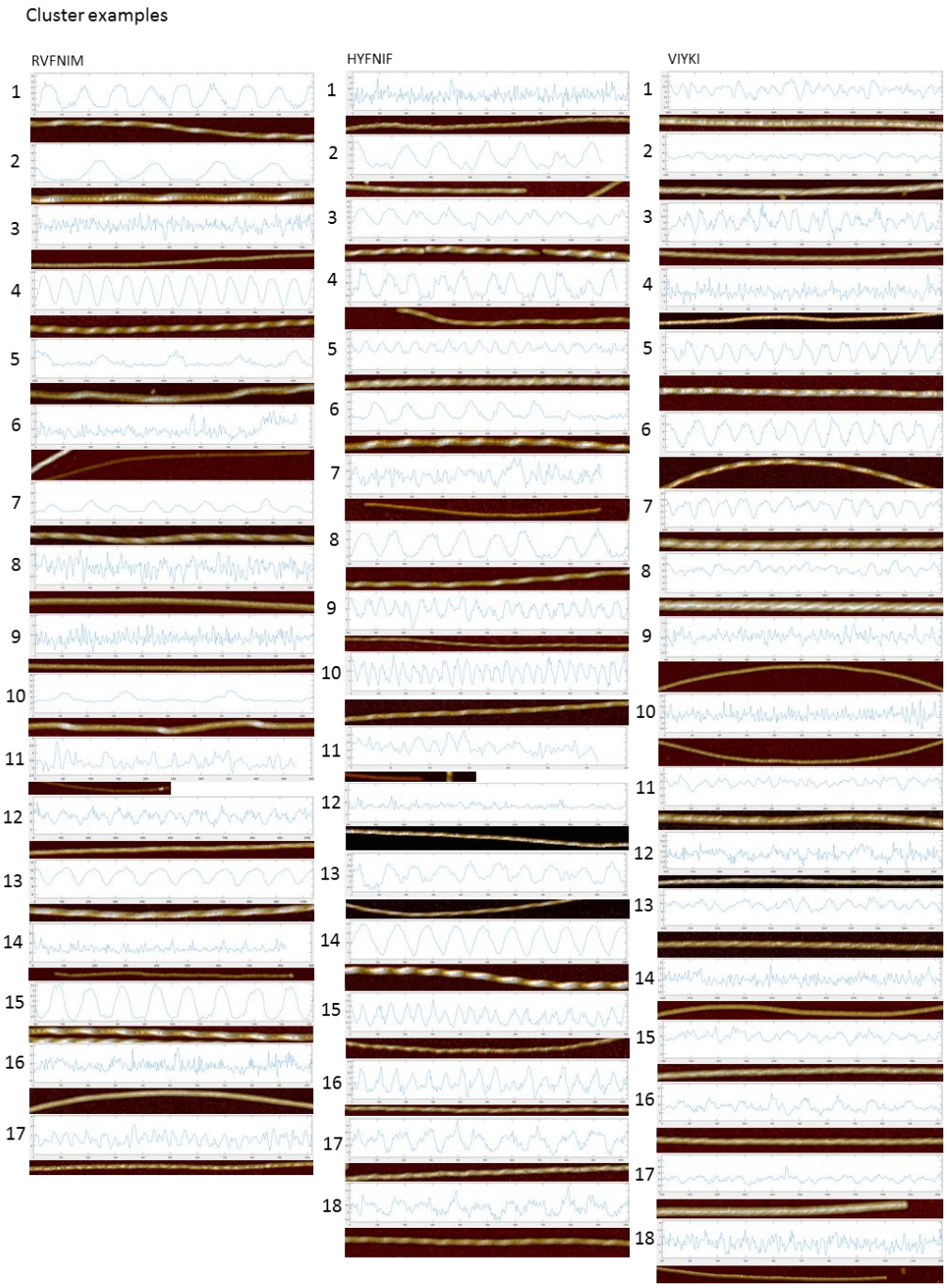


Figure 19. AFM images of morphologies detected by hierarchical clustering

Each of the clusters identified through the hierarchical analysis contain fibrils that present relative similarity. This figure displays AFM images of example fibrils for each of the clusters identified within the WALTZ peptide populations. Each image has a pixel density of 2.9nm/pixel and represents a 1µm stretch of the fibril particle. In addition, each fibril is accompanied by the corresponding height profile collected using an in house Matlab application. RVFNIM contains 17 example fibril morphologies corresponding to the maximum number of clusters to likely represent the population. HYFNIF contains 18 example morphologies, as does VIYKI.

RVFNIM and VIYKI beta strands interact via a steric zipper interface, whilst VIYKI displays an offset interaction in the direction of the beta strand in which the strands are parallel to each other and has a cell unit size of around 1.5nm x 2 nm. RVFNIM displays the only antiparallel interface and has a cell unit size of around 1nm x 2nm (Morris et al., 2013). Whilst the specific residue interactions are not defined, it is likely that this discrepancy between beta sheet interface between the 3 peptides, is sequence dependent. These values are therefore size approximations of respective protofilaments. The hierarchical structure of an amyloid fibril can be described as having two interacting protofilaments which are each formed from two or more interacting β-sheets. These protofilaments are the constituents of filaments; which are in turn the hierarchical unit that best describes the individual strands of a fibril (Fitzpatrick, Debelouchina, Dobson, et al., 2013). With AFM, derived height information in support of this method of hierarchical assembly, we constructed our models accordingly. During our data collection process, we did not observe any fibrillar species below the height of 3nm (figure 18) supporting the interaction of a pair of protofilaments to form an individual filament. Therefore, the minimum unit used to develop the models was a filament consisting of two laterally associated protofilaments. Each subsequent addition was of a filament, meaning the number of protofilaments per fibrils is double the number of filaments. The minimum number of filaments required to construct model morphologies that covered the range of heights seen in the data was 4; therefore, the initial model series was made with all possible arrangements of filaments up to a maximum of 4. These model morphologies were attributed with a twisted archetype (twisted ribbon or twisted rope). Experimental data was fed into the modelling process with subsequent models were added to the series based on observations within the respective fibril populations. These were models with more complex arrangements, such as nanotubes, helical structures, and

fibrils with greater than 4 filaments. The full series of models can be seen in figures 21 and 22. The averaged characteristics for each cluster of RVFNIM, HYFNIF and VIYKI can be seen in figure 25 which details the average height and periodic behaviour but also adds the regular minimum and maximum height along each fibril axis; again, averaged per cluster. The difference in maximum and minimum height provides valuable information eluding to the number of filaments per fibril. While the maximum height and periodicity alone can allow the determination of twisted ribbon morphologies; a twisted rope morphology with equal height and periodicity would be accompanied by a comparatively smaller difference in axial height and an even greater difference for a helical morphology.

Figure 20. Comparative height ratios for multistranded twisted ribbons

Example models representing a 2, 3 and 4 filaments twisted ribbon morphology. Each filament in this example is 4 x 4 nm. With each additional filament, the difference between maximum and minimum fibril height gradually increases. This relationship is implemented into characterisation process.

The ratio of maximum to minimum fibril height also changes with the addition of new filaments even within the same archetype of fibril. For example, the maximum height of a 3-filament ribbon is 1.5 times larger than the minimum height, whereas the maximum height of a 4-filament ribbon is 2 times larger than the minimum height if you consider a filament with equal X and Y dimensions. This suggests that a more complex

characterisation is required beyond single height values to accurately determine fibril morphologies present. The hierarchically clustered data, which was linked based on average height and periodicity is then used to fit the constructed model series based on maximum and minimum height. Therefore, adding an additional layer of class determination.

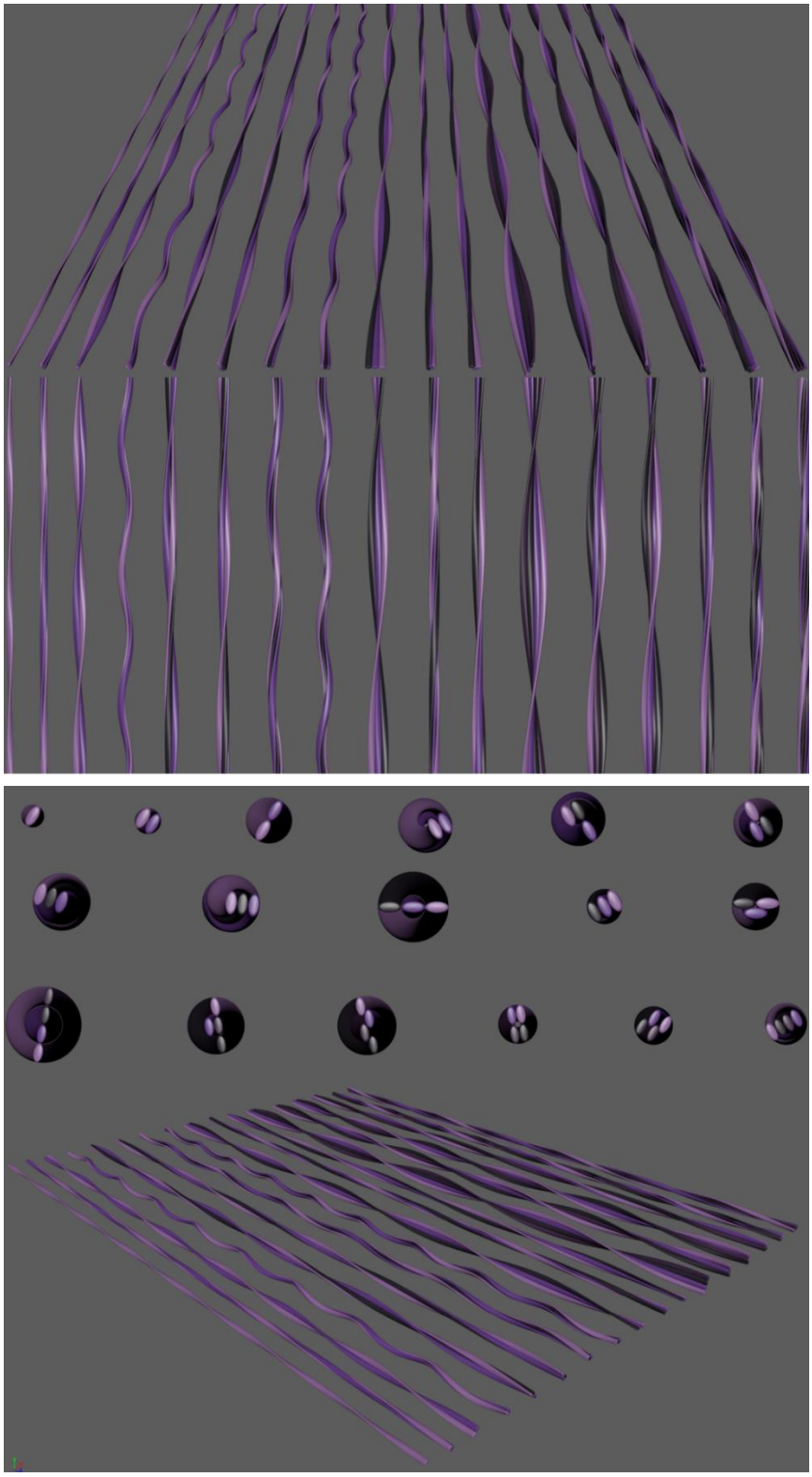


Figure 21. Complete model series created for RVFNIM fibrils

Multiple viewing angles of the original model set made to represent RVFNIM filament arrangement. Each model was made using MAYA (Autodesk) and represents the possible arrangement for the hierarchical assembly of additional filaments within a fibril.

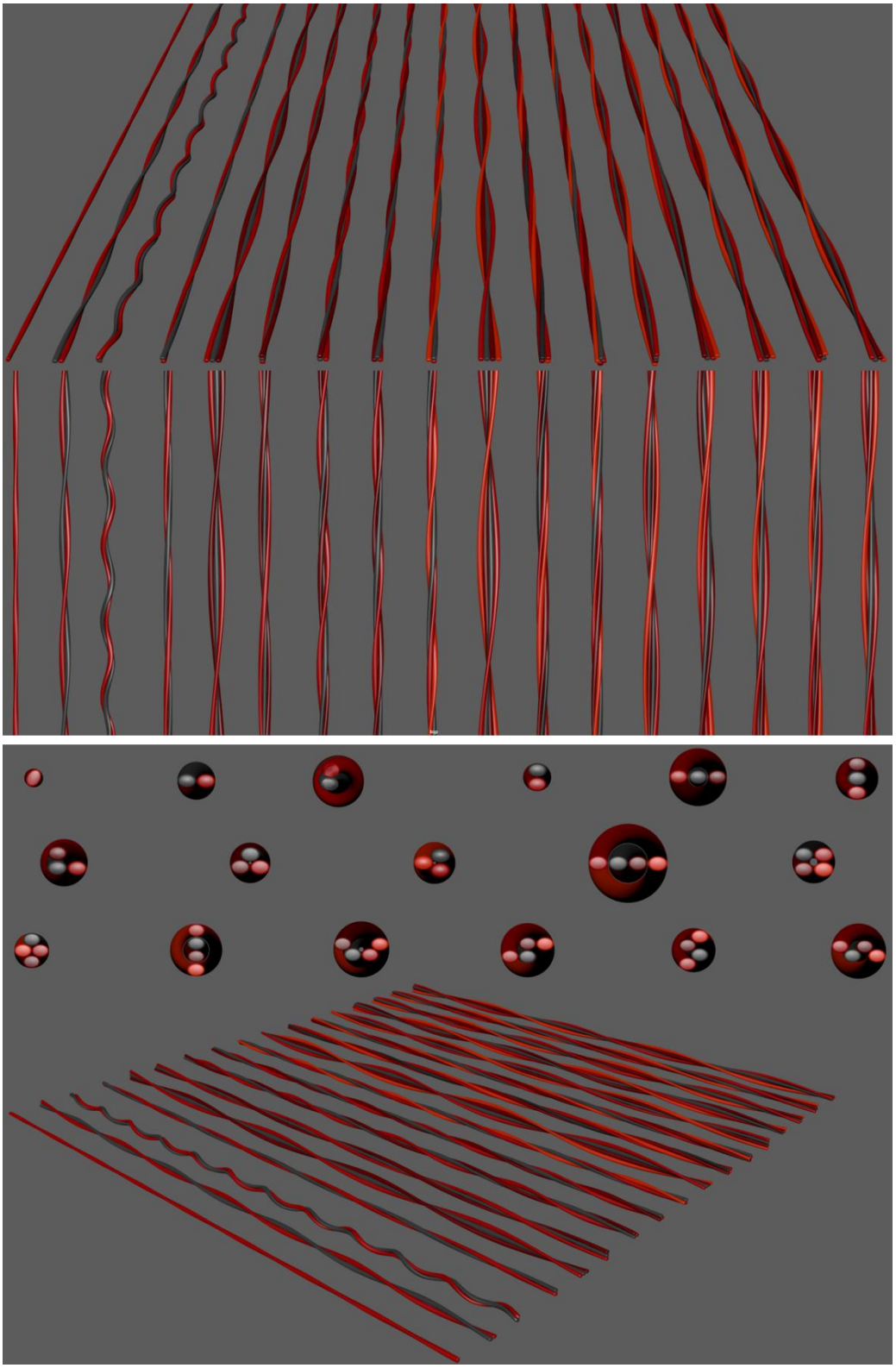


Figure 22. Complete model series created for VIYKI and HYFNIF fibrils

Multiple viewing angles of the original model set made to represent HYFNIF and VIYKI filament arrangement. Each model was made using MAYA (Autodesk) and represents the possible arrangement for the hierarchical assembly of additional filaments within a fibril.

	Fibril morphology	Max height (nm)	Min height (nm)
1	1 filament	4	3
2	2 filament ribbon	8	5
3	2 filament rope	4.5	4
4	2 filament helical ribbon	10	4.5
5	3 filament stacked rope	6	5
6	3 filament ribbon	12	7
7	3 filament L twisted rope	9.5	6.5
8	3 filament triangle rope	8	5.5
9	3 filament alternate triangle	8.5	6
10	3 filament helical rope (LH)	10	5.5
11	3 filament helical rope (RH)	10	5.5
12	4 filament twisted rope	8	6
13	4 filament twisted ribbon	16	9
14	4 filament lateral ribbon	8	5.5
15	4 filament 3 and 1	12	8.5
16	4 filament long offset ribbon	13	8
17	4 filament short offset ribbon	8	7
18	4 filament alternate rope	8	7

	Fibril morphology	Max height (nm)	Min Height (nm)
1	1 filament	4	3.5
2	2 filament ribbon	8	6
3	2 filament rope	6	5
4	2 filament helical ribbon	13.5	6
5	2 filament helical rope	10	6
6	3 filament lateral ribbon	9	6.5
7	3 filament ribbon	12	8
8	3 filament L twisted rope	10	7.5
9	3 filament triangle twisted rope	9	7
10	3 filament alternate triangle	9.5	7.5
11	4 filament twisted ribbon	16	10
12	4 filament lateral ribbon	12	8
13	4 filament twisted rope	9	8
14	4 filament U ribbon	14	10
15	4 filament staggered ribbon	13.5	9.5
16	4 filament zig zag ribbon	13	10
17	4 filament alternate stagger	11	9
18	4 filament diamond rope	8.5	8

Table 12

Tables in this figure detail the maximum and minimum heights displayed by the fibril models presented in figures 27 and 28. RVNIM models are on the left and VIYKI/HYFNIF models on the right.

Table 12 displays the dimensions of each constructed model for the 3 peptides (note the models for VIYKI and HYFNIF use the same sized filaments). Figure 23 shows the maximum and minimum heights of the fibril clusters plotted with the maximum and minimum heights of each constructed model. This allows the fit of the model data to the experimental data. As expected, each model does not represent a fibril class of its own, regardless there is a strong fit between the 2 sets of data suggesting that the models may accurately represent the polymorphism of each fibril population. The final column of figure 25 assigns each cluster the model(s) which most accurately represent them based on the dimensions of each construct in the series. The overlay suggests that while some models do not fit the data, other models may represent more than one cluster; suggesting a merging of clusters into the same model defined class. These relationships can be due to clusters displaying the same maximum and minimum height but different periodicity, which may suggest a flexibility in the periodicity adopted by a certain morphology. To support the correlation between models and experimental data, AFM images of example fibrils from each cluster and their respective height profiles

(figure 19) were used to verify the likelihood of a cluster being represented by a specific model. Some clusters are closely related to more than one model (linked in figure 23 with more than one loop) and visual inspection allowed the assignment of a fibril to its more likely model. Red loops indicate a model that upon visual inspection could potentially represent a cluster outside from its most closely related group. While some clusters overlap with models at high accuracy, some models can be linked to clusters if they lie within the variation of that cluster. Which is an ever-present issue with classification - the determination of definitive class boundaries given the evident intra-morphological variation.

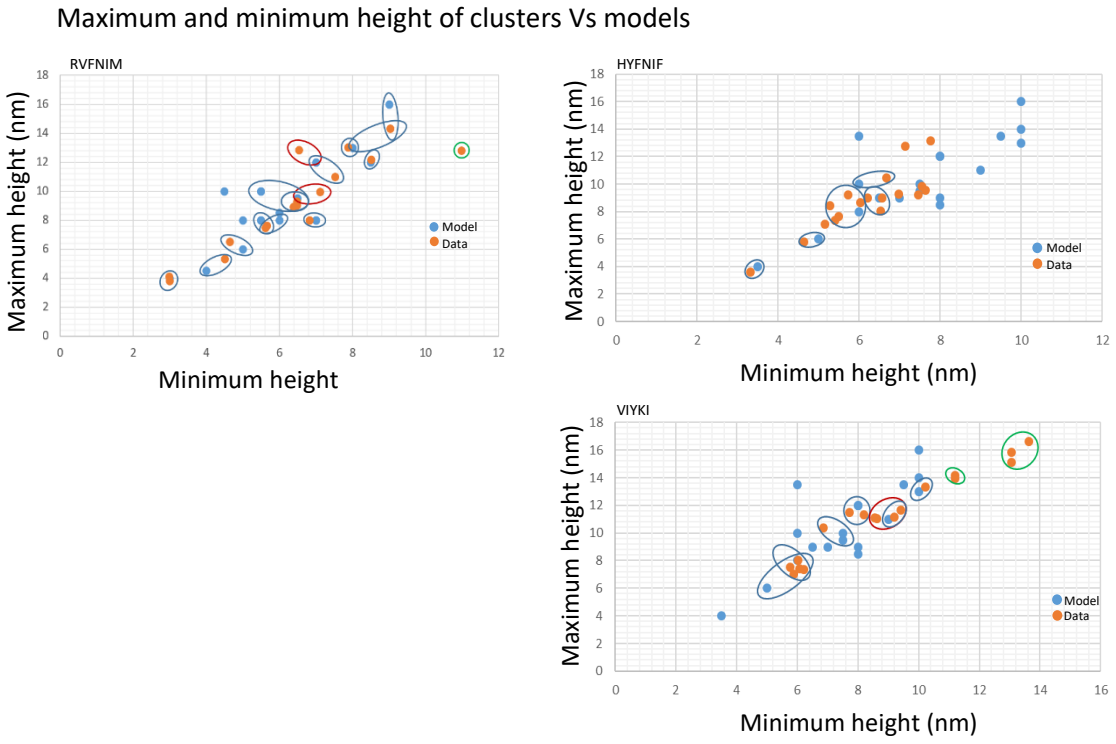


Figure 23. Fitting fibril models to the data

To identify the possible filament arrangements observed in the experimental data, the AFM and Matlab derived fibril data was plotted with the equivalent dimensions of the graphical representations. Scatter graphs with both sets of data points allow for correlation between the model dimensions and the experimental dimensions. Clusters in orange were plotted with models in blue; correlating fibrils and models are circled in blue to identify possible model representation. Red circles identify models that may represent more than one cluster. Additionally, multiple clusters may be represented by the same model as the number of clusters represent the maximum number of possible fibril classes. Green circles identify fibril clusters that do not correlate with the original set of model structures; therefore, requiring additional model development to present possible filament arrangement.

Finally, certain classes that do not fit the originally created series of models are circled in green on Figure 23. Such classes display heights that are more than the general population and may be sequentially developed morphologies that require an increased incubation time to appear (Usov et al., 2013). Regardless, figure 26 and 27 displays models to represent the likely morphologies adopted by the outliers seen for RVFNIM and VIYKI, respectively. RVFNIM displays a single outlying cluster, which is emphasised due to the similarity in minimum and maximum height and lack of periodicity. Upon further inspection of the AFM image (figure 26) a change in morphology along the fibril axis can be observed.

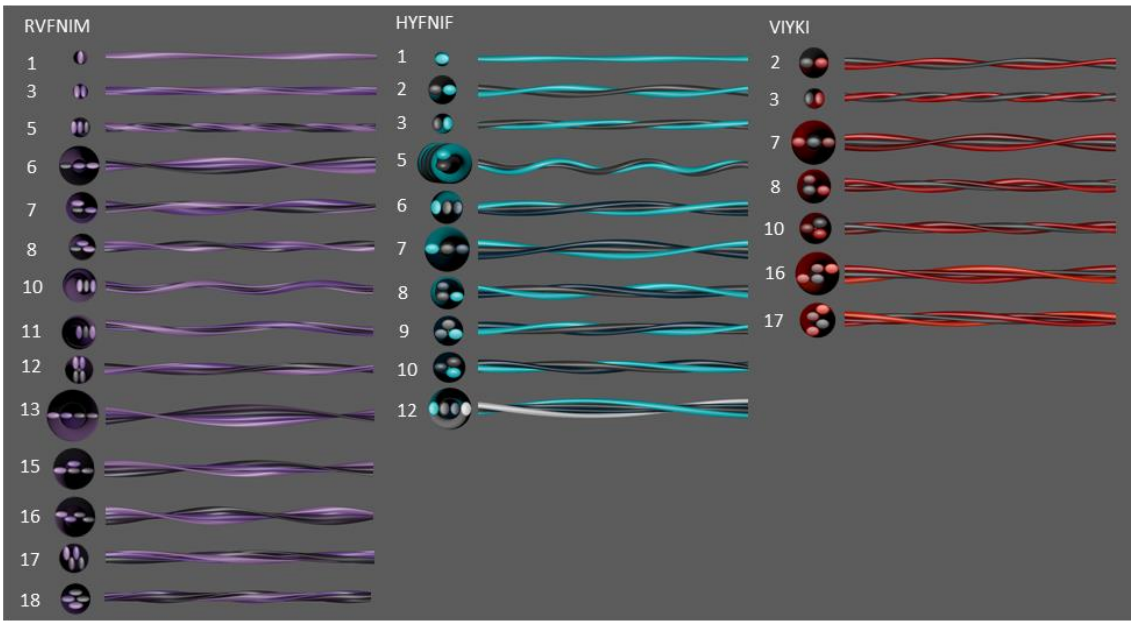


Figure 24. The range of model morphologies detected in fibril populations

From all models created to represent the data, those which are identified within the experimental data are depicted here for each respective fibril population. Through structural matching, the AFM collected data in conjunction with the graphical representations, suggest 14 possible morphologies observed within the RVFNIM fibril population. HYFNIF presents a possible 10 and VIYKI 7.

VIYKI

Cluster number	Frequency (nm)	Periodicity (nm)	Height (nm)	Max Height (nm)	Min Height (nm)	Model number
1	0.126	68.1	10.5	11.6	9.4	17
2	0.006	1426.9	12.5	13.9	11.2	19*
3	0.206	41.8	10.1	11.1	9.2	17
4	0.072	119.0	6.7	7.3	6.2	3
5	0.119	72.7	8.6	10.3	6.9	8 or 10
6	0.007	3169.6	9.7	11.4	7.7	7
7	0.093	93.6	12.8	14.1	11.2	19*
8	0.179	48.0	15.1	16.6	13.6	20*
9	0.138	62.8	6.5	7.5	5.8	3
10	0.499	17.2	6.7	7.4	6.1	3
11	0.131	67.5	11.8	13.3	10.2	16
12	0.022	748.3	14.0	15.1	13.0	21*
13	0.167	51.7	9.8	11.0	8.6	17
14	0.010	1954.1	6.7	8.0	6.0	2
15	0.110	79.7	9.7	11.0	8.5	17
16	0.053	164.0	10.0	11.3	8.2	7
17	0.106	85.3	14.3	15.8	13.1	21*
18	0.320	26.8	6.4	7.0	5.9	3

HYFNIF

Cluster number	Frequency (nm)	Periodicity (nm)	Height (nm)	Max Height (nm)	Min Height (nm)	Model number
1	0.006	2829.7	7.9	9.3	6.9	9
2	0.089	97.0	7.1	8.6	6.0	2
3	0.083	103.8	9.8	12.8	7.1	7 or 12
4	0.128	67.4	6.1	7.1	5.1	3
5	0.148	58.2	8.7	9.8	7.5	10
6	0.060	143.5	8.2	10.4	6.6	5
7	0.063	136.7	5.2	5.8	4.6	3
8	0.058	153.0	7.2	9.2	5.7	2
9	0.004	3064.6	6.8	8.4	5.2	2
10	0.205	42.0	7.3	8.0	6.5	2
11	0.136	63.3	3.5	3.6	3.3	1
12	0.034	256.3	8.0	9.2	7.4	8
13	0.093	93.2	7.6	9.0	6.5	6
14	0.099	87.6	10.4	13.2	7.7	7 or 12
15	0.159	54.4	6.3	7.3	5.4	2
16	0.138	62.5	7.8	9.0	6.2	6
17	0.102	85.5	8.7	9.6	7.6	10
18	0.086	101.6	6.5	7.7	5.4	2

RVFNIM

Cluster number	Frequency (nm)	Periodicity (nm)	Height (nm)	Max Height (nm)	Min Height (nm)	Model number
1	0.079	109.2	9.0	10.9	7.5	6
2	0.052	172.7	10.1	13.0	7.9	16
3	0.006	2494.9	5.6	6.5	4.6	5
4	0.111	77.5	7.7	9.0	6.5	10
5	0.003	2798.3	9.5	12.8	6.5	6
6	0.027	318.4	3.4	4.1	3.0	1
7	0.074	116.3	11.7	14.3	9.0	13
8	0.136	64.0	6.7	7.6	5.7	12
9	0.375	22.9	7.3	8.0	6.8	17 or 18
10	0.012	1071.0	7.4	8.9	6.4	7
11	0.452	19.0	3.4	3.8	3.0	1
12	0.078	116.2	8.5	9.9	7.1	7
13	0.079	109.5	10.4	12.1	8.5	15
14	0.039	243.3	4.9	5.3	4.5	3
15	0.073	118.5	7.6	9.2	6.5	11
16	0.004	2039.2	11.6	12.7	11.0	19*
17	0.247	34.9	6.6	7.47	5.6	8

Figure 25. Average fibril properties of identified clusters

The cluster analysis of the respective WALTZ peptide datasets, has each fibril assigned to a group. Each 'group' or 'cluster' contains multiple fibril particles from the AFM collected images; The tables in this figure present the average values for each cluster with each characteristic analysed using the purpose developed Matlab script. In addition to dimensional information, the table details the representative model for each cluster.

Using height measurements and the approximate size of an individual filament, the sequential addition of an additional filament and the subsequent transition into a nanotube like structure can be observed. This is a relatively uncommon structure due to its infrequency, and absence from the population dataset for HYFNIF however this does not mean that is an inaccessible morphology. VIYKI also displays morphologies that are larger than the limit of a 4-filament twisted ribbon. A further set of models constructed to represent VIYKI class outliers, is shown in figure 27. These models represent large helical ribbons, made up of 5 or 6 filaments, and a nanotube closure of the 5-filament helical ribbon.

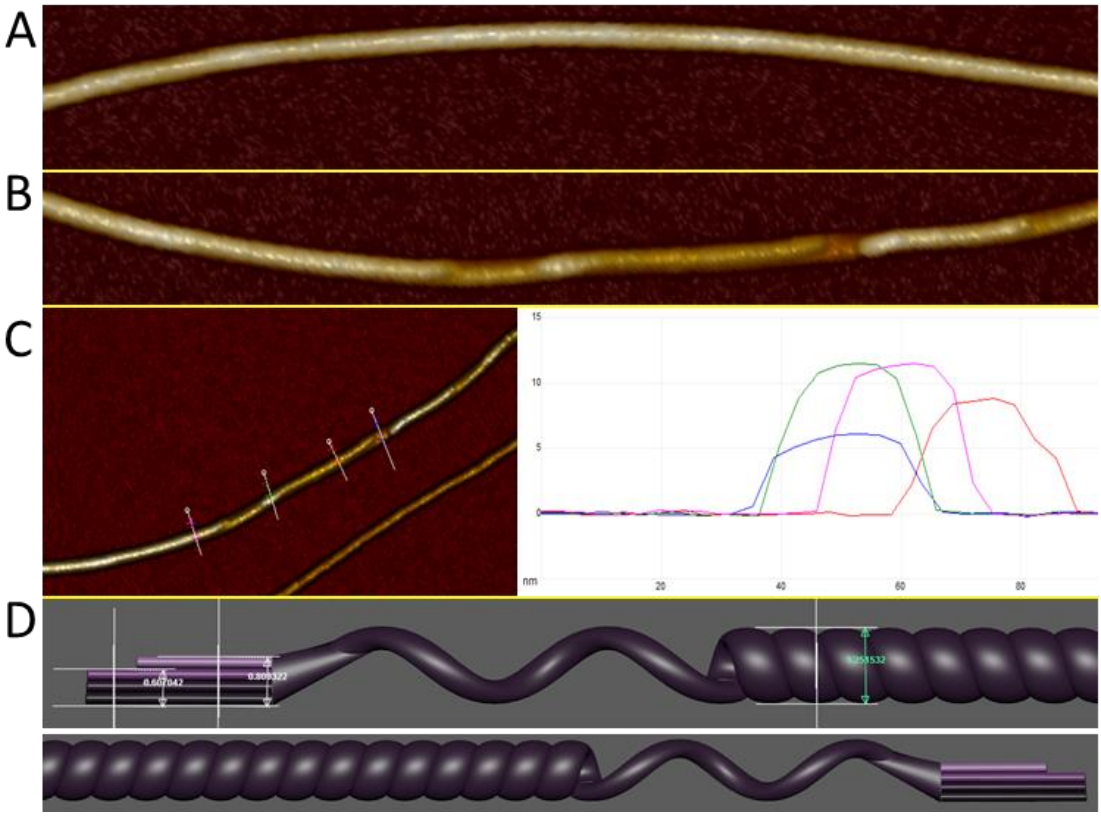


Figure 26. An observed change in morphology, into nanotube-like structures

Some fibril clusters have no match within the original set of models, therefore new models were developed per dimensional observations within the AFM images of fibrils belonging to respective classes. RVFNIM displays a fibril class which is ~12nm in height with minimal height fluctuation and undetectable periodicity; suggestive of nanotube like structures (A, 1µm stretch of the fibril particle). Some examples of such fibrils display a changing morphology across the polymer, observed through considerable and continuous changes in height (B). Height measurements were taken using Nanoscope analysis (Bruker); the changes in morphology are suggestive of additional filament association, given the height of an individual RVFNIM filament is ~2nm and the height changes from 6nm to 8nm. Following this, the increase in height to 12nm is indicative of the adoption of a helical twist which in addition to the lack in periodicity is likely nanotube-like (C). Finally, graphical representations created in Maya suggest the possible filament association and change in morphology with a final nanotube like structure that matches the height measurements seen in the AFM image (D). This represents model 19 in figure 25, which corresponds to cluster 16.*

This is the most likely twisting archetype and morphology of clusters with such a height profile. VIYKI clusters 2 and 7 are represented by model 19* (figure 27A) with max and min heights of 14nm and 11nm, VIYKI cluster 8 is represented by model 20* (figure 27B) with heights of ~16nm and ~13nm, respectively. Finally, VIYKI clusters 12 and 17 have similar max and minimum heights as cluster 8, however the average height is lower; on observation of their height profile (figure 19) the regular height fluctuations are minimal with a clear lack in periodicity. This suggests a flat and regular height along the fibril axis

that is likely representative of a nanotube like structure; that may have formed following the helical closure of the structures previously mentioned. It is a predetermined thought that helical ribbons such as these represent an intermediate stage prior to helical closure into a nanotube structure (Jozef Adamcik et al., 2011). It seems that this may be a common structure adopted by fibrils formed by VIYKI.

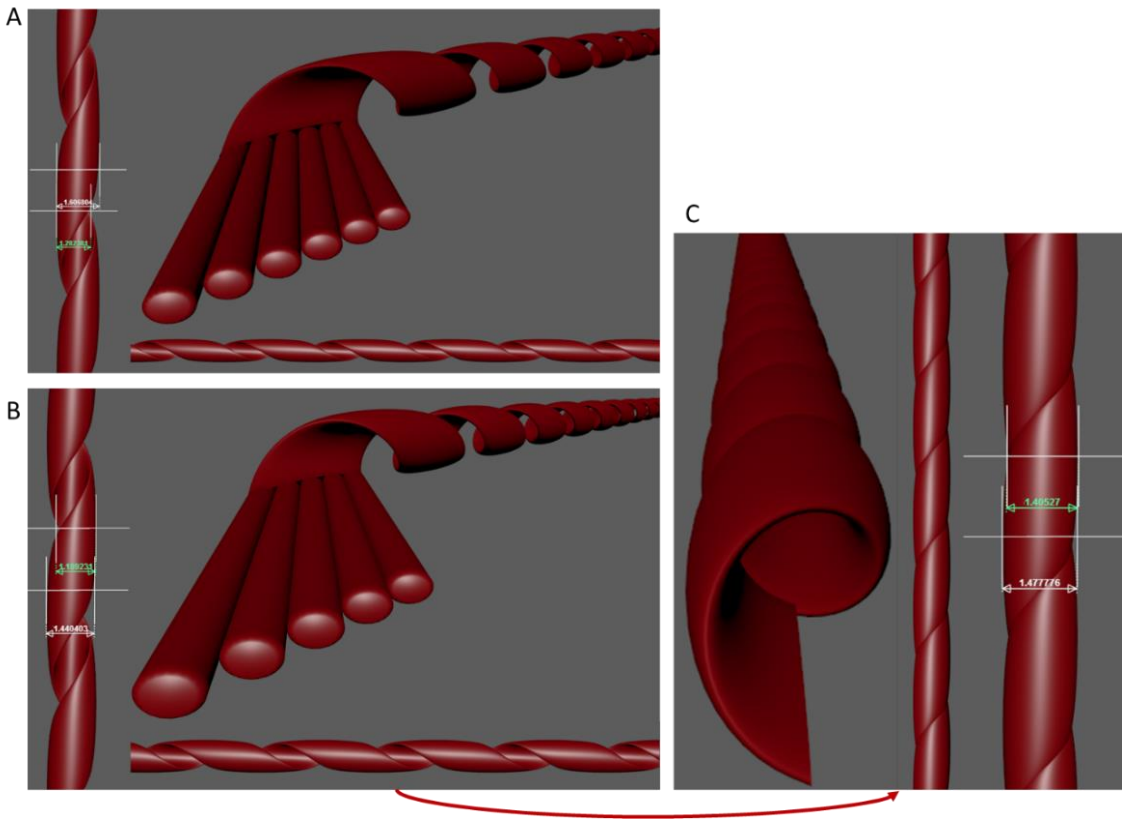


Figure 27. Additional morphologies adopted by VIYKI fibrils

VIYKI displays 3 clusters which do not correlate to the original model sets. (A) represents model 20 for VIYKI in figure 28 – cluster 8, and proposes a helical ribbon morphology with a maximum height of ~16nm and a minimum height of ~12nm. The model seen in (B) is number 19 which represents cluster 7 in the VIYKI dataset, presenting another helical ribbon morphology. However, this model proposes a 4-filament structure with a maximum height of ~14nm and a minimum height of ~11nm. Finally model 21 seen in (C) represents a possible closure transition from a helical ribbon morphology into a nanotube like structure. This morphology is likely adopted by clusters 12 and 17 in the VIYKI dataset, and has a maximum and minimum height of 14nm with little to no fluctuation along the profile.

It is important to note that we consider there is a discrepancy between helical ribbons and helical ropes; model's 6 and 11 for RVFNIM and the respective clusters 5 and 15 depict such a structure. Given the dimensional difference between laterally interacting filaments and those interacting 'head to tail' (Fitzpatrick, Debelouchina, Bayro, et al., 2013) the resulting fibril is more indicative of either a twisted rope or a twisted ribbon morphology. The Helical ropes observed for RVFNIM display a reduced helical pitch compared to the helical ribbons seen with VIYKI. It is likely that this helical rope is the precursor structure which upon addition of another filament, transitions into the nanotube structure seen in figure 27C. Suggesting a dynamic progression of morphologies into new ones. The closure of helical ribbons into a nanotube structure is thought to be an electrostatic charge mechanism, which is dependent on the charge of the specific residues of the protein or peptide (Castelletto et al., 2011). The Arginine in RVFNIM and the Lysine in VIYKI result in a net charge of 1 meaning that given enough time and in the absence of high salt concentration fibrils formed from these peptides can facilitate helical closure into nanotubes. The lack of helical structures beyond a 2-filament helical rope (model 5) for HYFNIF may be indicative of the lack of a morphological progression driven by net sequence charge; given that HYFNIF has is neutral.

The orientation in which filaments may attach is a variation that must be considered when predicting fibril morphologies. It raises the question whether a lateral side chain interaction between filaments is more favourable than an end to end interaction between terminal residues of separate beta strands within a filament. Both these possibilities have been considered in the construction of the models, and the data when matched to such suggests both these are potential filament arrangements. For example, model 6 and model 7 for HYFNIF represent 3 filament fibrils which are interacting via

both interfaces. Figure 19 shows example fibrils and height profiles corresponding to these cluster-matched models. These models display a maximum and minimum height of 9 and 6.5nm then 12 and 8nm, respectively. Therefore, these two 3-filament twisted ribbons, display different height profiles, and therefore can be attributed to separate morphologies. An example in which the intra-cluster variation presents classification issues is again using model 7 as an example with model 12 for VIYKI. Model 7 represents a 3-filament twisted ribbon orientated laterally, as the beta strand axis is perpendicular to the long dimension of the filament (the opposite is the case for HYFNIF ~4x3nm as opposed to 3x4 for VIYKI). Model 12 represents a 4-filament ribbon in which the filaments are interacting in a head to tail fashion. This means that both fibrils display a height profile with a maximum height of 12nm and a minimum height of 8nm. This means that clusters within this range of height profile can potentially be represented by either model. Figure 32 displays all the models that have been attributed to a cluster, showing that there is a lower number of represented models than there are clusters. The average fibril dimensions per cluster are tabled in figure 25 in addition to the matched model structure. This tightens the classification, by culminating predetermined clusters into the same class. The resolution of this become apparent when the previously linked dataset, was linked to the models based on additional height profile information alone; separate from the periodic information. Thus, fibril morphology has a high level of homogeneity regarding the height profile, however is more heterogenous within a class in regard to its periodicity. This could be attributed to transient changes in twisting that are indicative of fibrils at different stages of their elongation and/or association with additional filaments.

Figure 32 and the additional models represented in figures 32-33, suggest that RVFNIM fibrils may adopt 15 different morphologies including a single filament structure. HYFNIF may be able to adopt at least 10 morphologies including a single filament structure, and VIYKI may be able to adopt 11 morphologies including a single filament structure (not observed in the dataset). While the results suggest a wide range of accessible morphologies, it also shows that certain structures are more likely and therefore more favourable for the given peptide sequence. Figure 28 shows the number of fibrils that are assigned to a cluster, with each cluster assigned a morphology. Around ~30% of RVFNIM fibrils are assigned to cluster 13, which is assigned a 4-filament morphology (model 15) that is similar in structure to model 6. Both present a 3-filament lateral assembly however cluster 13 displays a minimum height indicative of an additional central filament. A maximum and minimum height of 12nm and 8nm requires at least a central filament of 4 nm x 4 nm. A possible alteration to this model would be a 6-filament fibril in which 2 layers of 3 filaments are laterally associated. Such a structure however, would require 12 protofilaments to interact given the determined size of an RVFNIM protofibril (Morris et al., 2013). Regardless, the most common adopted archetype is that of a twisted ribbon with variation in the overall morphology which requires a minimum of 3 filaments across. In addition to this RVFNIM display the prevalence of both left and right-handed twists, in the absence of any chiral inverting conditions (Kurouski et al., 2013). HYFNIF on the other hand presents a lower average height, with its most common morphologies represented by models 2 and 3. These models are both 2 filament twists, with a difference in the orientation of interacting filaments. The reduced favour for larger fibril structures could be attributed to its lack of charge driven morphological advancement. The next most common structure is again a 2-filament helical rope, suggesting that HYFNIF favours the adoption of 2-filament

morphologies. Finally, VIYKI cluster 5 contains approximately 19% of the overall fibril population. This cluster is represented by a 3-filament rope (model 8 or 10); however, it is unclear due to the similarity which filament arrangement is more likely. The preference for a 3-filament twisted rope differentiates the polymorphism between the 3 WALTZ peptides, further supporting a sequence dependent driving force of polymorphism amongst amyloid species.

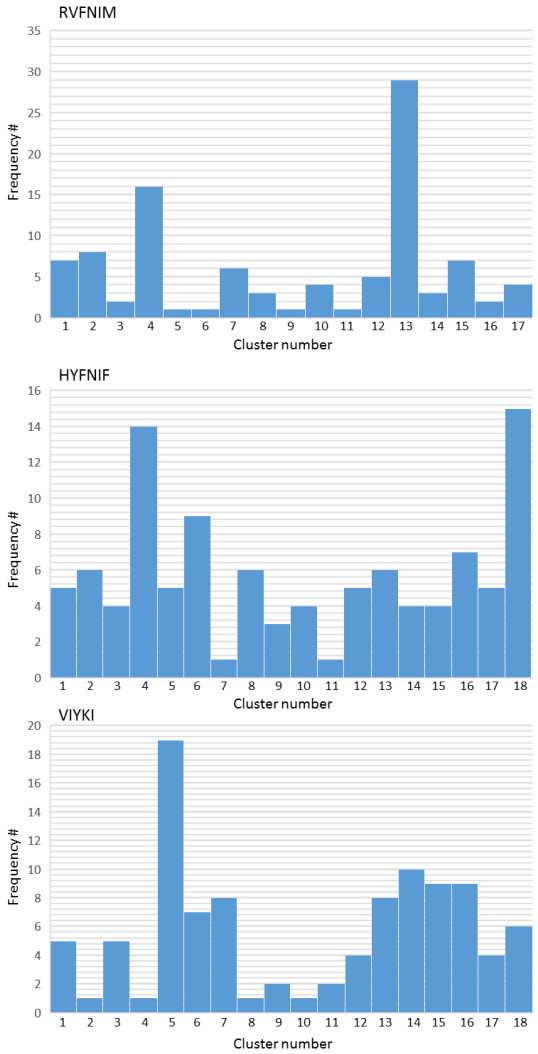


Figure 28. Frequency of morpholgy

While the cluster analysis identifies multiple fibril classes within the input data set, not all classes are represented equally within the fibril populations. The histograms in this figure quantify the respective numbers of fibrils per cluster for each of the WALTZ peptide datasets. Each of the fibril populations suggest dominant morphologies with RVNIM and HYFNIF displaying 2 relatively dominant clusters, and VIYKI 1 dominant cluster. Other clusters either possess similar distributions of fibrils or are uncommon with only 1 or 2 example fibrils within the dataset. The dominant clusters for RVFNIM are 4 and 13 which correspond to models 10 and 15 respectively. For HYFNIF, the dominant clusters are 4 and 18 which correspond to models 3 and 2 respectively; both are alternative arrangements of a 2-filament fibril. Finally, cluster 5 is dominant for VIYKI which corresponds to both model 8 and 10; 3-filament twisted rope structures with a slightly altered arrangement.

The models identified in Figure 32 represent those which match experimental data, so are likely adopted within the population. However, in the case of models 8, 9 and 10 (represent clusters 5, 1 and 12 respectively) for HYFNIF; all are very similar morphologies, making differentiation within the data difficult. In addition, the cluster

tree shows high similarity for clusters 1 and 12 supporting the similarity between the two models. However, cluster 5 shows a high level of dissimilarity to the other two clusters (figure 18 C). The separation on the cluster tree is likely due to periodicity difference which may be due to additional variation in the filament arrangement, causing different twisting patterns. an inaccurate periodicity value is usually due to an irregularity of the periodicity when there are fluctuations however they are very erratic. Such height profiles may be indicative of rope like morphologies in which the changes in height are minimal, and in fact very like the likely height profile of a nanotube like structure (height changes of ~1nm, difficult to detect and the fluctuations that would be seen in some rope like structures). Alternatively, each of the three models (8,9 and 10) are simply possible structures adopted by each of the clusters, as opposed to a definitive orientation and arrangement of filaments. The observations shown in figure 26 display morphological switching, due to the change in morphology along the fibril axis. Not only is another filament associating with the fibril, but the twisting pattern changes. This supports the idea of morphological switching from the original nucleating particle, and a dynamic mechanism for fibril growth. This is an idea that has been reported with $A\beta_{1-42}$ fibrils, in which monomer seeded growth resulted in a change in fibril morphology to a 'hybrid' one (Watanabe-Nakayama et al., 2016). The high degree of variation in fibril height profiles reported here, further supports the idea of morphological flexibility as opposed to rigid size constraints for characterised structures. Morphology switching, either by monomer addition (Watanabe-Nakayama et al., 2016), or additional fibril-fibril interactions; it is possible that all the models presented here are accessible to the respective fibril populations. However, some are more common than others.

To our knowledge there has not been a quantitative analysis previously performed that details the range of morphologies present in a fibril population regarding the number and arrangement of filaments for each type of structural archetype (twisted ribbon, twisted rope etc.). Previous studies have determined polymorphism present in regards to archetype alone, however (Jozef Adamcik et al., 2011; Annamalai et al., 2016; Usov et al., 2013). For specific twisting archetypes, such as the twisted ribbon; extensive morphological characterisation has been conducted due to its common occurrence as a structure adopted by amyloid fibrils, with a clear relationship between filament number, size and periodicity (Jozef Adamcik et al., 2010; Fitzpatrick, Debelouchina, Bayro, et al., 2013; Meinhardt et al., 2009). This work expands on the currently understood morphology landscape of amyloid fibrils; showing that there is a more dynamic relationship between fibril size and morphology. I.e. archetypes such as helical ribbons and twisted ropes vary in morphology in a filament dependent manner akin to twisted ribbons. This morphological modelling defines the different twisting archetypes adopted by fibrils formed from WALTZ peptides. A heterogenous filament arrangement is present throughout each twisting archetype, defining separate sub classes. Therefore, this work characterises the morphological variation present for all known twisting archetypes, in terms of the both number and arrangement of filaments.

Conclusion

The fibril morphological landscape is likely a dynamic and progressive one and not static. Meaning that the fibril morphologies observed in a period are not indicative of the entire range of structures accessible to an amyloidgenic protein or peptide. As fibril elongation occurs, it is possible that fibrils are formed from oligomers in a multi filament arrangement resulting in twisting as the fibril elongates (Barnhart & Chapman, 2006). They may also elongate as single protofilaments, which hierarchically associate into

mature fibrils which over time recruit additional filaments resulting in a transition to a new morphology (Kad et al., 2003). It is likely that both occur simultaneously, however. Additionally, in a biological environment polymorphism does exist amongst clinically relevant amyloid showing that amyloid fibrils maintain their heterogeneity *in vivo* by analysing the polymorphism displayed by fibrils extracted from a patient suffering from light chain amyloidosis (Annamalai 2016). In addition, It has been shown that Different Alzheimer's patients present variation in the polymorphism of individual $A\beta_{1-40}$ fibrils, and there is variation in the structure and distribution of amyloid material within plaques (Liu et al., 2016; Lu et al., 2013). The manipulation of morphology has also been demonstrated by altering the fibril forming conditions, and the conditions of mature fibrils (Usov et al., 2013). Therefore, disease associated fibrils may have access to a wide range of morphologies such as the ones presented here through a condition dependent mechanism. However, the size of the morphology landscape may vary for larger protein-based amyloids as opposed to peptide-based ones, as presented here. In that the number of possible morphologies may be reduced due to a limitation in the number of thermodynamically favourable orientations of interacting filaments or even protofilaments. While the types of core interactions amongst all amyloid species are shared, such as steric zipper based hydrophobic or van der Waals interactions, aromatic and polar stacking or stronger hydrogen bonding(J. Park, Kahng, & Hwang, 2009b). The specific selection of these interactions is dependent on sequence, and the larger the monomer unit sequence the longer the repeating array of these interactions (J. Park, Kahng, & Hwang, 2009b). This potentially makes the balance between attractive and repulsive forces within the fibril structure more rigid due to the added complexity of a larger interacting sequence(Jozef Adamcik & Mezzenga, 2012). Meaning that to stabilise structures formed from these larger repeating monomer units with a likely higher

degree of internal repulsion, there is a narrower range of morphologies which are thermodynamically favourable. On the opposite side to this, peptide based amyloid structures such as those presented here, have a smaller core interface, meaning that the overall level of repulsion at the monomer unit level is likely much lower. This may put a weaker dependence on the specific type twisting pattern required to achieve thermodynamic stability within the fibril structure. Multiple structures will have similar Gibbs free energy, meaning the morphology landscapes global minimum will consist of more morphologies with similar energy conformations, than a landscape of fibrils formed from larger more complex proteins. This is a potential reason for reduced observed polymorphism in clinically relevant amyloid proteins.

It is important to understand the structural morphology of amyloid fibrils, as different filament arrangements may grant variation in the residue side chain profile along a fibril axis and at the fibril ends; which may be involved in cytotoxicity in a sequence dependent manner (Karen E Marshall, Marchante, Xue, & Serpell, 2014). This concept, is further complicated based on a sequence dependent polymorphism such as that presented here. Raising the question, is sequence related cytotoxicity dependent on the residue specific interactions, or the number of possible cell membrane interactions for a given morphology with varying rigidity? Additionally, the number of exposed residue side chains around a fibril circumference will vary depending on the number of filaments associated. It may also be likely that cytotoxicity is a size dependent mechanism, that is affected by particle number (W. F. Xue, Hellewell, et al., 2009). Different amyloid proteins may have a propensity to form fibrils displaying structures at different points along the hierarchical tree, with some more likely than others to form larger fibril structures such as nanotubes which incorporate multiple fibrils into a single polymer. A nanotube is the most likely structure adopted by a fibril of the respective size and lack

of detectable periodicity, however we cannot confirm there is a hollow centre. For two fibril populations with the same concentration of monomer precursor; the population that achieves a higher frequency of nanotube or other large structures, will display a smaller population of individual fibrils. Suggesting that there will be a reduced number of available fibril ends which may interact with cell membranes. Further to this, if cytotoxicity was indeed facilitated through a lateral association (Monsellier, Bousset, & Melki, 2016); the mechanical properties of the fibrils with enhanced size would affect the ability of the fibrils to effectively buckle and interact with a cell membrane due to changes in rigidity. In both instances, the simple organization of larger fibrils interacting with cell membranes would result in an increased difficulty of interaction and lower number of successful fibrils. Therefore, it is likely that the morphology of fibrils plays an important role in the level of cytotoxicity displayed by a fibril population and understanding quantitatively, the varying levels of polymorphism present may elude to an enhanced understanding of a sequence – structure relationship towards disease progression.

Chapter 4

The mechanical stability of amyloid fibrils formed from WALTZ peptides

Introduction

Amyloid can be attributed to the pathophysiology of various diseases; including both intra and extra - cranial diseases. Alzheimer's disease is a key example of a neurological disorder that involves the aggregation and accumulation of amyloid, resulting in characteristic and distinct cognitive impairment and eventually death. This disease involves the cleaving of amyloid precursor protein (APP) and subsequent mis-folding of the resulting Aβ 40-42 peptide, which elongates into mature amyloid fibrils. Parkinson's disease is an additional neurological disorder and its pathology is defined by the presence of alpha-synuclein deposits. While these two diseases display different symptoms, the presence of amyloid deposits within the brain is a hallmark feature and the method of cytotoxicity is a key area of interest. All proteins are thought to have the potential to form amyloid, given the ideal conditions, and all amyloid has a shared core cross β structure.

The 'life cycle' of amyloid yields the production of various intermediate assemblies during the elongation of fibrils; such as oligomeric prefibrillar structures of different sizes as well as the monomer units of the aggregated protein itself. Fragmentation of fibrils via mechanical perturbation or enzymatic fragmentation such as Hsp104 chaperone mediated fragmentation (Tyedmers et al., 2010), results in fibrillar material of reduced length. Together with the large deposits of accumulated fibrils into senile plaques or Neural fibrillary tangles (NFT) The amyloid “life-cycle” is a distinctly complex and heterogeneous process, making it difficult to identify the species involved in cytotoxicity. Conversely, some amyloid is not linked to disease and is deemed “functional” amyloid. For example, Curli fibers of *E. coli* and other bacteria produced amyloid has been linked to cell signalling via the retention of quorum signalling molecules within biofilms (Seviour et al., 2015).

The attributing characteristics and how amyloid conveys its cytotoxic potential is unclear, however various studies have shown that it is a sequence – structure relationship (Campioni et al., 2010; Karen E Marshall et al., 2014) which has implications on both the intermediate elongation species that form and the fragmentation or dissociation of fibrils, producing toxic species (W. Xue et al. 2010). How fibrils interact with biological membranes may be a mechanism shared amongst all amyloid (Williams & Serpell, 2011) and not necessarily due to the specific residues of a given amyloidgenic protein. The specific types of aggregates that form during elongation such as prefibrillar oligomers, are common structures accessible to all amyloid and these are linked to cytotoxicity (Bucciantini et al., 2004a). This has been previously suggested by showing that the elongation of Amylin fibrils results in Liposome disruption, likely due to the prefibrillar intermediates (Engel et al., 2008).

Fibril ends have been proposed as the active sites for membrane interaction (W. F. Xue, Hellewell, et al., 2009) and this may be due to the exposure of hydrophobic regions which act as “sticky” sites which additionally allow fibril – fibril interactions and the formation of clumps which in turn can act as sources of new oligomeric fragments (Williams & Serpell, 2011). This is a potential link between sequence and cytotoxicity, as proteins or peptides with a high number of hydrophobic residues may display an increased affinity for membrane interaction. In addition to this, fragmentation of fibrils formed from human β2-micro-globulin results in fibrils with enhanced cytotoxic properties, attributed to the reduction in length (W. F. Xue, Hellewell, et al., 2009). Together this represents a size dependent mechanism for cytotoxicity in both the forward and reverse directions of fibril growth, in which the sequence – structure relationship can influence the stability and therefore longevity of toxic prefibrillar oligomers as well as the mechanical properties of fibrils that may affect the

fragmentation or dissociation of fibrils. Given the correlation between reduction in length and cytotoxicity, the sequence and structure related mechanical properties of fibrils and how they affect the stability towards fragmentation is an important area of research.

HYFNIF VIYKI and RVFNIM are amyloidgenic short peptide sequences, identified via the WALTZ algorithm; a residue position scoring system with the ability to isolate amyloidgenic sequences within larger proteins (Maurer-Stroh et al., 2010). HYFNIF, VIYKI and RVNIM are sequences found within the Human Bloom syndrome protein, *Drosophila* Chorion protein and Human eIF-2, respectively (Morris et al., 2013). They are ideal models for exploring the sequence – structure relationship as they each form a different distribution of polymorphic fibrils. Assembly polymorphism is differences in the hierarchal arrangement of proto-filaments within a fibril (chapter 3) and has been shown to be sequence dependant, with the residue specific core interactions providing the molecular architecture for distinct mechanical properties (Yoon et al., 2014). Amyloid fibrils have been shown to have impressive mechanical properties, with tensile strength like steel and a bending rigidity comparable to spider silk (Smith et al. 2006). The interactions attributed to β sheet is thought to provide amyloid with such mechanical properties; observed when the level of β sheet structure is reduced within amyloid fibrils by the replacement of β strands with α- helices (Vandenakker et al. 2011). This resulted in a reduction in persistence length, and therefore a reduction in rigidity. Additionally, the directionality of interacting β sheets has been shown to affect the rigidity of amyloid fibrils, with anti-parallel β sheets granting an increased rigidity. The same team also suggested a mechanical dependence on the steric zipper interface, meaning the amino acid sequence is also responsible for the mechanical stability of amyloid fibrils (Yoon et al. 2014).

Here, using Atomic force microscopy (AFM) and sonication we present a quantitative analysis of the fragmentation of HYFNIF, VIYKI and RVFNIM showing a relationship between the mechanical properties of distinct populations of polymorphic fibrils and their fragmentation behaviour. Using alpha- synuclein, an amyloid that forms fibrils which display relatively little assembly polymorphism we further support a sequence dependent relationship towards the mechanical properties of amyloid.

All three of the WALTZ identified peptides form fibril populations with differences in assembly polymorphism and display different fragmentation behaviour, potentially due to differences in the net mechanical properties of the adopted structures within the population. An increase in persistence length conferred by fibrils with an increased number of associated filaments, results in fibrils which are more resistant to fragmentation than the fibril morphologies adopted by alpha-synuclein; albeit to different extents. The steric zipper interface within protofilaments and molecular interactions between filaments, attributed by the amino acid sequence; is thought to determine the mechanical properties of amyloid fibrils. Additionally, the presence of an anti-parallel β sheet packing interface for α-synuclein and RVFNIM has been shown to reduce fragmentation stability of the polymers despite a potentially increased hydrogen bond density; providing an environment for enhanced protofilament rigidity (Yoon et al., 2014). Taken together with our results here, the mechanical properties of the core protofilament structures likely transition to the hierarchically assembled mature fibril, where differences in assembly polymorphism (determined by filament arrangement) result in differences in the persistent length of the fully assembled polymer, and stability towards fragmentation. Additionally, the relationship between persistence length and fragmentation stability is likely one influenced by contour length. Wherein the difference between the two values is indicative of displayed mechanical behaviour, and

therefore likely-hood of breakage through buckling. However, when the persistence length of 2 fibrils is considered at equal contour length, it seems that increased rigidity provides enhanced stability. While we can only speculate as to whether the breakage along a fibril axis is influenced by specific protein or peptide sequences; sequence determined polymorphic populations display differences in fragmentation behaviour, supporting a sequence-structure relationship for fragmentation stability.

Materials and methods

WALTZ Peptide synthesis

Each of the WALTZ peptides were synthesised with an N-terminal acetylation and a C-terminal amidation. Multistage synthesis using an Fmoc protection mechanism created lyophilised powder which had a HPLC achieved purity >95% (JPT peptide technologies. Separate batches were synthesised on site using the same method (Kevin Howland, University of Kent).

Protein purification

E. coli cultures expressing Sup35NM were sonicated in a Sup35NM Lysis Buffer (SLB) suspension. Lysates were clarified via centrifugation prior to Ni-NTA resin incubation. Ni-NTA attached material was eluted in 1ml of Melki Lysis Buffer Denaturing (SLBG) and subsequently purified by size exclusion chromatography using a GE Healthcare HiLoadTM 16/600 SuperdexTM 200pg, from GE Healthcare (28-9893-35), using Denaturing Sizing Buffer (DSB) on an ÄktaPrime Plus system at a flow rate of 0.3ml/minute. For α-synuclein, E. coli expression cultures were sonicated in an α - Synuclein Lysis Buffer (SLB) + Protease Inhibitor suspension (Complete Protease Inhibitor tablets Roche, 11697498001). HCL based protein precipitation and subsequent NaOH neutralisation was performed prior to clarification via centrifugation at 13000rpm

and treatment with (NH4)2SO4. The Precipitate was collected via centrifugation and resuspended in α-Synuclein Ion Exchange Entry Buffer (αEB). Purification was by ion exchange chromatography using an ÄktaPrime Plus system with a strong anion exchange HiTrapTM Q FF 5ml cartridge (GE Healthcare, 17-5156-01).

In vitro polymerisation

To prepare populations of WALTZ peptide fibrils 1mg of the respective lyophilised powder is suspended in 100ul of 0.2µm syringe filtered filter sterilized milli-Q water to a final concentration of 10mg/ml. Due to the potential instantaneous formation of fibrils, the solution is not vortexed to mix; to avoid potential fragmentation. The solution was mixed by pipette, completely dissolving the peptide powder. The solution is incubated at room temperature for 1 week prior to experiments and imaging. This time allows for 'maturation' of fibrils providing equal time for morphological adoption across all 3 WALTZ samples prior to imaging. All manipulations of the WALTZ peptide solutions were with low bind tips (VWR 53503-294 (with SpidersilkTM surface)), and Protein LoBind 1.5ml tubes (Eppendorf, 022431081). For α-synuclein, 2.5ml of protein in α -Synuclein Ion Exchange Elution Buffer (SElB) was buffer exchanged with α-synuclein fibril forming buffer (25mM sodium phosphate buffer, pH7.5) using a PD10 column per the manufacturer's instructions. In addition, 0.2µm syringe filtered sodium azide was added and the sample was incubated at a concentration >200µM and at 37°C with orbital agitation (160rpm) for 2 weeks. For Sup35NM, 2.5ml of protein in DSB was buffer exchanged with 3.5ml of Sup35 fibril forming buffer (25mM sodium phosphate 50mM NaCl) using a PD10 column according to the manufacturer's instructions. Fibril forming solutions were made at 10µm and required incubation at 37°C without agitation. All manipulations of the Sup35 solutions were with low bind tips (VWR 53503-

294 (with SpidersilkTM surface)), and Protein LoBind 1.5ml tubes (Eppendorf, 022431081).

Fibril Fragmentation

Fragmentations of α-synuclein and the WALTZ peptides were conducted in parallel; therefore, the same time points were used (1s, 2s, 3s, 4s, 5s, 6s, 12s, 24s, 48s). Sonication was used with a 20% amplitude and samples were on ice for each time point's duration. Fibril solutions were diluted to a concentration of 0.1mg/ml (7uM) using 25mM sodium phosphate buffer (pH7.5) for α-synuclein, and 25mM sodium phosphate and pH1 HCL for WALTZ peptide fibrils. WALTZ fragmentation solutions were a final concentration of 2.5mM sodium phosphate at pH2. 10 µm Sup35NM in fibril forming buffer (25mM sodium phosphate 50mM NaCl) was sonicated for 4minutes in 5 second bursts at 20% power. Fibril solution was inside Protein LoBind 1.5ml tubes (Eppendorf, 022431081), and kept on ice for the duration of fragmentation.

AFM sample preparation

Following sonication, 20µl of the respective sample is deposited on a freshly cleaved mica surface (Agar scientific, F7013) attached to steel discs (Agar scientific). The WALTZ peptides are deposited with no dilution; a concentration of 0.1mg/ml and are incubated on the surface for 45 minutes. α-synuclein Is diluted prior to deposition; to a concentration of 0.003mg/ml and incubated on the mica surface for 5 minutes. Sup35NM is not diluted following fragmentation and is deposited at 10µM with a 5-minute incubation. Following appropriate incubation, the sample is washed with 1ml of filter sterilized milli-Q water, then dried using a stream of nitrogen gas with filter paper to aid in the removal of residual liquid from the mica surface. The samples were then imaged using a Bruker Multimode 8 AFM with a Nanoscope V controller operating under peak force tapping mode. With a ScanAsyst probe (silicon nitride triangular tip with tip

height = 2.5-8µm, nominal tip radius = 2nm, nominal spring constant 0.4N/m and nominal resonant frequency 70kHz. The imaging software used is Nanoscope 8.15. Images were collected with a scan size of 30 x 30µm and at a resolution of 2048 x 2048 pixels. A scan rate of 0.305Hz was used with a noise threshold of 0.2nm (manually set within the nanoscope parameter selection workspace). Additionally, Z limit was reduced to 1.5µmand the peak force set point set to 0.05- 0.15mV (depending on sample).

Image analysis

AFM images of fragmented fibrils were analysed using an in-house Matlab application (W.-F. Xue, 2014). The application automatically searches for particles within AFM images based on input parameters including a z height (nm) threshold, apparent width of the particles (pixels) and search angle Phi (°). Data extracted from fibril particles was subsequently quantified in regard to particle number, height, and contour length. Both per-particle and average population height eludes to the relationship between morphology and fibril breakage. Therefore, each time point was analysed on both a per-particle and population basis. For each sample, time point data is comprised of 3 images per time point, and two independent repeats, resulting in six images per time point in total. Extracted particle data is filtered to remove particles where regions of clumped fibrils >20nm and traced particles that are comprised of <2 pixels, which cannot be confirmed as fibrils. Removal of these particles from data set prevents length and height bias that may not be representative of morphology dependent reduction in fibril length. All distribution plots are constructed using Matlab, and represent >1000 particles per time point. Each samples data set (WALTZ peptides and α - synuclein) is comprised of >30,000 particles; providing sufficient data to determine reduction in length representative of each respective amyloid species.

To determine the Persistence length (L_p) of each sample (fibrils formed from RVFNIM, HYFNIF, VIYKI and α-synuclein) unfragmented fibrils were traced using an in-house Matlab application (W.-F. Xue, 2014). Images used for analysis were collected using AFM at a resolution of 2048 x 2048 pixels and a scan size of 30 x 30 µm, with a pixel density of 14.65nm/pixel. Each image was post-processed to remove sample tilt and bow, in addition to the removal of image streaks which may identify as polymers. Each dataset consisted of 2000-10000 individual fibrils collected from multiple images and was analysed in Matlab using a script (W.-F. Xue, 2014) to determine the population L_p. The script utilises equation 5, which is a probability based method for determining the maximum likelihood estimation of persistence length. This determines the most likely L_p value.

$$p\left(\theta_{2D}(L)\right) = \sqrt{\frac{L_P}{2\pi L}}\, e^{-\frac{L_p \theta_{2D}^2}{2L}}$$

Equation 5.

Equation based on the Probability distribution of bending angles against length; determining persistence length (L_p) of a flexible polymer. Here L_p is defined by the probability of bending angles at a segment length (L) up to the full contour length] This allows for determination of the rate of decay of tangent vectors with increasing segment length (L).

To determine standard error L_p for each sample, datasets were resampled 500 times using a bootstrap method. Each resampled dataset was analysed using the L_p Matlab script in the same manner, then the Mean value and standard deviation was calculated

from all resampled datasets. Providing mean values to determine the standard error of the L_p estimates.

Results and discussion

Mechanical fragmentation of amyloid fibrils

Fibrils were formed from monomer of the WALTZ identified peptide sequences RVFNIM, HYFNIF, VIYKI and from α-synuclein. Fragmentation by controlled sonication was carried out across an initial 1-6 second time course with subsequent doubling time point of 12, 24 and 48 seconds. This resulted in a reduction in length over time of the respective fibril populations. The WALTZ peptide samples were fragmented at 0.1mg/ml, however across initial experiments they displayed differences in fibril distribution on the mica surface. This difference could be due to a change in interactive behaviour amongst individual fibrils and mica. As fibrils fragment, hydrophobic regions of the core structure may become exposed which can result in an increased clumping between fibrils(Karen E Marshall et al., 2014). This may be protective in that the fibrils length is preserved over time as the equivalent force required to sheer an individual fibril cannot penetrate and have the same effect on an aggregate of fibril clusters. Equivalent techniques of fragmentation result in an enhanced display of fragmentation resistance, which may in part be attributed to the clumping nature of the fibrils (Figure 29). Figure 29 presents a population of fibrils fragmented for 4minutes; where the average contour length is like that of the WALTZ peptides when fragmented for under a minute. The ability of a fibril to interact with other fibrils and form intertangled networks of fibrils is possibly due to the specific sequence of amyloid species. Differences in hydrophobicity of the exposed residues on a fibril surface likely contribute to the likelihood of fibril clumping. Therefore, to limit the clumping potential of the three WALTZ peptides the conditions of fragmentation were tuned. A reduction to pH2 and the addition of 25mM sodium

phosphate reduced the clumping of the fibrils, allowing a quantitative analysis of the fibrils on an individual basis. The conditions were only changed at the point of fragmentation; elongation is simply in water.

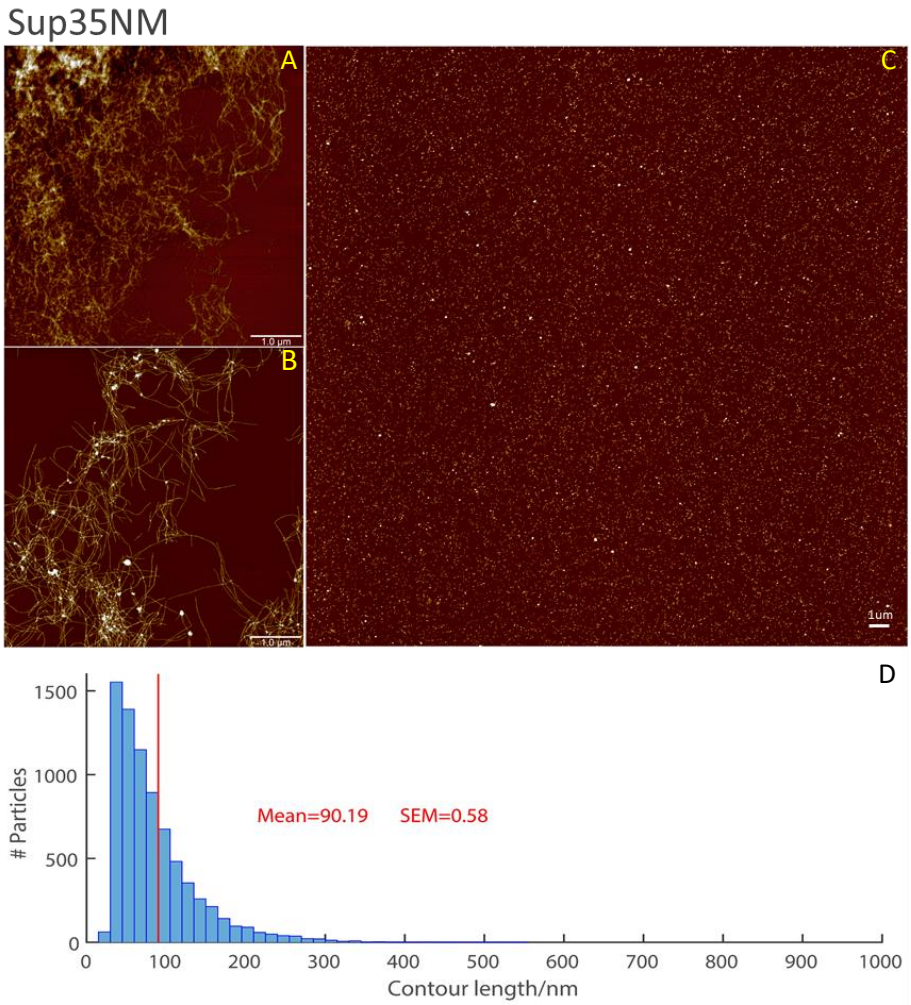


Figure 29. Comparative fragmentation of Sup35NM fibrils

AFM images of Sup35NM, used for comparison with the WALTZ peptides. The surfaces seen in (A) and (B) are 6 x6 µm and contain unfragmented fibrils captured at a resolution of 2048 x 2048. This conforms to the image collection technique used to characterise the WALTZ fibrils, allowing similar detail to be obtained for comparison. For a fragmentation comparison with the WALTZ fibril populations, Sup35NM fibrils were fragmented for 4 minutes by sonication (C). This time course is considerably longer than the WALTZ peptides, however Sup35NM fibril populations present an enhanced stability towards fragmentation. The average particle length of fragmented Sup35NM was quantified in Matlab (D) with an average length similar to the final length of the WALTZ peptides; which take considerably less time to fragment under the same conditions.

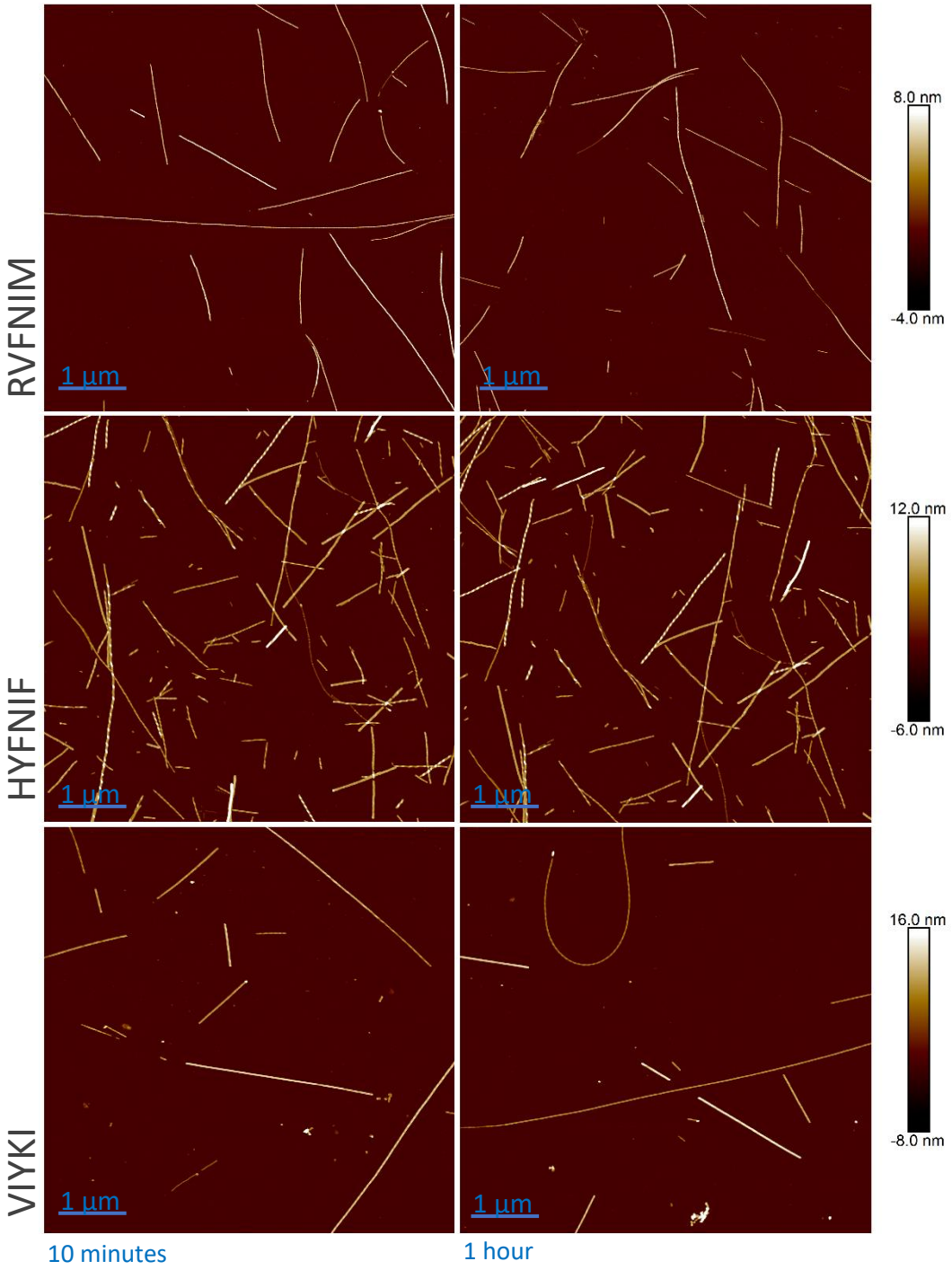
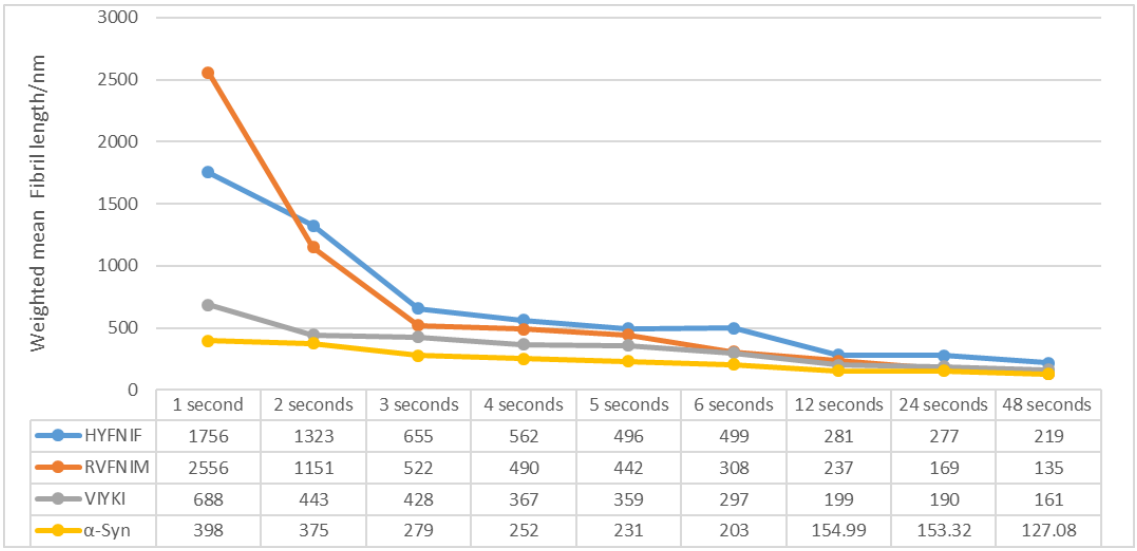


Figure 30. Determining similar morphologies with a change in conditions

The fragmentation of amyloid fibrils is likely affected by structure, which in turn may be affected by environment condition. The alterations of pH and ionic strength used for the fragmentation and deposition of the WALTZ peptide fibril samples open a potential for morphological manipulation that may affect fragmentation behaviour. This figure presents a control comparison of the respective fibril samples. Images on the left represent fibrils that have been prepared with a 10-minute incubation of fibrils which were placed in a pH2 2.5mM Sodium phosphate solution at the time of deposition. Images on the right represent fibrils that have been incubated in the same buffer conditions for 1 hour prior to deposition and standard sample preparation. Fibril samples appear identical regardless of an increased incubation time in the deposition buffer. Surfaces were subsequently imaged by AFM for a scan size of 6 x6 µm and a resolution of 2048 x 2048 pixel. Each WALTZ peptide is scaled with a colour bar defining the height contrast.

Each timepoint was fragmented as a separate sample, therefore not allowing the fibrils to rest in-between timepoints. Additionally, each timepoint sample was at the same volume, meaning that the fibril mass within each sample is conserved as much as possible across the time course. Each time point was imaged via atomic force microscopy(AFM). Each time course was run twice for the final data set, allowing a final particle number in the thousands for each time point; which constitutes a quantitative analysis, likely representative of the respective fibril populations.

The height and length information are key for this analysis, therefore this information and more was extracted from each image using an in house automated tracing application. This allows for the collection of individual fibril dimensions, which were subsequently analysed using Matlab. The data itself was corrected for any bias towards small potentially non fibrillar particles present within the unfragmented samples, and for large clumped aggregates over a certain height which do not offer accurate length determination of individual fibrils. Therefore, the average height and length values are determined independently of such particles that may weigh the data in a way that is unrepresentative of the reduction in length over time. Figure 32 displays the reduction in length over time for fibrils formed from the 3 WALTZ peptides and fibrils formed from α-synuclein. The images show the overall reduction in fibril length, but most importantly displays a difference in length between the samples; the timepoints shown are 1 second, 6 seconds and 48 seconds respectively. The initial timepoint here is 1 second, and this is representive of the least fragmented state within this analysis with the initial unfragmented length difficult to accurately determine.



	1 second	2 seconds	3 seconds	4 seconds	5 seconds	6 seconds	12 seconds	24 seconds	48 seconds
HYFNIF	1756	1323	655	562	496	499	281	277	219
RVFNIM	2556	1151	522	490	442	308	237	169	135
VIYKI	688	443	428	367	359	297	199	190	161
α-Syn	398	375	279	252	231	203	154.99	153.32	127.08

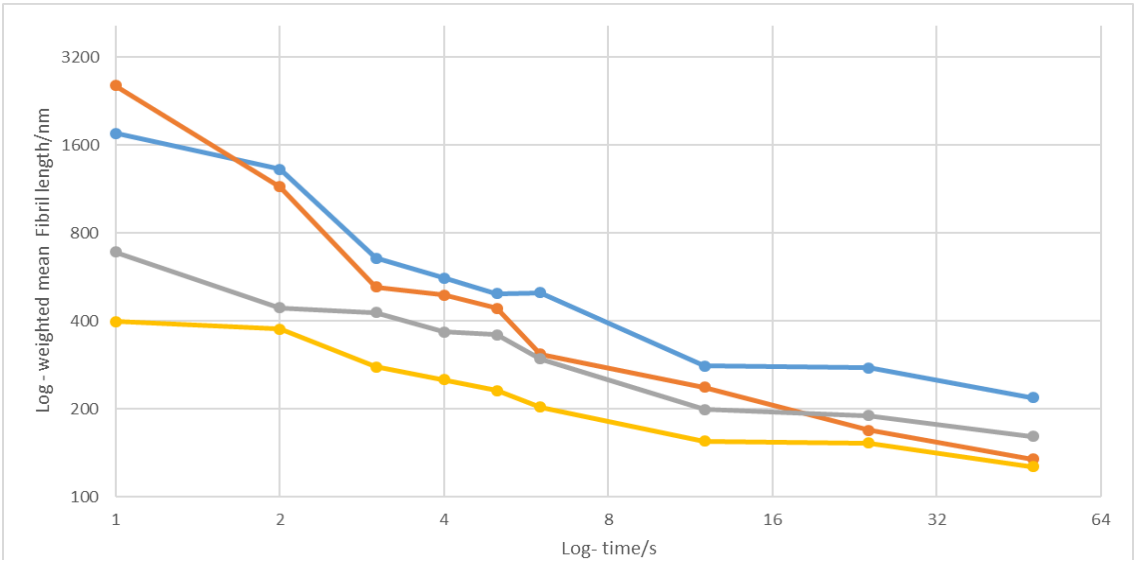


Figure 31. Quantifying the reduction in length over time of each sample

The reduction in length over time, represented by the weighted mean length per time point for fibrils formed from each of the WALTZ peptides and α – synuclein on linear scale (A) and Logarithmic scale (B). The graphs show that HYFNIF displays the longest final length and α – synuclein the shortest; RVFNIM displays a similar final length to α – synuclein however maintains a longer fibril length through the initial phase of the fragmentation, almost mirroring the reduction in length observed in HYFNIF. VIYKI displays a fragmentation behaviour like α – synuclein, with an initial large reduction in length and proceeds to maintains a relative plateau. Generally, for all fibril samples as the particles reduce in length their stability towards further fragmentation increases suggesting a length dependent involvement in fragmentation behaviour. However, at equal lengths, HYFNIF for example; still displays a higher stability towards fragmentation than α – synuclein and the other WALTZ peptides. This suggests that at equal lengths, the sequence and subsequent structure influences mechanical stability. Each time point value on the graph is comprised of 1000's of analysed particles from multiple AFM images and repeated fragmentations.

Fibrils formed from the WALTZ peptides were observed to consistently exceed 20µm in length, which means that due to the difficulty of containing sufficient numbers of individual fibrils within single images; unfragmented length was not used as the initial timepoint. Subsequent analysis follows the reduction in length from this point. The unweighted mean lengths in Figure 32 are represented by the red vertical line and represent the mean as observed by the images collected. Weighted reduction in length can be seen in Figure 31 and regardless of the weighted values, the comparative stability patterns are the same. Therefore, the observations seen within these experiments accurately represent the fragmentation stability of the WALTZ peptides and α-synuclein. To determine a potential change in fibril structures, qualitative AFM analysis was conducted on samples that had been incubated post elongation in the fragmentation buffer comparatively to control fibrils which had not had their conditions modified (Figure 30). The results show that during this short incubation in altered conditions, fibril morphology does not appear to change.

Figure 32. The reduction in length of amyloid fibril following fragmentation

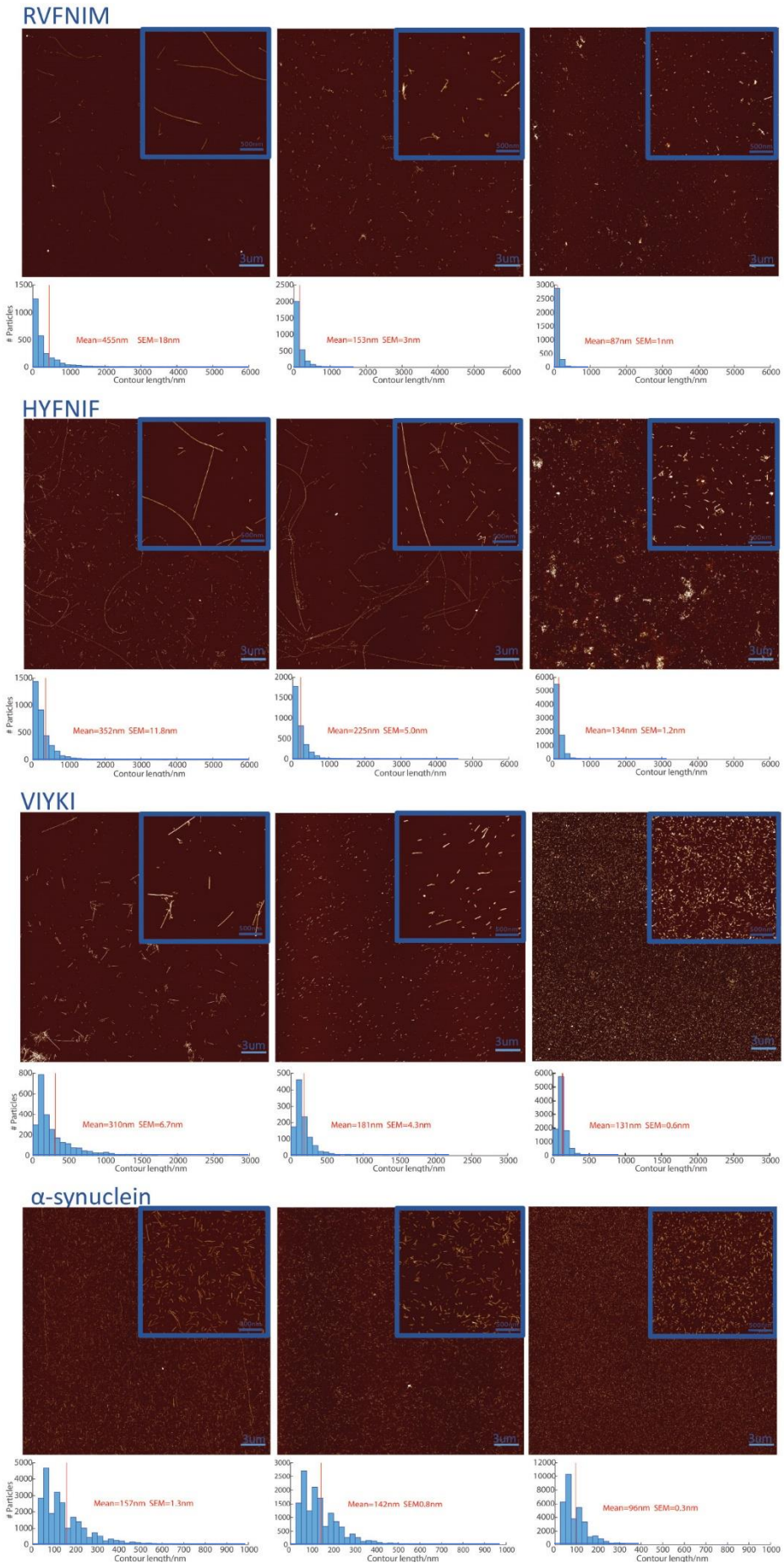


Figure 32 legend.

AFM images displaying the reduction in length of fibrils formed from α – synuclein and WALTZ peptides. Fragmentation is by sonication and across a 48 second time course; images displayed are for 1 second, 6 seconds and 48 seconds. Images are 30 x 30 µm and were captured with a resolution of 2048 x 2048. The scale bar represents 3µm and the blue area is a crop equal in size. In addition to each image set is the corresponding contour length histogram which displays the distribution of mean length across the fragmentation time course. The mean contour length varies slightly from the weighted mean, in that HYFNIF and VIYKI are similar in final length; with both being longer than α – synuclein. RVFNIM however displays the largest reduction in mean length. The fibrils analysed represent the observed distribution and not necessarily the entire population; therefore, weighted mean lengths were also used for the final determination of reduction in length (Figure 31).

As seen in Figure 32 Each of the WALTZ peptides display a higher initial length than α-synuclein, with RVFNIM displaying the longest fibrils after 1 second of fragmentation. Considering the similarity observed of unfragmented fibrils, this suggests that RVFNIM displays fibrils that are more resistant to fragmentation initially. The difference between HYFNIF and VIYKI is relatively small, however the initial reduction in length compared to RVFNIM is considerable and suggests that as an observed population, the fibrils formed by VIYKI and HYFNIF are initially less resistant to fragmentation than RVFNIM. In addition to this initial comparison, α-synuclein displays an even larger reduction in initial length after 1 second suggesting a further reduction in mechanical stability to the WALTZ peptides. Subsequent time points however show a change in the relationship between the samples and present a potentially increased rate of fragmentation for RVFNIM. Despite the longer initial length, RVFNIM after 6 seconds of sonication displays an average fibril length of 153nm which is less than both HYFNIF and VIYKI with average lengths of 225nm and 181nm respectively. Here, RVFNIM displays an average length similar to α-synuclein (142nm). Following the initial 6 seconds, doubling time points are conducted which expose the fibrils to increasing amounts of sonication. The final timepoint of 48seconds shows HYFNIF as displaying the longest fibrils with an average length of 134nm which is seemingly indistinguishable from VIYKI at 131nm, suggesting a similar overall mechanical stability between these two populations of fibrils. RVFNIM

however, despite showing an initial longer length results in an average length comparable to α-synuclein, 87nm and 96nm respectively. HYFNIF and VIYKI seemingly fragment at similar rates, with similar initial and final lengths. However, length distributions seen in Figure 32 suggest that there is a difference in fibril size distribution. In each timepoint, HYFNIF displays fibrils of increased length compared to VIYKI whilst also displaying a higher number of fibrils that are relatively short in length. The histograms seen in Figure 32 display blue trails along the x-axis which represent bins of relatively low frequency, with each bin containing few particles for the respective length. However, the length in which the blue trail extends, represents the boundaries for each distribution profile. HYFNF's length distribution extends further than VIYKI suggesting that even though they are few, HYFNIF's population contains fibrils of increased length; potentially due to a specific morphology or class with increased fragmentation resistance. Conversely to this, α-synuclein displays a relatively tight distribution which is to be expected as the morphological landscape for this amyloid is comparatively narrow. The length distribution of RVFNIM is more dynamic in that the initial range of length is comparable to that of HYFNIF with potential fibril classes displaying an extended contour length. However, as the time course proceeds, the range of lengths within the distribution narrows, suggesting that all displayed fibril classes have arrived at a relatively similar length. The dynamic behaviour of RVFNIM is of interest here, and Is suggestive of a difference in fragmentation rate amongst the amyloid species. Such a difference in fragmentation rate and final lengths, regardless of the fact the 3 WALTZ peptides form fibrils of similar length and from monomer that is highly comparable in size(Morris et al., 2013), is indicative of a morphological difference amongst the respective fibril populations. Therefore, the fibrils height and length

distributions across the time course can shed some light on how different morphologies may be fragmenting at different times within the fibril populations.

Structural relationship to fragmentation

Each of the WALTZ peptides form fibrils that are polymorphic. In general, amyloid is known to be polymorphic in regards to the assembly of filaments within a fibril (chapter 3) (Annamalai et al., 2016; Meinhardt et al., 2009; Tycko, 2014). The degree in which an amyloid species is polymorphic does however vary, and this can allow access to a different range of morphologies depending on the protein or peptide sequence. Fibrils formed from the WALTZ peptides here have been seen to be polymorphic (Morris et al., 2013). We have characterised them with a focus on the assembly polymorphism displayed (chapter 3) and have suggested the range of possible morphologies displayed by each of the fibril populations(myself). This is regarding the number of filaments within a fibril and the twisting and arrangement archetypes that they adopt. What we found is that each of the WALTZ peptides, different in sequence but similar in number of residues per monomer, form fibril populations that display different types and different distributions of morphological classes. Therefore, it seems likely that these different classes present inter-class differences in mechanical properties due to changes in overall fibril width. In addition to this, the sequences themselves may provide mechanical discrepancies due to different residue profiles and subsequent inter and intra-filament interactions. The Height distribution of a fibril population allows the basic determination of the morphological landscape. The height of a fibril alone is not enough to determine the morphological class, as different twisting archetypes can be adopted with the same average height.

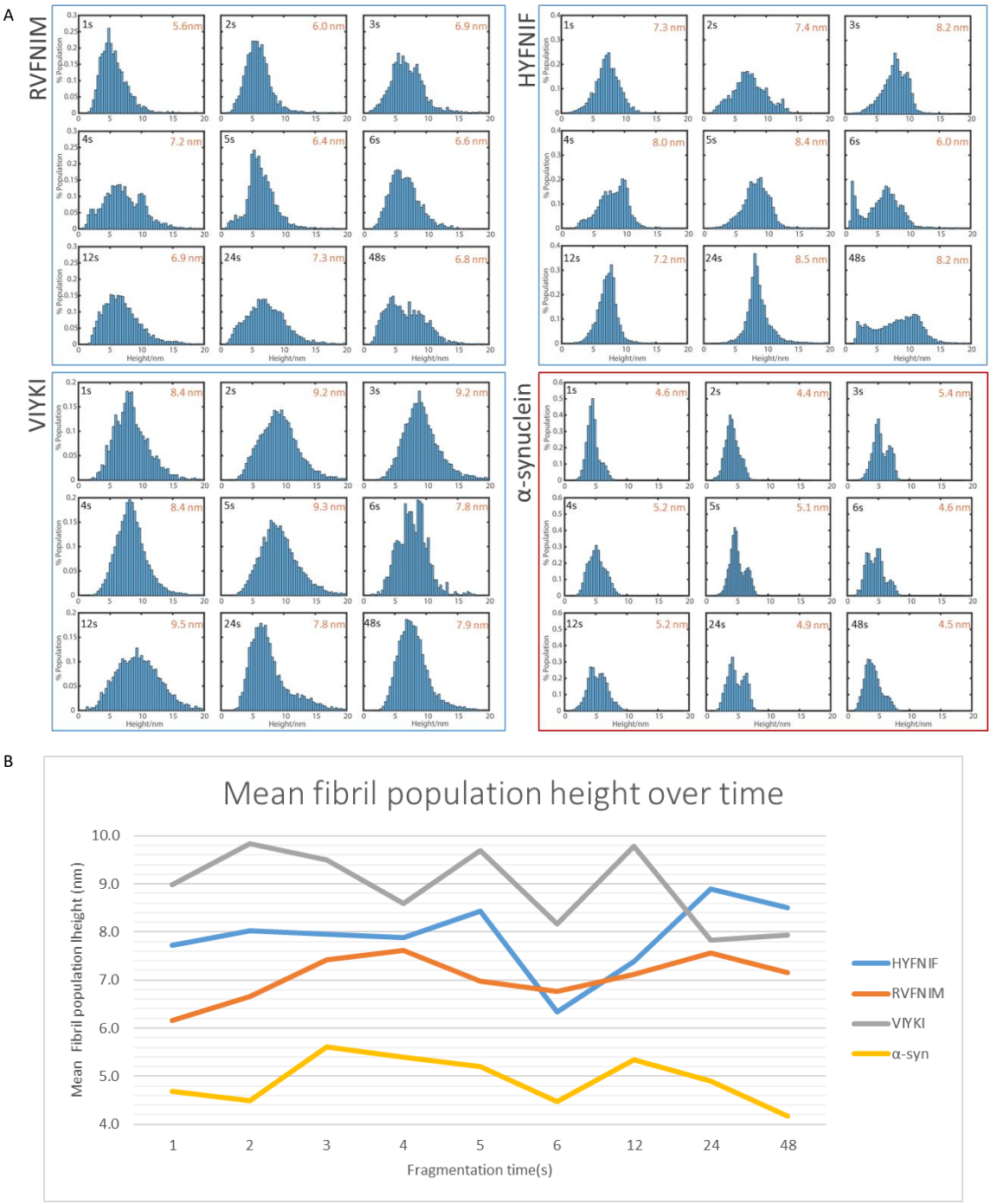


Figure 33. Observing the change in particle height distribution

A series of histograms (A) displaying the shift in fibril height distribution during fragmentation across 48 seconds. α-synuclein maintains a relative static height distribution during fragmentation; which contrasts with the WALTZ peptides that display slight change in height distribution. The shift in height distribution is indicative of morphology dependent fragmentation, which differs depending on the respective peptides adopted fibril morphologies. Distribution shift while present, is only slight with an average height relatively maintained per time point; this shift would be expected to be larger given the polymorphous heterogeneity between the WALTZ fibril populations. This suggests a sequence dependent influence on fragmentation stability in addition to the likely structural relationship. Distribution plots were constructed in Matlab. (B) displays the change in average particle height over time.

However, the height alone is representative of the fibrils thickness meaning that an analysis of the height distribution shift during fragmentation allows for the determination of height class dependent fragmentation. Fibrils with different heights, and therefore mechanical properties may be fragmenting at different times within the population and this can be observed in Figure 33. Here, height distributions are shown for each sample population across the full time-course. α-synuclein serves as a control here, as the fibrils formed are more homogenous in regard to morphology than the WALTZ peptides. This can be seen by the relatively close height distribution at each of the time points, which does not shift considerably from the mean population height. Therefore, α-synuclein fibrils within a population all likely fragment within a similar time frame; requiring a relatively equal force exposure. A shift in height distribution is indicative of a height class fragmenting more readily than other classes, resulting in an effective increase in the population of fibril particles displaying that height. The Height profiles for each of the WALTZ samples show a shift in distribution across the fragmentation time-course. The shift is not in one direction, with the average height of the population fluctuating at different time points suggesting the differential breakage of fibril classes. The unfragmented average height of RVFNIM is 8.9nm (chapter 3) and the prominent morphology displayed within the population is 10nm in height suggesting a favourable fibril height class. In fact, each of the WALTZ peptides display a favoured fibril morphology, with a defined average height (chapter 3). HYFNIF has an average height of 7.4nm and a prominent height class of 6nm, VIYKI displays 9.7nm and 9nm respectively. It should also be highlighted that per our characterisation work, VIYKI displays a relatively high height distribution with only few numbers of fibrils with low height (~6nm) therefore displaying a population with the largest overall height. HYFNIF

is the converse to this, with the lowest population height and a high frequency of low height fibril classes. RVFNIM is effectively in the middle of this, with a high frequency of large fibrils (prominent class is 10nm) but also a relatively increased frequency of low height fibrils compared to VIYKI. This difference in height landscape amongst the 3 populations can be translated into the height shift seen in Figure 33. RVFNIM at 1 second, displays a decrease in average height representing an increase in the number of 'thin' fibrils because of an increased fragmentation sensitivity for this class. However, seeing as this is not the prominent fibril height class, it is likely that this class of fibrils is relatively intact which explains RVFNIM's initial and apparent resistance to fragmentation (Figure 31). At 3 seconds, the Height distribution displays an increased shift in average height which coincides with the comparatively (to HYFNIF) rapid decrease in contour length following in initial display of resistance like seen with HYFNIF. This point may represent the subsequent breakage of larger fibril classes which display an intra-population increase in fragmentation stability but ultimately do not display the same stability in comparison with HYFNIF. This suggests a collaborative relationship between sequence and structure towards fragmentation stability wherein both the sequence and the assembly morphology contribute to mechanical stability, but neither one is independently responsible for the mechanical properties of amyloid. The initial resistance, and subsequent decline in length to the lowest contour length of the WALTZ peptides may be a result of the wider range of morphologies displayed by RVFNIM (chapter 3). This could be attributed to a sequence that is more promiscuous in its adopted polymorphism, granting access to a higher number of morphologies that are mechanically fragile. Meaning that while certain morphologies may display enhanced fragmentation resistance, overall the average length is lower than VIYKI and HYFNIF which display a tighter range of polymorphic fibrils. HYFNIF displays a gradual increase

in average height, with an initial resistance indicated by a relatively large average A length and no considerable shift in average height, meaning that at this stage the population has fragmented equally. However, as fragmentation proceeds you see an increased shift in average height suggesting an increase in the population of fibrils larger than the average height, indicative of a height class less resistant to fragmentation than the prominent 6nm height class seen by HYFNIF. The Final time point of 48 seconds displays an increased average height in comparison to the initial time point, fitting with an increase in population of larger fibrils. The opposite behaviour is observed for the fragmentation of VIYKI in that the initial height distribution is preserved and B representive of the unfragmented average population height but the height gradually decreases in distribution. Suggesting an increase in population of fibrils of a height lower than the most prominent fibril class (9nm). Together the height profiles suggest that for VIYKI and HYFNIF initially the population fragments relatively evenly up to a certain point in which the respective prominent class displays an enhanced resistance to fragmentation represented by an increase in initially lower frequency height classes. RVFNIM however does not seem to fragment as an equal population at any point, with inconsistency in the displayed fragmentation resistance and fluctuations in the height distribution. Indicative of a low frequency presence of highly resistant fibril classes which initially weigh the population length to similar values of that seen with HYFNIF (figure). However, ultimately the population displays an overall reduction in length that is more considerable than HYFNIF and comparable to α-synuclein.

Figure 31 represents the mean fibril length over time, and displays the rate in which each of the fibril populations reach their respective final length. The weighted mean allows for the meaningful representation of lower frequency contour lengths observed in the sample areas; enhancing the representation of the entire population. The initial

time point suggests RVFNIM as having the longest average contour length out of all samples, however after an additional second of fragmentation declines rapidly relative to the other samples. HYFNIF and RVFNIM generally display enhanced initial length accompanied by a greater rate of reduction in length than VIYKI and α-synuclein. This itself is suggestive of a length dependent basis for fragmentation stability in that the longer a fibril is the more susceptible to breakage it is. For polymers of similar mechanical properties this is to be expected and is likely representive of the number of possible points of breakage along the fibril axis i.e. a longer fibril will have a larger number of possible points of breakage than a shorter one (W. F. Xue & Radford, 2013). Meaning that as a fibril reduces in length, it becomes increasingly more resistant to further breakage. However, differences in the overall mechanical properties are likely present due to the polymorphism present and sequence dependent properties. Therefore, for a representive comparison of stability between the samples; subsequent changes in length from similar lengths should be analysed. For example, HYFNIF the seemingly most resistant sample; at a time of 12 seconds has a comparable length to α-synuclein at 3 seconds (~280nm). An additional 3 seconds for α-synuclein reduces the length to 203nm, while an additional 24 seconds of fragmentation is required to reduce the length of HYFNIF to 219. Suggesting that at comparable lengths, each sample displays different stabilities towards fragmentation.

HYFNIF with a shorter monomer sequence and enhanced morphological landscape displays an increased overall stability towards fragmentation than α-synuclein. VIYKI which displays a similar degree of polymorphism as HYFNIF, Fragments to a considerably lower initial length, however subsequently fragments less rapidly. The final length of VIYKI is lower than that of HYFNIF at 161nm which is also longer than α-synuclein. When Both populations are around 650nm (1s for VIYKI and 3s for HYFNIF) an additional 3

seconds of fragmentation results in a similar reduction in length to ~400nm. Albeit HYFNIF is 499nm, almost 100nm longer; at these length scales 100nm is less considerable a difference than when they are at the end of fragmentation. However, HYFNIF is expected to be longer generally, considering its longer final length. This does show however, that VIYKI fragments more initially and displays a shorter final length despite the similarity in polymorphism. This may suggest a stronger sequence dependent basis for fragmentation stability, however while the polymorphism is similar regarding the types of morphology present; the predominant classes do differ. Still presenting a complex relationship between sequence, structure and mechanical stability in which it is likely that each characteristic contributes to a fibrils resistance towards fragmentation. A potential factor towards the differences in stability between HYFNIF and VIYKI is the fact that VIYKI is a residue shorter than HYFNIF which may confer a difference in breakage point given the larger number of monomer interfaces along an extended polymer axis. Regions of a fibril such as the monomer-monomer interface may present regions of fibril breakage due to a weaker bonding interface of hydrogen bonds than the bonds within the peptide chain itself (Fitzpatrick, Vanacore, & Zewail, 2015). However contrary to this is the comparison with α-synuclein, in which the monomer unit is considerably longer than the WALTZ peptides. Meaning, the adopted morphologies likely confer a lower overall fragmentation stability.

RVFNIM displays a larger range of morphologies than HYFNIF and has the same number of residues in its peptide sequence (chapter 3). The displayed behaviour for RVFNIM throughout the fragmentation is again different to HYFNIF and VIYKI. The initial length of RVFNIM is in fact longer than HYFNIF but similar regardless. However, the final length is the shortest out of the 3 WALTZ peptides and is similar to α-synuclein. Throughout the initial 6 seconds the length of RVFNIM is similar to that seen with HYFNIF, however

the subsequent doubling times separate the populations as the fibrils become shorter. This is potentially when the number of potential breakage points are similar and the differentiating factor is the morphological differences present (W. F. Xue & Radford, 2013). The difference in overall number of potential breakage points along a fibril axis is more prominent on fibrils of increased contour length i.e. unfragmented. However, as the fibrils become shorter the number of points of breakage along an axis of reduced length will become more similar despite initial differences (W. F. Xue & Radford, 2013). It is possible that at increased lengths, the overriding property contributing to fragmentation stability is the number of breakage points along the fibril. However, as the fibril reduces in length and the number of points per particle becomes similar, the morphological class differences between the respective populations contributes more as a differentiating factor for mechanical stability. This is indicated by the behaviour seen in comparison to HYFNIF of the other 2 WALTZ peptides. RVFNIM with the same monomer residue number initially displays similar resistance as HYFNIF, while VIYKI with a shorter sequence appears more fragile. However, once the fibrils fragment to similar lengths, VIYKI with a similar morphological landscape as HYFNIF displays an enhanced resistance to fragmentation than RVFNIM which subsequently fragments with higher sensitivity than either HYFNIF or VIYKI. These observations further suggest a complex but dynamic relationship between sequence structure and mechanical stability.

Determining height class influence on fragmentation

To determine the time frame in which the height classes begin to differentiate regarding their mechanical stability, we looked at the height distribution of each population plotted vs per-particle contour lengths (Figure 34). While each fibril population displays a predominant class (chapter 3), the distributions in figure 34 show that this class may not necessarily be indicative of a stability defining class. This is because at the earlier

timepoints, there is not a dominant height class which displays the longest fibrils. In-fact the particles which display the longest lengths can be seen at multiple height classes. Additionally, the fibrils of a set height class display a wide range of contour lengths suggesting that not all fibrils of a set height fragment at the same time. However, timepoints 6 and then 48 seconds in Figure 34, show widening of height distributions.

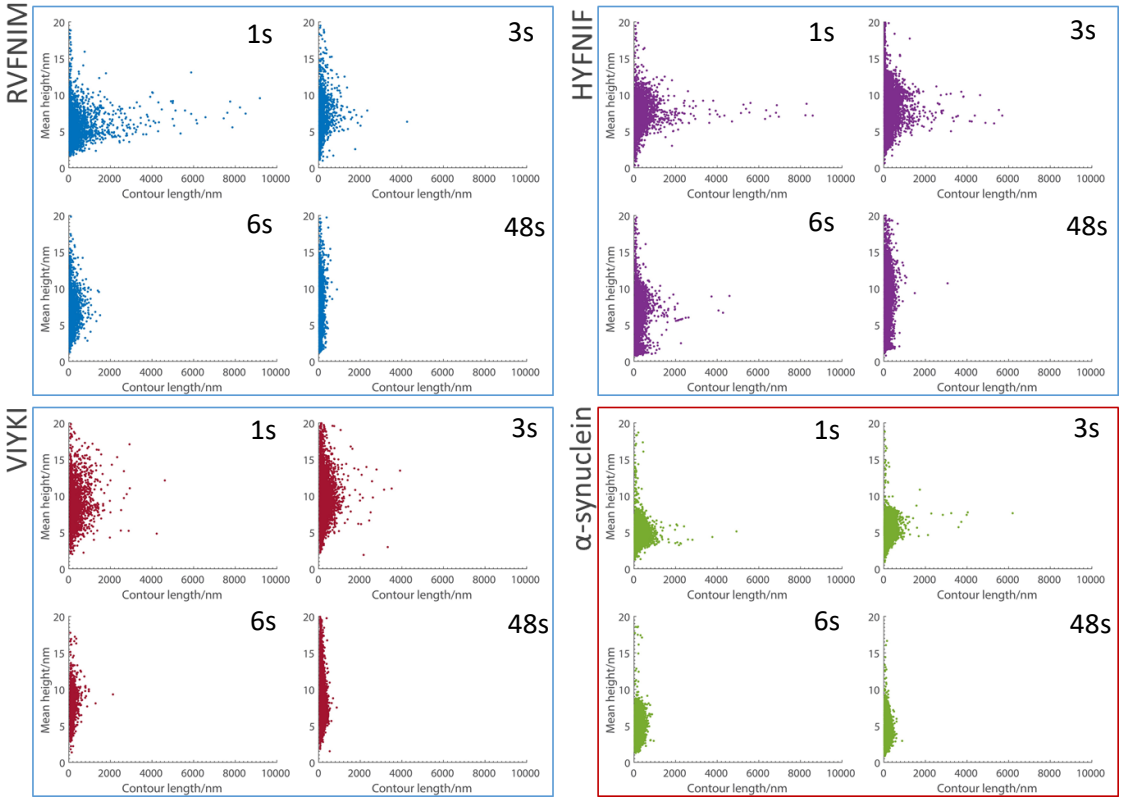


Figure 34. Observing the height dependent reduction in L_c

Plotting the per particle height with particle length allows for the identification of morphology dependent trends within the fragmentation data. Generally, amongst the distribution plots a reduction in length can be observed as well as a shift in height. α – synuclein displays a maintenance of height distribution as previously seen in Figure 33. However, for the WALTZ peptides while both a shift in height distribution and length can be observed; a strict relationship between height and length is difficult to identify. This is because as seen in the plots in this figure, at a set height fibrils display a range of lengths and do not all present the same height dependent length at a set time point. However, the longest fibrils seen in each time point for the respective populations display heights within a relatively tight range as fragmentation progresses. The height range which displays particles with a larger population of longer fibrils within the distribution is 5-10nm suggesting that the 2 ends of the morphological scale may present the lowest fragmentation stability. Distribution plots were constructed in Matlab, and represent the 1s, 3s, 6s and 48s timepoints.

This can be seen the clearest for VIYKI in which the 1s and 3s timepoints show a wide distribution of both height and corresponding particle length; then the subsequent 6s timepoint shows a narrowing in the distribution of particles, with the longest fibrils being seen at around 9nm. The final 48s timepoint shows a general recession of all height classes to a similar length scale, with a slight bulge in the distribution between 5 and 10nm suggesting fibrils within this height range maintain the highest contour length. As seen in Figure 33 the initial timepoint of RVFNIM has an average height of 5.6 nm, Figure 34 also shows a distribution supporting this height however contradicts previous observations. While it seemed that this initial average height was indicative of an increase in population of fibrils within this height range, Figure 34 suggests that in-fact fibrils within this height class display fibrils of a contour length equal to other thicker height classes. Although there is still a large population of short fibrils at this height range. Generally, for the 1s and 3s RVFNIM timepoints the fibrils with the longest contour length are between 5 and 10nm. At 6s a similar situation is present, with the longest fibrils still within 5 and 10nm and both long and short fibrils at multiple height classes within this range. The final 48s timepoint for RVFNIM however, shows a relatively uniform recession of all height classes to a similar contour length. This suggests a class independent relationship towards fragmentation stability at this timepoint, in which the exposure of sonication is sufficient to equally fragment all height classes regardless of mechanical differences. HYFNIF initially displays fibrils with the longest contour length at around 7nm which is the average height of the unfragmented population. However, this initial range includes fibrils that are of course above and below this height with fibrils of 6nm and 8nm at the same contour length. At 3s figure 34 displays fibrils at equivalent extended lengths ranging from 6nm up to 10nm. Therefore, also supporting the idea that fibrils appear to fragment mechanically

independent of their height class, to a certain extent. At 6s the distribution displays fibrils extending to around 2um at multiple height classes within the same 5-10nm boundary. Although at this time point there is a potential indication of class dependence due to the presence of a small cluster of fibrils around 4µm in length, which range in height from 6-8nm; a tighter range than much of the population. However, contrary to this apparent class dependent increase in fragmentation stability, at 48s (figure 34) the fibril population displays fibrils at around 10nm as having the longest contour length. Generally, the fibril population has receded to a relatively equal contour length regardless of height class. Overall, the height range which displays particles with a larger population of longer fibrils within the distribution is 5-10nm suggesting that the 2 ends of the morphological scale may present the lowest fragmentation stability.

Finally, in comparison to the polymorphic populations of the WALTZ peptides, α-synuclein displays a similar height linked contour length distribution despite having a lower level of morphological difference. While α-synuclein displays a tighter height distribution and subsequent level of polymorphism compared to the waltz peptides, there is still variation in height within the fibril population. The reduction in length during the fragmentation time-course is accompanied by a similar height distribution throughout. Further supporting an equivalent fragmentation of height classes. Therefore, despite the differences in polymorphism the patterns seen by the height and length distributions for both the WALTZ peptides and α-synuclein are similar eluding to a sequence dependent relationship towards the observed differences in mechanical stability. It is likely that the fluctuations in height distributions, and narrowing of the height range in fibrils with an extended contour length during fragmentation is based on the population and probability of certain height classes at an observed length. In other words, they are observations influenced by the probability of seeing a particle at

a certain height or length in which the population of each height class varies due to morphological favouring by each WALTZ sequence; and is not indicative of classes with an increased mechanical stability. This does not suggest that there is not a mechanical variability between the different morphologies adopted by the individual WALTZ peptides and it is still likely that fragmentation stability is still influenced by such differences. However, the method of fragmentation presented is not intricate enough to differentiate between possible class dependent mechanical properties. Which suggests that the differences are small in comparison to the sequence dependent mechanical properties between the WALTZ peptides and A-synuclein that have been observed here.

Mechanical properties influence fragmentation stability

The height of the fibril is indicative of the overall thickness, and it is likely that thickness and rigidity are related. The rigidity of a polymer itself can be defined by its persistence length (L_p), which is a rate measurement of the correlation decay between 2 tangent vectors along a contour length. Meaning that a very rigid polymer would maintain tangent vector correlation for an increased length than a flexible polymer. This mechanical property can be related to the physical length of a polymer to help understand the bending behaviour of said polymer. Meaning that if the physical length of a polymer exceeds the persistence length, then the mechanical agitation will cause the polymer to buckle and bend in a flexible manner. Conversely, if the physical length is below the persistence length value then the polymer will behave rigidly. These mechanical properties have been previously characterised for different fibril morphologies (Jozef Adamcik & Mezzenga, 2012)(Vandenakker et al., 2011). Additionally it has been suggested with molecular dynamics simulations that the specific

residue sequence and the type of steric zipper interface can influence and control the rigidity of amyloid fibrils (Yoon et al., 2014).

	RVFNIM	HYFNIF	VIYKI	α-synuclein
L_p/nm	14018	17675	15827	6638
SE/nm	4922	3506	2949	613

Table 13

Persistence length values for fibril populations formed from RVNIM, HYFNI, VIYKI and α-synuclein. Values were determined from datasets containing 2000-1000 individual fibrils. Bootstrap resampling was repeated 500 times to provide final average values, standard deviation (σ) and standard error of L_p (SE)

Fibrils formed from each of the WALTZ peptides α-synuclein, have been analysed using Matlab to determine the persistence length from datasets containing between 2000-10000 particles. Therefore, the values seen in table 12 are the estimated persistence length values of respective sample populations and do not consider the morphological variation. HYFNIF displays the highest L_p of 17675nm, which is comparable to that of VIYKI and RVFNIM fibrils with values of 15827nm and 14018nm, respectively. This difference is likely due to the different morphological landscapes displayed by fibrils formed from each of the WALTZ peptides. Each of the WALTZ fibril populations however, display L_p values considerably higher than α-synuclein; 6638nm. This difference is expected given the structural difference between WALTZ fibrils an α-synuclein fibrils. It has been previously shown that there is an expected relationship between persistence length and fibril height (Jozef Adamcik et al., 2010; Usov et al., 2013). Here we observe an expected relationship between fibril height and persistence length (figure 35) In that the WALTZ fibrils which display a larger average height, indeed have higher persistence length values. This suggests that as width increases, so does

rigidity. However, a discrepancy in this relationship can be observed amongst the WALTZ fibrils in that the HYFNIF population does not display the largest average height, while it does have the largest L_p. The reason for this may be due to the morphological landscape for HYFNIF, which produces more rigid fibril morphologies, despite a lower average particle height.

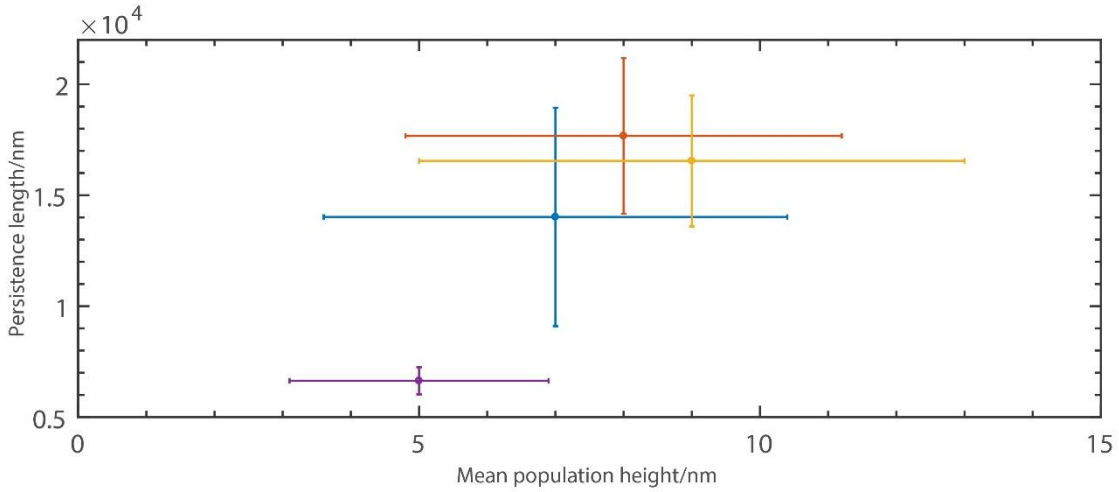


Figure 35. The relationship between height and persistence length

*Persistence length (L_p) values plotted against average height for fibril populations formed from RVNIM (blue), HYFNIF (orange), VIYKI (yellow) and α-synuclein (purple). Error bars represent standard deviation for both values. All 3 of the WALTZ fibril populations display comparatively high persistence length, which is likely attributed to fibril structures that provide an increased average height. Average height is not the only contributory factor with HYFNIF having the largest L_p value, suggesting a morphology influence. Mean L_p values were determined from resampled (*500) datasets, containing 2000-10000 individually traced fibrils (Trace-y, Matlab application)*

While height often correlates with an increase in rigidity, this is likely a parallel observation that accompanies the adoption of more complex morphologies. Changes in twisting pattern and the increased complexity of additional filaments interacting within a fibril is the probable cause of the increase in rigidity. Therefore, it is likely that morphologies with a lower average height may still present a filament arrangement that provides increased rigidity. Nanotube like structures and helical ribbons formed from the end-capped heptapeptide CH_3CONH- bAbAKLVFF-$CONH_2$ have a lower average

particle height than twisted ribbons formed from the same heptapeptide, however the nanotube likes structures display a considerably increased persistence length (Jozef Adamcik et al., 2012). Therefore, it is possible that the overall morphological landscape displayed by HYFNIF consists of structures that are more rigid despite an average height lower than VIYKI fibrils. Regardless, the multi-filament fibril morphologies adopted by all three of the WALTZ peptides are considerably more rigid than the thinner, more flexible structures adopted by α-synuclein. The relationship between persistence length and height can be defined through the relationship between the Young's modulus of a polymer and its persistence length. This relationship is defined by equation 6.

$$L_p = \frac{EI}{k_B T}$$

Equation 6.

Relationship between Persistence length (L_p) and Young's modulus (E). L_p of a polymer is scalable with the cross-sectional moment of inertia (I). section (for rigid rods of diameter R, I = 1/4πR), κB=Boltzmann constant and T = temperature (K).

For fibrils that consist of multiple filaments, cross section moment of inertia *I* is dependent on how the individual filaments are arranged within the fibril. Therefore, differences in the filament packing mechanism and the size of the fibril cross section itself are influencing factors on a fibrils mechanical properties. While the rigidity of a polymer is determined by the persistence length, the elasticity is defined by the Young's modulus. Both these mechanical properties are affected by the morphology where the cross-sectional moment of inertia *I* affects a polymers persistence length (equation 6) and the Young's modulus (equation 7)

$$E = \left(\frac{L_P}{I}\right) k_B T$$

Equation 7.

Relationship between Persistence length (L_p) and Young's modulus (E). L_p of a polymer is scalable with the cross-sectional moment of inertia (I). section (for rigid rods of diameter R, I = 1/4πR), κB=Boltzmann constant and T = temperature (K).

This relationship has been explored previously (Jozef Adamcik et al., 2010) and it was suggested that the relationship between fibril height and persistence length does not necessarily conform to the relationship defined by equation 9. Indeed, generally fibrils with increased height do display increased L_p, as seen in figure 35. However, this relationship is also likely dependent on twisting patterns and filament arrangement, which explains why HYFNIF displays a higher L_p regardless of reduced height from VIYKI. The two main filament arrangements to consider here are those in a twisted ribbon and those in a close packed twisted rope. The relationship of the former can be defined by equation 8.

$$L_P = \frac{n r_0^4 E}{k_B T}$$

Equation 8.

For a multifilament fibril, (I) is defined by the accumulated value of constituent filaments. Bending of Fibrils with a ribbon morphology occurs through the value of (I) with the lowest possible energy requirement (k_BT). Lateral assembly of filaments mean that the lowest value of I can be determined by the number of filaments (n) and the radius of an individual filament (r).

For this equation, the persistence length can be determined for a fibril consisting of *n* filaments and r is the radius of a single filament. For this twisted ribbon conformation, two cross-sectional moments of inertia are present. Due to this asymmetric geometry, this relationship (equation 8) considers a force exertion on the smallest cross section

which reflects the lowest energy bending mode. Therefore, this remains the same with the addition of a new filament and the L_p scaling behaviour becomes linear through the addition of a filament (Jozef Adamcik et al., 2010). Alternatively, when a fibril presents a close-packed filament arrangement such as with a twisted rope, the scaling behaviour is considerably increased as the cross-sectional moment of inertia *I* scales in addition to the number of filaments *n.* This relationship can be defined by equation 9.

$$L_P = \frac{n^2 E A_0^2}{k_B T}$$

Equation 9.

For a multifilament fibril, (I) is defined by the accumulated value of constituent filaments. Bending of Fibrils with asymmetrical morphologies occurs through a value of (I) which scales quadratically with a fibril surface area (A). Asymmetrical assembly of filaments means that the energy required(k_BT) to bend a fibril is greater than a ribbon morphology with identical filament number (n).

In this equation, A represents the cylindrical cross sectional area comprised of n filaments with a radius of r. In this relationship, each additional filament increases the value of both *A* and *n* leading to an enhanced scaling behaviour. This behaviour is supported through observations of increased persistence length in morphologies such as helical ribbons compared to twisted ribbons (Jozef Adamcik et al., 2012) as when a force is applied to a polymer, bending will occur through the structural area that requires minimum energy. Fibrils which do not display considerably periodicity, observed through height fluctuations across its axis, have likely adopted a tight packed filament arrangement. Said fibrils will therefore offer the same cross sectional moment of inertia *I* across the fibril axis. The reasons discussed, could explain why HYFNIF fibrils display an increased persistence length, despite displaying a lower average height. Distribution of morphologies within a large population likely favours those with minimal

periodicity, such as twisted ropes or helical structures. Indeed, results from chapter 3 do suggest prominent fibril structures for HYFNIF that are either helical or have a close-packing interface. However, VIYKI fibrils also adopt these types of structures most commonly, with the additional presence of nanotube-like structures. The difference in persistence length between HYFNIF and VIYKI fibrils could however be considered as minimal, relative to a comparison with α-synuclein fibrils. RVFNIM fibrils, which display a higher probability of adopting a classical twisted ribbon structure, in fact have the lowest persistence length value out of the three WALTZ peptides. These observations are in support of both a height and morphology relationship towards persistence length. An additional reason for RVFNIM fibrils displaying the lowest L_p value, out of the WALTZ peptides, is possibly due to differences in the core β-sheet packing interface. RVFNIM adopts an anti-parallel interface, as does α-synuclein. Previously it has been suggested that the specific sequence and direction of the steric zipper interface can be manipulated to enhance the bending rigidity of hIAPP fibrils (Yoon et al., 2014).

While the Persistence Length Is an important mechanical property, the mechanical behaviour of a polymer is dynamic and affected by particle length. The persistence length is a theoretical value which if exceeded by the physical polymer length, results in flexible behaviour. Therefore, polymers that display a contour length below the L_p can behave more rigid. Rigidity of a polymer is a property that can enhance its resistance to breakage, but in a length dependent manner. The fragmentation data presented here (Figure 31) displays multiple patterns that suggest a length dependence on the rate of fragmentation. There is no linear reduction in length during fragmentation, and relative plateaus are displayed by all fibril samples once they have reduced considerably in length. This suggests that as a fibril becomes shorter, it becomes more resistant to fragmentation. As previously mentioned, this could be due to a reduction in the number

of possible breakage points along a fibril axis as the particles become shorter. However, it may also be due to the dynamic behaviour of a polymer resulting from an L_c: L_p ratio. The suggested relationship is that when a polymer displays rigid behaviour i.e. Contour length is below the L_p, the larger the difference between these two values, the more resistant to breakage the polymer will be. This differs from a simple size dependent relationship, in that two polymers of the same length but different L_p values will display different fragmentation stability. Therefore, L_p alone is not sufficient to determine how a polymer will resist fragmentation unless determination is from a contour length independent experiment. Of course, contour length of a fibril is a difficult characteristic to control, unless using specialist conditions (Kumar, Tsumatori, Yuasa, Kawai, & Nakashima, 2015). Therefore, in most instances studying the structure or fragmentation of amyloid, length is not controlled and is likely affecting a fibrils mechanical stability in relation to its defined persistence length. To identify this relationship within the fragmentation data, it was split into 3 sections for analysis. The initial fragmentation from 0-1 seconds, the fragmentation full course 1-48 seconds, and fragmentation following the stage in which all samples display lengths within a similar range (400-500nm). Table 14 outlines these sections. For the initial fragmentation, the lengths were determined for unfragmented fibrils using the Matlab application above. These 0s values were compared with the 1s values (contour lengths determined using the same method) to determine how much each respective fibril sample had reduced in length by. This considers the initial fibril length, which varies amongst the samples; VIYKI= 7920nm, HYFNIF=5853nm, RVFNIM =2956nm, α-synuclein = 2964nm.

	RVFNIM	HYFNIF	VIYKI	α-synuclein
L_P/nm	14018	**17675**	15827	6638
L_C(0s)/nm	2956	5853	7920	2964
L_C (0s): L_P	**21%**	33%	50%	44%
Initial % L_C reduction (0-1s)	**14%**	70%	91%	87%
L_C (1s)/nm	**2556**	1756	688	398
L_C (1s): Lp	18%	10%	**5%**	**5%**
1-48s %L_C reduction	95%	88%	77%	**68%**
L_C (48s)/nm	135	**219**	161	127
Time/s: 50% reduction in L_C	8	**42**	10	5

Table 14

Table containing persistence length values for all sample populations. Additionally, values for respective contour length (L_C), (L_C: Lp %), (% reduction in L_C) and time taken to reach 50% L_C are detailed; providing information towards a L_C: Lp relationship. The 3 fragmentation periods of interest are: initial 0-1s, 1-48s full time course and fragmentation from an initial length of 400-500nm.

Table 14 displays the percentage length reduction during the initial fragmentation period. RVFNIM fibrils displayed the lowest reduction in length, only reducing by 14% which is considerably less than HYFNIF and VIYKI which reduced their length by 70% and 91%, respectively. α-synuclein, initially displayed an average contour length similar to RVFNIM however, after 1s of fragmentation had reduced in length by 87%. This is likely due to RVFNIM having a much larger persistence length than α-synuclein, at the same contour length. However, out of the WALTZ peptides RVFNIM has the lowest persistence length, meaning that higher persistence length alone does not account for fragmentation behaviour. The CL: L_p ratio in table 14 shows that RVFNIM fibrils have an average contour length that is only 21% of the L_p value. This is the largest CL: L_p ratio out of all samples, suggesting that RVFNIM fibrils are initially the most resistant towards

fragmentation. This is likely attributed to the displayed contour length conveying maximum stability towards breakage at the given persistence length. The closer a polymers contour length is to its persistence length, the more likely it is to break following mechanical perturbation. Indeed, a higher persistence length suggests a higher rigidity, however a polymer that is long enough will buckle regardless of rigidity. The buckling of a rigid polymer (i.e. one with a L_p higher than any observed contour length) will result in breakage. VIYKI fibrils display the largest initial contour length, and then experience the largest initial reduction in size. Therefore, it is likely that VIYKI fibrils at this initial length and the given persistence length, are long enough to buckle more than the other samples during fragmentation. The next observed period of fragmentation is between 1-48 seconds on Figure 31. At this point all species fibrils have been fragmented for 1 second only, however the initial reductions in contour length varied. Initial resistance, meant that RVFNIM at 1 second had the longest fibril population. Despite the largest reduction in length, VIYKI fibrils were not the shortest at 688nm; α-synuclein fibrils were, at 398nm. The CL: L_p ratio for each sample has also changed at this section (figure 36), VIYKI and α-synuclein fibrils which are the shortest two fibril populations are at 5% of the persistence length. RVFNIM and HYFNIF are at 18% and 10%, respectively. During the 1-48s period, reduction in contour length is consistent with CL: L_p ratio. RVFNIM, which had a contour length closest to the persistence length, reduced in length by 95%.

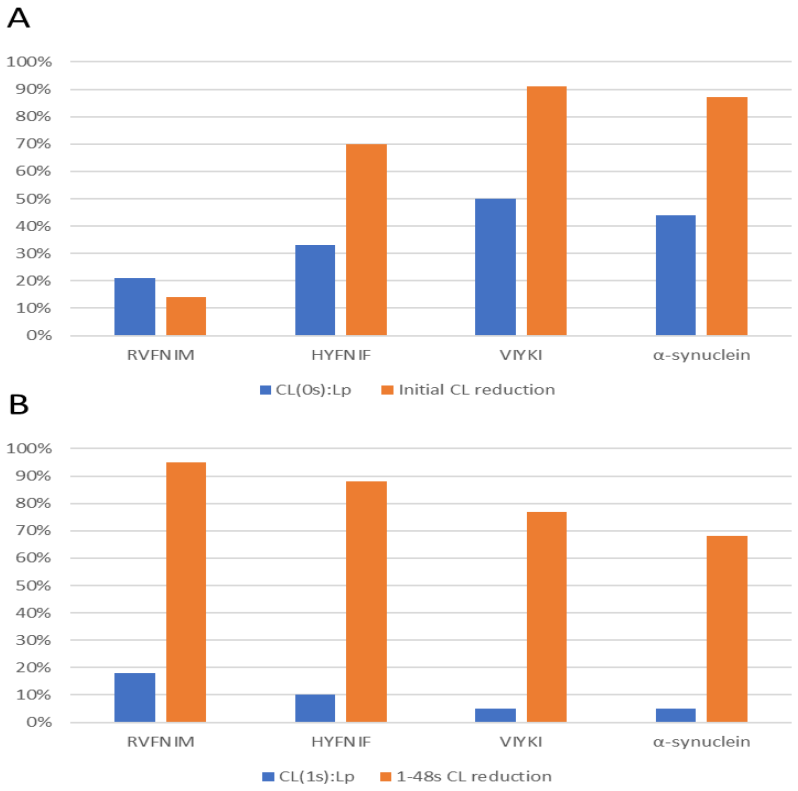


Figure 36. The relationship between contour length and persistence length

The L_C: L_p is visualised in these graphs. Generally, the higher the L_C: L_p difference (represented by L_C % of L_p) the less the fibril population fragments. This is true at both the initial 0-1s period (A) and the 1-48s fragmentation time course (B). This suggests a L_C influence on fragmentation behaviour, in which the L_C: L_p relationship may affect likelihood of fibril buckling upon a force exertion. Initially RVFNIM fibrils are relatively short (compared to HYFNIF and VIYKI) but has a high persistence length (compared to α-synuclein), resulting in minimal length reduction. α-synuclein displays the lowest reduction in length during the 1-48s time course, further supporting a L_C: L_p relationship.

While α-synuclein reduced by 68%. Indeed, both VIYKI and α-synuclein had the same L_C: L_p ratio, however the shorter overall length of α-synuclein at the same ratio may have allowed for increased resistance. An explanation for this may be in previously discussed reasons, such as the number of possible points of breakage along a fibril axis is reduced with length. VIYKI fibrils were however the next most resistant in terms of percentage reduction in fibril length (77%).

Fibril length is possibly limited by mechanical behaviour. I.e flexible fibrils with relatively low L_p will generally be able to reach lengths greater than rigid fibrils with higher L_p values before breaking. However, this is a complex relationship and the variation in both observed fibril length and L_p is considerable across amyloid species, and a fibril with high enough L_p will likely display a rigidness that allows relatively high contour length. It has been suggested that as amyloid fibrils become longer, they become more rigid and therefore more brittle and prone to fracture through breakage of longitudinal hydrogen bonds(Fitzpatrick et al., 2015). The reason for increased susceptibility to fibril breakage is due to an increased rigidity through a reduction in intersheet and interprotofilament sliding. Conversely, when a fibril reduces in length the intersheet and interprotofilament sliding increases, known as shearing. This allows for bending of the fibril to occur, which is accommodated by the longitudinal elasticity of hydrogen bonds. Therefore, it has been suggested that as a fibril reduces in length it becomes more resistant to breakage due to a reduction in stiffness(Fitzpatrick et al., 2015).

Of course, an important aspect of the fragmentation experiment is the length at the final time point. HYFNIF displays the longest final length and was intermediate regarding L_C: L_p ratio and percentage reduction in length. With the highest persistence length value and the longest final length, it is highly suggestive that HYFNIF fibrils display the most stability towards fragmentation due to an enhanced rigidity. However, due to the suggested role of contour length towards fragmentation we looked at the reduction in length following a point in which all sample populations were at similar average lengths. The initial range was 400-500nm and the time taken to reach ~50% contour length was determined. This was achieved by post-analysis of the fragmentation data presented in Figure 31. Therefore, the initial contour length was at different timepoints for each sample: HYFNIF=499nm; 6s, RVFNIM=490nm; 4s, VIYKI=428nm; 3s, α-synuclein=398nm;

1s. By looking at the time in which fibrils take to reduce in length from a similar starting point, determination of stability becomes independent from initial length. In addition to the L_C: L_p relationship the number of likely breakage points along a fibril axis, it will require more time to fragment a polymer with an increased initial length assuming mechanical properties are identical. Therefore, to identify the relationship between a variable such as mechanical properties (L_p) or structural height and the stability towards fragmentation, it is more conclusive to analyse with an independence from contour length. Of course, we assume that contour length between samples is similar, and not identical. In this scenario (Figure 31), HYFNIF indeed displayed the largest resistance towards fragmentation taking 42 seconds to reach ~50% of 499nm. This is considerably longer than all other samples. VIYKI and RVFNIM fibrils took 10s and 8s, respectively and α-synuclein took 5s. This trend in time taken to reach ~50% contour length agrees with the respective persistence length values.

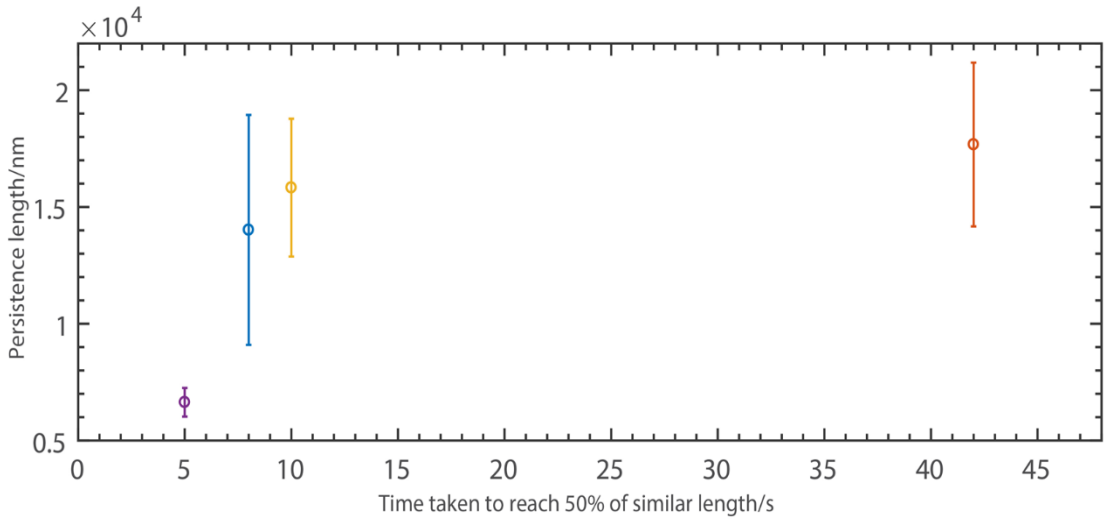


Figure 37. How persistence length affects reduction in length

To determine the influence of persistence length (L_p) on fibril fragmentation independent of variable L_C: L_p relationships, fragmentation from similar lengths was analysed. The time taken for the average fibril length to reduce by 50% was determined from populations of fibrils at an average initial length of 400-500nm. Time taken is plotted against average L_p; RVFNIM (blue), HYFNIF (orange), VIYKI (yellow), α-synuclein (purple). These times were determined post-fragmentation from current data. At similar initial lengths, HYFNIF fibrils take considerably longer to reach 50% contour length (L_C). HYFNIF fibrils have the highest L_p value, suggesting enhanced rigidity provides increased stability towards breakage amongst fibrils of a similar length (table 14).

HYFNIF fibrils are more rigid than α-synuclein fibrils, and take considerably more time to fragment. VIYKI and RVFNIM fibril were comparatively more similar to α-synuclein than HYFNIF despite closer similarities in structure and mechanical properties. This could be due to the prevalence of morphologies within the HYFNIF fibril population that are more resistant to fragmentation, such as a higher number of rope-like or helical structures. Alternatively, the core sequence and β-sheet interface could determine an amyloid fibrils stability towards fragmentation. Indeed, HYFNIF has been modelled as having an interface considerably more offset than both VIYKI and RVRNIM (Morris et al., 2013). HYFNIF protofilaments consist of 2 antiparallel β-sheets which are offset across the long axis, almost 2-fold in comparison with VIYKI and RVFNIM β-sheet interfaces. This results in a single cell unit, (monomer containing 2 interfaced β-strands) [HYFNIF-HYFNIF]) that is ~4nm wide; compared to VIYKI and RVFNIM that are each ~2nm. The interaction between 2 protofilaments is therefore occurring across a wider stretch of residues, as each β-strand may be conformationally available to interact with an adjacent strand. The comparison of this is with RVFNIM and VIYKI, where each protofilament unit can either interact end to end with an adjacent protofilament (which recruits only the terminal residues from each strand). Or they may interact laterally which increases the number of residues interacting from a β-strand, however only one can interact. Therefore, the net interacting residues at the core interface of HYFNIF fibrils is likely higher than that of RVFNIM and VIYKI. This increased core interaction, may provide enhanced stability towards fragmentation, despite mature populations of relatively similar morphologies. As already mentioned, when a fibril reduces in length (<100nm) there is an increase in intersheet and interprotofilament sliding which enhances stability through a reduced bending rigidity (Fitzpatrick et al., 2015). HYFNIF fibrils which are below this proposed threshold therefore may become more flexible

through intersheet and interprotofilament sliding, allowing them to deform through shear. With an enhanced core interface, this may provide the increased stability compare to VIYKI and RVFNIM.

Conclusion

The relationship between fibril structure and stability towards fragmentation is both complex and dynamic. Mechanically, characteristics such as persistence length help understand resistance towards fragmentation. Generally, it seems that rigid fibrils present an enhanced stability towards fragmentation. However, with a defined persistence length, fibril stability is likely influenced by the contour length. It seems that the closer a fibrils contour length is to the persistence length, the more likely the polymer is to buckle. Persistence length and Young's modulus of a polymer are scalable to one another, meaning that rigid polymers are likely relatively inelastic. Buckling of a rigid polymer results in breakage, probably at the longitudinal hydrogen bonds which do not provide the elasticity needed to bend. Longer fibrils are considered more brittle due to this reason, in addition to a reduction in interprotofilament and intersheet sliding (Fitzpatrick et al., 2015). However, at similar lengths fibrils with high L_p values are less likely to buckle and subsequently break. While this relationship can be defined using the CL:L_p values, structurally it is less clear as to how stability is enhanced. Fibril height and L_p generally increases with a strong correlation, meaning that thicker fibrils are more rigid. Additionally, fibril populations with similar polymorphism do not adhere to this relationship. HYFNIF displays the highest L_p value as a population but does not have the highest average height. Therefore, specific fibril arrangement is likely influential towards rigidity. The data presented here eludes to a morphology shift during fragmentation, meaning that different structures within the WALTZ fibril populations are fragmenting at different times. This can explain the slight differences seen between

WALTZ fibrils, wherein HYFNIF likely adopts rigid structures with a high cross-sectional area such as helical or rope-like fibrils. However, it is unclear within this work as to exactly which morphologies are the most resistant towards fragmentation. Additional work would be required to determine such. The relationship between fibril size and cytotoxicity (W. F. Xue, Hellewell, et al., 2009) means that the stability towards fragmentation is an important factor to consider in disease progression. This is because fibril structures less resistant to breakage will generate a higher number of shorter particles than those more resistant. Alternatively, regarding the potential lateral association of fibrils (Monsellier et al., 2016), rigidity of fibrils would affect the ability of the fibrils to effectively buckle and interact with a cell membranes. Meaning that short rigid particles would provide more membrane interactions as the equivalent monomer mass of unfragmented fibrils of high rigidity.

Ultimately however, the larger and more rigid morphologies adopted by the WALTZ fibril populations present an enhanced stability towards fragmentation, compared to α-synuclein. While sequence certainly influences adopted structures, it is probable that fragmentation stability is also influenced by the sequence-dependent, core structural interface. Subsequent work would be required to determine the exact relationship, such as stability enhancing residues/number of residues or β-sheet interface conformations.

The relationship between fibril length, L_p and stability towards breakage, presented here can provide valuable insight into the likely breakage stability of clinically relevant amyloid fibrils. α-synuclein, an amyloid involved in Parkinson's disease presents an overall lower resistance to fragmentation than the fibrils formed from all three WALTZ peptides. Understanding the core structural motifs or morphologies that grant WALTZ fibrils enhanced rigidity and fragmentation stability will likely aid in the furtherment of what we understand about amyloid associated cytotoxicity.

Chapter 5

Structural comparison of Sup35NM fibrils formed *in vitro* and on *E. coli* cells

Introduction

The term functional amyloid is often used to describe amyloid that is not involved in disease; instead displaying functional activity that is beneficial to a certain organism. It is still a concept of much debate, in that what defines an amyloid as functional? Given the core β-sheet rich structure common to all amyloid and the potential cytotoxicity associated with amyloid, the likely applications of 'functional' amyloid may be limited with the requirement of high levels of structural and/or assembly control. Regardless, the inherent mechanical properties of amyloid are what make its utilisation in various nanotechnology applications of keen interest. Amyloid fibrils can be seen as flexible but mechanically stable polymers comparable to spider silk (Askarieh et al., 2011). They are highly resistant to proteolysis and possess the ability to self-assemble and self-propagate independently (Makin et al., 2005). Functional amyloid can be observed in nature in various organisms, for example *Saccharomyces cerevisiae* which display the [*PSI+*] phenotype, accumulate misfolded Sup35 prion protein. Another example is the yeast protein Ure2p which propagates the [*URE3*] prion state. A protein of *Saccharomyces cerevisiae;* Ure2p regulates nitrogen catabolism through the interactive function of transcription factor Gln3p GATA (Wickner, 1994).

Curli fibrils and their function

One of the most well-known functional amyloid systems is the formation of curli fibrils by *E. coli.* The biogenesis and role of curli fibrils has been extensively researched, which has resulted in the determination of its role in biofilm formation (Barnhart & Chapman, 2006). Specifically, Curli fibrils act as the mechanical framework for biofilm which aids in the colonisation and continued survival in various environments. Aside from the curli fibers, the bacterial communities are encompassed by a complex network of proteins, nucleic acids and polysaccharides which provide protection from various chemical and

physical stresses (Friedman, Kolter, & Branda, 2005). The components of the biofilm can be culminated by the term 'extracellular polymeric substances' (EPS). Initially this term was used to denote extracellular polysaccharides, which are known to be important in biofilm production such as with cellulose. Additionally, extra cellular polysaccharides have been linked to virulence through their role in bacterial capsules(Friedman et al., 2005). However, upon expansion of the known contents of bacterial biofilms the term EPS was adjusted to define all polymeric species including DNA species and of course, protein structures such as curli fibrils. Ultimately, the complex EPX matrix present in biofilms is vital for the successful nature of many bacteria communities, by providing various physical and enzymatic factors that aid in the colonisation of harsh, niche environments. The Presence of biofilms within areas prone to colonisation present an increased health risk, as biofilms can enable prolonged infections with an enhanced resistance to both host immune responses and treatment with bactericidal treatments (Hall-stoodley & Stoodley, 2009). The matrix of curli fibrils within biofilm aids the expansion into pellicles, a biofilm floating at the liquid-air interface of liquid cultures; displaying its mechanically suited role as a supportive structure outside the suspension of a liquid culture.

Curli biogenesis is a multistage mechanism conducted by a group of proteins and protein complexes which are encoded by two curli operons csgBAC and csgDEFG. Each operon contains multiple genes which encode a respective member of this coordinated system. Until recently the first of the two operons was considered to be csgAB and was known to encode the proteins csgA and csgB which are well defined in their function of the curli system (Chapman et al., 2010). However, relatively recent findings have expanded the operon to incorporate csgC as the function of the subsequent protein has been characterised after years of elusiveness (Evans et al., 2015). Mutant strains of *E. coli*

which are deficient in CsgF and G (two protein complexes involved in curli biogenies) subsequently do not produce curli fibrils and therefore do not form pellicle biofilms. This inability can be rescued upon the addition of a CsgFG containing plasmid (H et al., 2010). In addition to this, the fibrils aid in the adhesion and colonisation of otherwise difficult surfaces, such as stainless steel (Ryu & Beuchat, 2005). Additionally, it was shown that *E. coli* display enhanced resistance to chlorine which is dependent on the maturity and growth of biofilm (Ryu & Beuchat, 2005). CsgA is the primary polymerising component of curli fibrils, and mutant strains deficient in this display a reduced attachment efficiency to various abiotic surfaces such as Teflon (Taran, Gentle, Gobius, & Dykes, 2011). These curli fibrils present an example of how amyloid can be utilised by nature to be beneficial to an organism. Despite the potentially cytotoxic nature of these fibrils, *E. coli* takes advantage of amyloids inherently suited mechanical properties and ability to self-assemble to aid in its colonisation and prolonged survival. The primary component of the curli fibrils themselves is the polymerising csgA subunit. CsgA forms the core structural basis for the amyloid like fibrils that result from curli biogenesis. There are three domains that define the csgA sequence which include an N-terminal 22-amino acid targeting sequence that directs the protein outside of the cell via the csgG outer membrane channel complex, a C-terminal amyloid core domain and a signal peptide (Van Gerven et al., 2015). The C-terminal amyloid core domain is the region responsible for the characteristic amyloid fold by providing the required inter-residue interactions that enable this β-sheet rich conformation and subsequent thermodynamically favourable polymerisation. that CsgA is targeted towards the periplasm by its signalling sequence and via the SecYEG translocon, however once it has arrived in the periplasm the signalling sequence is proteolytically cleaved (Costa et al., 2015). Following this event, CsgA requires CsgG to be exported across the outer

membrane and requires for efficiency but is not dependent on; the nucleation by CsgB (Chapman et al., 2010). CsgG is vital for curli biogenesis as it forms a channel complex that provides a means for csgA and csgB to cross the outer membrane. Specifically, CsgG is a 262-residue lipoprotein that forms a nonameric anticodon-binding domain complex, culminating to a 36-stranded-β-barrel (Goyal et al., 2014). As already stated, this allows the diffusion of csgA and csgB into the extracellular space. CsgB, is the second component of mature curli fibrils; providing a slight discrepancy to most amyloid fibrils in that the fibrils themselves are heterogenous in regards to the polymerising monomer units. CsgB is however similar in structure to csgA; it contains a signal peptide, a targeting sequence, and an amyloid forming domain. However, its role in the formation of curli fibrils differs from CsgA in that it is not necessarily required for the formation of fibrils, however it is required for the cell surface associated elongation of fibrils. This is because CsgA is able to polymerise into fibrils through a localised concentration mechanism, resulting in fibrils that are not attached the cell surface (Dueholm et al., 2013). Additional similarities are present in the structure of the amyloid forming domain for csgA and csgB, which consist of 5 repeating units. However, from this arises a difference within the 5^{th} repeating unit as for csgB this unit contains four charged amino acids which are likely responsible for the fibril association to the cell surface (Hammer, Schmidt, & Chapman, 2007). This R5 dependent cell surface nucleation is thought to be attributed to the interaction between csgB and an additional curli factor, csgF. CsgF is an accessory protein that is bound to the csgG transmembrane protein complex (Van Gerven et al., 2015). This protein is thought to be responsible for chaperone like behaviour and the aiding in csgB nucleation, providing an important role in maintaining control of the csgA extracellular polymerisation. CsgF like most of the curli factors is transported to the periplasm via the SEC pathway. It is then directed to the csgG

transmembrane channel protein complex, where it associates extracellularly to provide an immediate anchor for csgB upon its own diffusion from the periplasm (Nenninger, Robinson, & Hultgren, 2009). An additional accessory protein within the curli machinery is the csgE curli secretion specificity factor, which has multiple functions during curli biogenesis. Its first function is its contribution to the silencing of premature polymerisation of csgA. CsgE interacts with CsgA monomers in a 1:1 ratio, and inhibits the potentially toxic premature intracellular fibril formation (Nenninger et al., 2011). In addition to this, csgE provides a specificity to csgG secretion which has been defined by the lack of specificity seen by cells overexpressing csgG and lacking csgE. In such strains, there is a non-specific efflux of periplasmic proteins however the specificity for csgA and csgB can be restored by co-overexpressing csgE (Nenninger et al., 2011). In addition, a general lack in CsgG mediated protein transport has been seen in strains lacking csgE but expressing normal levels of csgG. Specificity arises from interactions between csgE and the N22 or N23 regions of CsgA and CsgB respectively. CsgE then displays further interactions with csgG in which it forms a conformationally compatible nonameric complex and binds to csgG (Goyal et al., 2014). Therefore, CsgE acts as a mutually binding accessory protein that acts as a mediator which brings csgA and csgB towards the csgG complex. Through its nonameric binding, CsgE provides a gating mechanism to the otherwise un-gated csgG complex subsequently forming a chamber that facilitates the accumulation of the targeted proteins csgA and csgB. This accumulation provides an entropically favourable diffusion of the respective proteins from the periplasmic space to the extracellular space, adding specificity to an otherwise simple diffusion channel(Van Gerven et al., 2015). As previously mentioned, until recently the function of the third protein encoded by the csgBAC operon, csgC has been relatively unknown. Aforementioned research has identified a function of csgC, which has possibly been

masked previously by the similar polymerisation inhibiting activity of csgE. While the exact mechanism is unknown, it seems that csgC binds to and inhibits the progression of prefibrillar intermediate species that accumulate in the periplasm. This was observed in strains deleted in both csgG and csgE, to facilitate accumulation and prevent alternative inhibition, respectively. In these strains, there was no progression from oligomeric species into mature fibrils which would likely result in cell death. Subsequent deletion of csgC in these strains resulted in the polymerisation of csgA into fibrils(Evans et al., 2015). A proposed mechanism of this inhibitory effect is the ability of csgC to interfere with the transition of amorphous csgA oligomers into oligomers with a rich and ordered β-sheet structure, that may subsequently seed fibril formation. This attributes csgC with a chaperone like behaviour, albeit not a true chaperone due inability to detect stable interacting complexes of these two curli factors; suggesting a transient interaction mechanism between csgC and pools of oligomeric csgA(Van Gerven et al., 2015). Finally, the master regulator csgD is responsible for the transcriptional regulation of the csgBAC operon(Van Gerven et al., 2015).

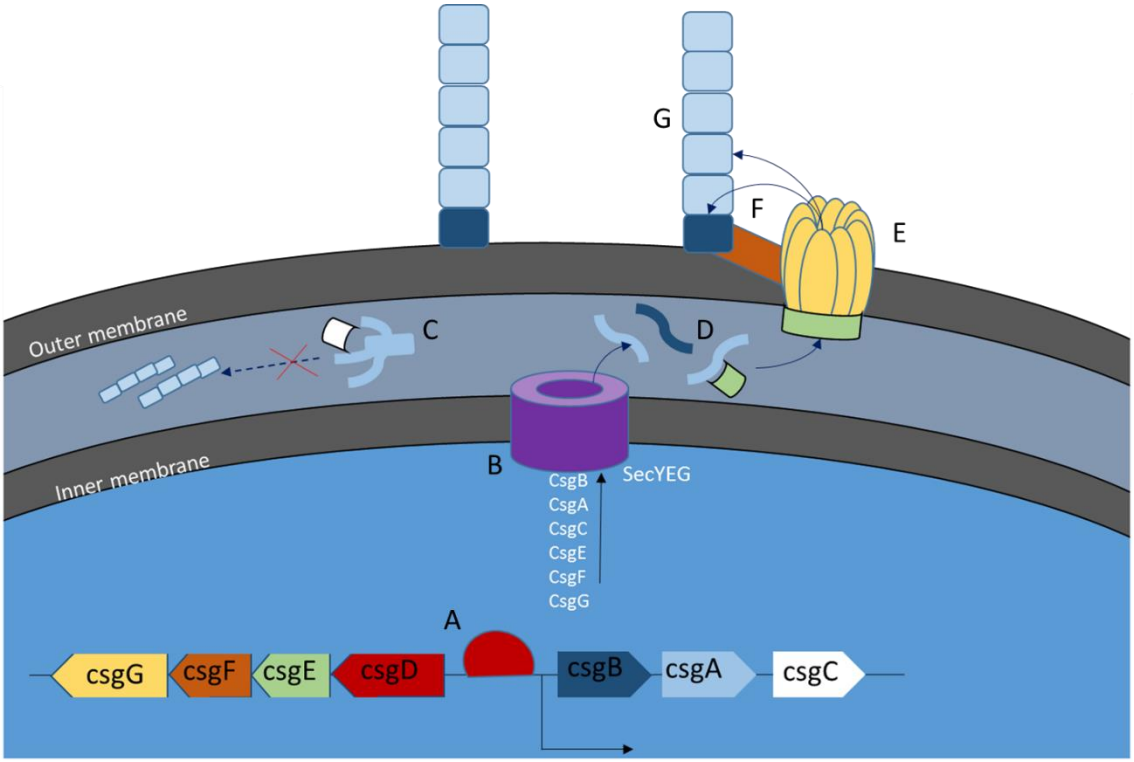


Figure 38. Cellular machinery required for curli biogenesis

Diagram detailing the machinery involved in curli biogenesis. the master regulator csgD is responsible for the transcriptional regulation of the csgBAC operon. (B) accessory proteins involved in curli biogenesis are translocated to the periplasm via the Sec pathway. (C) CsgC csgC binds to and inhibits the progression of prefibrillar intermediate species that accumulate in the periplasm. (D) CsgE interacts with CsgA monomers, and inhibits premature intracellular fibril formation. In addition to providing CsgG secretion specificity. (E) CsgG forms a channel complex that provides a means for csgA and csgB to cross the outer membrane. (F) CsgF is an accessory protein that is bound to the csgG complex. It is also thought to provide chaperone like behavior and aid in csgB nucleation. (G) CsgB is anchored to the cell surface via CsgF interaction and nucleates CsgA polymerization.

Curli dependent amyloid generator

A deep understanding of the curli biogenesis machinery has allowed for its functional manipulation into an amyloid generating system in which the amyloid core domain of CsgA is essentially swapped for potentially any amyloid. This system, developed by Viknesh Sivanathan and Ann Hochschild is known as the curli dependent amyloid generator (Sivanathan & Hochschild, 2013). It exploits the fact that the targeting system of the curli machinery is a separate sequence from the amyloid forming region of the amyloidgenic proteins involved. Coupled with the intrinsic ability for amyloid to self-polymerise, this system allows *E. coli* to facilitate the biogenesis and fibrillization of any

amyloidgenic protein. The core requirement of this system is the conjugation of an amyloidgenic protein or peptide sequence to the CsgA full signal sequence; a bipartite sequence that includes the 20-residue SEC targeting signal sequence and the 22-residue outer membrane targeting sequence, thus facilitating the targeting of exogenous proteins to the CsgG transmembrane channel protein complex (Sivanathan & Hochschild, 2012). In addition to a CsgAss coupled protein, the system requires a modified strain of *E. coli* that is deficient in csgBAC through a deletion of the csgBAC operon. As the csgG targeting sequence can be attached to any protein, the specificity of the interaction between accessory proteins and CsgAss may be retained. Therefore, in a system lacking csgBAC, accessory proteins such as CsgE may transition their function to the exogenous protein, possibly maintaining a specific and efficient secretion of the newly expressed protein. For example, CsgE would continue to gate the CsgG complex providing specificity for the csgAss. However, since the exact interactions between csgE and CsgA and CsgB is not fully understood there is a chance that some specificity is lost, and CsgE will facilitate the export of the exogenous protein from the cell as efficiently, regardless of the targeting sequence that directs the protein towards CsgG. To overcome this potential lack in efficiency or specificity for csgG transport, the strain is also modified to overexpress csgG (Sivanathan & Hochschild, 2013). This overexpression has been shown to enable the export of csgA in strains lacking the required accessory proteins such as CsgE, however there was a lack in specificity with a general diffusion of endogenous periplasmic proteins. CsgA specificity can subsequently be restored in such strains with the co-expression of CsgE (Nenninger et al., 2011). Therefore, the C-DAG system requires a ΔCsgBAC strain which has been supplemented with an overexpression of csgG. The resulting strain is one that does not produce csgA, CsgB or CsgC however overexpresses csgG to allow for an increase in protein export and a means to overcome

a potential reduction in CsgE mediated specificity that may restrict the export of exogenous proteins utilizing this curli machinery. However, as the exogenous protein is a fusion with the CsgAss the CsgE that is present likely maintains a level of specificity that will ensure a predominant efflux of target protein; not the general diffusion of periplasmic proteins, albeit not eradicating this likelihood entirely. Strains lacking CsgC may be prone to premature fibril formation, however due to the remaining presence of CsgE this is likely controlled given the chaperone like behaviour of CsgE (Nenninger et al., 2011).

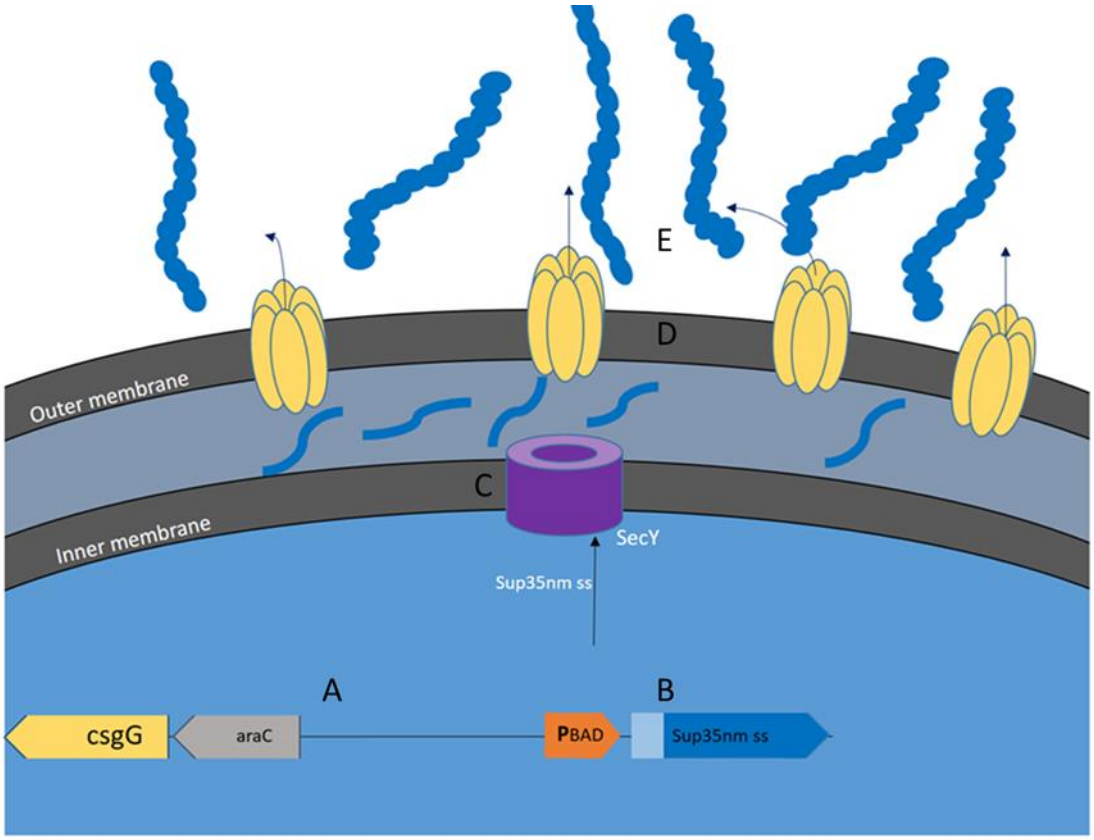


Figure 39. Cellular machinery of the C-DAG system

The curli machinery is modified for the C-DAG system. (A) The system is deficient in csgBAC through a deletion of the csgBAC operon. Also required is the arabinose induced overexpression of CsgG. (B) in place of the CsgA and CsgB proteins, is the CsgAss fusion protein; an exogenous protein of choice attached to the CsgA signaling sequence. (C) CsgG and the fusion protein are transported to the periplasm by the SEC pathway. (D) overexpression of CsgG enables the periplasmic export of the fusion protein, absent specificity providing accessory proteins. (E) Minus CsgF to anchor the nucleating monomer, localized concentration of amyloidgenic protein enables the proximal polymerization of the fusion protein outside of the cell.

The yeast prion protein Sup35

Amyloid is considered self-propagating, however when it is also transmissible and able to propagate the amyloid fold in non amyloidgenic populations of precursor material it is considered to have prion like behaviour. Simply put, Prion proteins are infectious amyloid and they are implicated in various mammalian diseases such as human Creutzfeldt – Jacob and bovine spongiform encephalopathy. While these diseases involve the transmission and aggregation of PrP, much of prion research focuses on prion proteins found I yeast which itself serves as a good model for understanding the behaviour of prion proteins. Sup35p is a translation termination and release factor in *Saccharomyces cerevisiae* and its aggregation results in the [*PSI+*] Prion phenotype. The transition into a [*PSI+*] phenotype is considered as a loss of function mutation, in which there is an increase in nonsense codon readthrough, which characteristically results in the readthrough of adenine stop codons, allowing for its production. This in turn allows for the growth of [*PSI+*] strains on adenine deficient media while [*PSI-*] strains do not grow due to efficient translation termination which produces truncated adenine. This loss of function is visualised on YPD media in which [*PSI+*] strains will grow as white colonies and [*PSI-*] strains will grow as red colonies (B. S. Cox, 1965). The observation of variation in the degree of pigmentation on YPD media (gradient of red to white) suggested the prevalence of differing strengths of the [*PSI+*] phenotype. In other words the ratio of aggregated against non-aggregated Sup35 protein is responsible for different levels of stop codon readthrough and subsequent characterisation of strong and weak [*PSI+*] variants (Uptain, Sawicki, Caughey, & Lindquist, 2001). In addition to this variation in phenotypic strength, it has been observed that each variant is stable in its expression and mating of alternate strains result in progeny the same as the parent strain which aggregates the strongest, suggesting an aggregation and propagation

dependent dominance amongst strain variants (Bradley, Edskes, Hong, Wickner, & Liebman, 2002). Considerable behavioural and structural characterisation has been invested into the validation of the core amyloid traits within prion proteins. Amyloid is known to be proteinase K resistant, and Sup35 conforms to this resistance with the prion form displaying the characteristic proteinase K resistance (Paushkin, Kushnirov, Smirnov, & Ter-Avanesyan, 1996). In addition to this, treatment of aggregate containing cell lysates with detergent at room temperature Sup35 aggregates are retained in filter trap assays because of their amyloid conforming resistance to detergent. Subsequent and additional treatment of filter trapped aggregates results in the separation into distinct fibrils which can be separated on an agarose gel and it was found that the strong prion variant displayed fibrils smaller in size than the weak variant; suggesting a possible size dependence on the propagation ability of [*PSI+*] strains (Kryndushkin, Alexandrov, Ter-Avanesyan, & Kushnirov, 2003).

The size dependence for prion propagation is supported by the well characterised Hsp104 fragmentation of Sup35 fibrils. This chaperone is considered necessary for the successful propagation of the [*PSI+*] phenotype, as mutations rendering a strain Hsp104 negative result in an inactivation of [*PSI+*] and an inability for its propagation into daughter cells (Shorter & Lindquist, 2004). Hsp104 is known to be active in the disaggregation of certain endogenous proteins, and has additionally been shown to enact similar behaviour on Sup35 which provides a rationale for its role in prion propagation. Hsp104 facilitates the fragmentation of Sup35 fibrils which creates fibrils of reduced length which firstly provide seeds for subsequent fibril formation and secondly allows for the size dependent transfer of amyloid material into daughter cells when it may be too large to be efficiently transferred at unfragmented sizes (Kushnirov & Ter-Avanesyan, 1998)(Kryndushkin et al., 2003). A proposed mechanism for Hsp104

mediated fragmentation is the monomer removal from the polymer chain via the central pore of the Hsp104 complex therefore breaking the fibril at that location (Liebman & Chernoff, 2012). This process may gradually disassemble Sup35 polymers into individual monomer. Contrary to this, strains that overexpress HSP104 also no longer present the [*PSI+*] phenotype suggesting a propagation enhancing mechanism that depends on the balance of HSP104 (Malato, Reis, Benkemoun, Sabate, & Saupe, 2007). Previously, it had been suggested that overexpression HSP104 may in fact result in fibrils of increased length (Liebman & Chernoff, 2012). However more recently a twostep disaggregation process has been suggested for HSP104 which upon overexpression results in a curing of [*PSI+*] and not just the propagation enhancing fragmentation of polymers. This twostep process involves the monomer level disaggregation reducing the sizes of the prion polymers, followed by the elimination of the amyloid core which is possibly facilitated by proteolysis on the seeds of reduced length (Y. N. Park et al., 2014). Ultimately however, the relationship between HSP104 is complex and likely not an isolated one; with additional interaction from other heat shock proteins such as HSP70 and HSP40. This creates a dynamic relationship suggesting enhanced activity from HSP104 overexpression alone is not sufficient to reliably cure [*PSI+*] (Liebman & Chernoff, 2012). The propagation enhancing activity of HSP104 can be inhibited through the treatment of Guanidine hydrochloride (GuHCL), resulting in the curing of [*PSI+*] in proliferating cells (Brian S Cox, Byrne, & Tuite, 2007). Fragmentation is inhibited and new prion units are not formed upon treatment with GuHCL; however, Sup35 aggregation is not affected and the aggregates are not destroyed. This observation is what eluded to the [*PSI+*] curing activity of GuHCL being a HSP104 targeted mechanism (Ferreira, Ness, Edwards, Cox, & Tuite, 2001). The blocking of HSP104 mediated fragmentation results in a lack of seeds for subsequent rounds of

elongation, therefore maintaining the population of fibrils within each cell. In a seedless environment, cell division leads to a reduction in population density as the original cells fibril population is serially diluted amongst the progeny until daughter cells no longer contain prion protein.

Structurally, prions share the same amyloid core as all amyloid. The regions within the protein sequence that are responsible for fibril formation and propagation are known as prion domains (PrDs). Much like many amyloid proteins, the formation of fibrils is facilitated by the amyloid core domain which maintains the ability to form fibrils when isolated from the flanking regions of the protein sequence. For Sup35p the N-terminal domain is responsible for the formation of prion particles and subsequent propagation. The C-terminal domain is not required for fibril formation however is functional in non-amyloid Sup35p through translation termination. Finally, the middle domain links the N and C domain, and contains charged residues that may have a role in the maintenance of the aggregation landscape (Luckgei et al., 2013). Neither the M or the C domain alone can form amyloid fibrils, however any combination incorporating the N domain displays the ability to form fibrils.

Prion forming domains have been characterised as having sequences rich in Glutamine (Q) and asparagine (N). Within these regions, the N residues are potentially more critical in the role of prion propagation. This arises from the observation that prion domain substitution of Q residues for N residues enhances propagation while the Converse of this reduces propagation (Halfmann et al., 2011). Additionally, the composition of the sequence i.e. the number of respective residues within a sequence and not the order of the sequence itself, is dominant in the role of prion propagation and fibril formation. This is based on the observation that scrambled sequences that maintain the same residue composition are still able to form amyloid fibrils and propagate the prion state

(Ross, Edskes, Terry, & Wickner, 2005). The N terminal domain of Sup35p Is no exception, and has high Q and N content within the first 40 residues. The QN-rich regions is considered critical for fibril formation, while the subsequent residues containing oligopeptide repeats are responsible for prion propagation and this regions extends to residue 97 of the N terminal domain (Liebman & Chernoff, 2012). In terms of structural characteristic of the resulting amyloid fibrils, Sup35p forms fibrils that are relatively polymorphic with fibrils that vary in both size and rigidity (Castro, Dong, Boyce, Lindquist, & Lang, 2011). The β-sheet rich core, is identified by a characteristic X-ray diffraction profile displaying a 4.7 Å meridional reflection representing the separation between stacked β- strands within a β-sheet (Castro et al., 2011; Tycko, 2014). The variation in structure is related to the strength of the prion phenotype, with strong and weak [*PSI+*] strains producing fibrils with different properties. In these studies, [*PSI+*] was propagated in cell lines using fibrils formed at either 4 °C or 37 °C which induced a strong and weak phenotype, respectively. The strong phenotype describes the ability of most endogenous Sup35p to aggregate upon transfection with exogenous Sup35NM and the converse for the weak phenotype. The weak phenotype was found to display fibrils that were 2-fold more rigid than the strong phenotype (Castro et al., 2011). This provides a potential explanation for the 'weak' phenotype In that the increased stability reduces that rate of fragmentation and therefore reduces the effects of size dependent propagation (Tanaka, Collins, Toyama, & Weissman, 2006). This agrees with the length based observations that show longer fibril species present within the weak [*PSI+*] phenotype. The detection of class specific fibril structure within the progeny of propagation that correspond to the original seeds (fibrils formed at 37 °C or 4 °C) suggest the templating nature of prion propagation in that subsequent cycles of fibril formation will adopt the seeded structure (Verges, Smith, Toyama, & Weissman, 2011).

The ability of Sup35p to form fibrils following N-domain scrambling, is in support of a parallel in-register β-sheet structure that is arguably the most common core structural arrangement (Ross et al., 2005). Even amongst a scrambled sequence, conformational shift into the β-fold would allow key amyloidgenic residues to still align forming an in-register β-sheet. This structure likely adopts a steric zipper interface which grants the displayed stability through axially repeating interactions between interdigitating residue. Expanding from the amyloid forming N-terminal prion domain, the M and C domains are not contained within the core β-sheet structure; instead forming globular appendages that hang from the stacked N domains. While the core fibril regions has been found to be around 8nm on average, the globular and cloud like C-domain containing region has been characterised as ~65nm which conformationally allows for the interaction of fibrils with ribosomes *in vivo* (Baxa, Keller, Cheng, Wall, & Steven, 2011). Amyloid defining characteristics such as protease resistance does not extend to this globular region, as upon protease digestion only the core N-domain remains within the prion polymers (Liebman & Chernoff, 2012). Alternatively, it has been suggested that Sup35 forms fibrils that adopt a β-helix structure similar to HET-s (Krishnan & Lindquist, 2005). This model suggests shorter core domains that polymerise that interact in a head to tail fashion where the N and C terminal regions of each monomer unit interact. Therefore, there is no repeating inter-sheet array of interactions within the core region of each monomer. Due to the detection of both structural archetypes, it is possible that there is a strain dependent variation within the structure and both β-helix and parallel β-sheet structures can be adopted by Sup35 fibrils.

C-DAG generated Sup35NM

The C-DAG system has been characterised extensively using Sup35NM fibril formation. The advantage of using Sup35 for this system is its conformational adaptability given its multi domain sequence. The N domain of Sup35, aggregates regardless of its attachment to non-amyloidgenic sequences. This displays its ability to retain its amyloid forming potential regardless of flanking amino acids which have been shown to interfere with the amyloid forming potential of certain sequences (Reumers et al., 2009). Considering the machinery of the C-DAG system requires a fusion protein containing the CsgAss, this intrinsic propensity of Sup35 to form fibrils and maintain attached regions that are not incorporated into the fibril core; makes it an ideal protein sequence to incorporate into the system. Upon export of the Sup35NM monomer from the cells, the localised concentration in proximity to the cell surface encourages the favourable polymerisation into fibrils, differing the method of fibril formation from the CsgB nucleated formation of curli fibrils. While this means that the resulting fibrils will not necessarily associate with the cells through accessory protein mediated attachment, the exogenous Sup35NM may still interact with the cells but it will be on a charge basis. This is due to the highly-charged M region of the fibrils, which has the potential to facilitate interaction with the bacterial outer membrane; therefore, enhancing the cellular support provided from a Sup35NM containing biofilm. Using AFM, we seek to qualitatively determine any morphological and structural discrepancies between Sup35NM fibrils formed *in vitro* and Sup35NM-Csg-SS fibrils formed using the C-DAG machinery. Considering the exclusion of attached non-amyloidgenic from the amyloid core in Sup35p fibrils, it seems likely that additional attached sequences would not affect the core fibril region. Our observations here support this, as despite a seemingly increased level of fibril to fibril association, the individual fibrils conform to the same

height range. Therefore, suggesting a preservation of the filament arrangement adopted by *in vitro* formed fibrils.

A key principle of the C-DAG kit, is the incorporation of any amyloid sequence into the PVS72 plasmid in place of Sup35NM. Therefore, this system can potentially manipulate the resulting biofilm, providing bacterial cultures with some interesting abilities. This highlights the potential for the C-DAG system to be used in a nanotechnology role that may aim to modify the biofilm through additional function of the constituent amyloid fibrils. To enhance the potential of this role, the ability to control the mechanical properties of the produced fibrils is important. For example, you may wish to produce fibrils that are rigid or fibrils that generally conform to a particular size, charge, or aromaticity. Synthetic peptide sequences, such as the WALTZ identified sequences RVFNIM, HYFNIF and VIYKI provide an ability to control such characteristics and here we have incorporated them into the C-DAG system to characterise their ability to form amyloid. For all applications of this system, the role of the implemented amyloid is based on the structural and functional knowledge of *in vitro* formed fibrils. However, with the addition of the signalling sequence and an alternate environment for fibril formation it is likely that fibril properties may change. Similar to understanding the possible discrepancy between *in vivo* and *in vitro* formed amyloid fibrils, it is important to determine possible issues with using the C-DAG system. This understanding is required to draw meaningful conclusions that can apply to *in vitro* and *in vivo* formed fibrils. Here we report a structural comparison between fibrils formed from Sup35NM *in vitro* and in using the C-DAG system. The resulting similarities support the application of this system to produce fibrils based on *in vitro* formed properties. However, implementation of the WALTZ peptides into the C-DAG system highlight issues with the use of short peptides.

Results and discussion

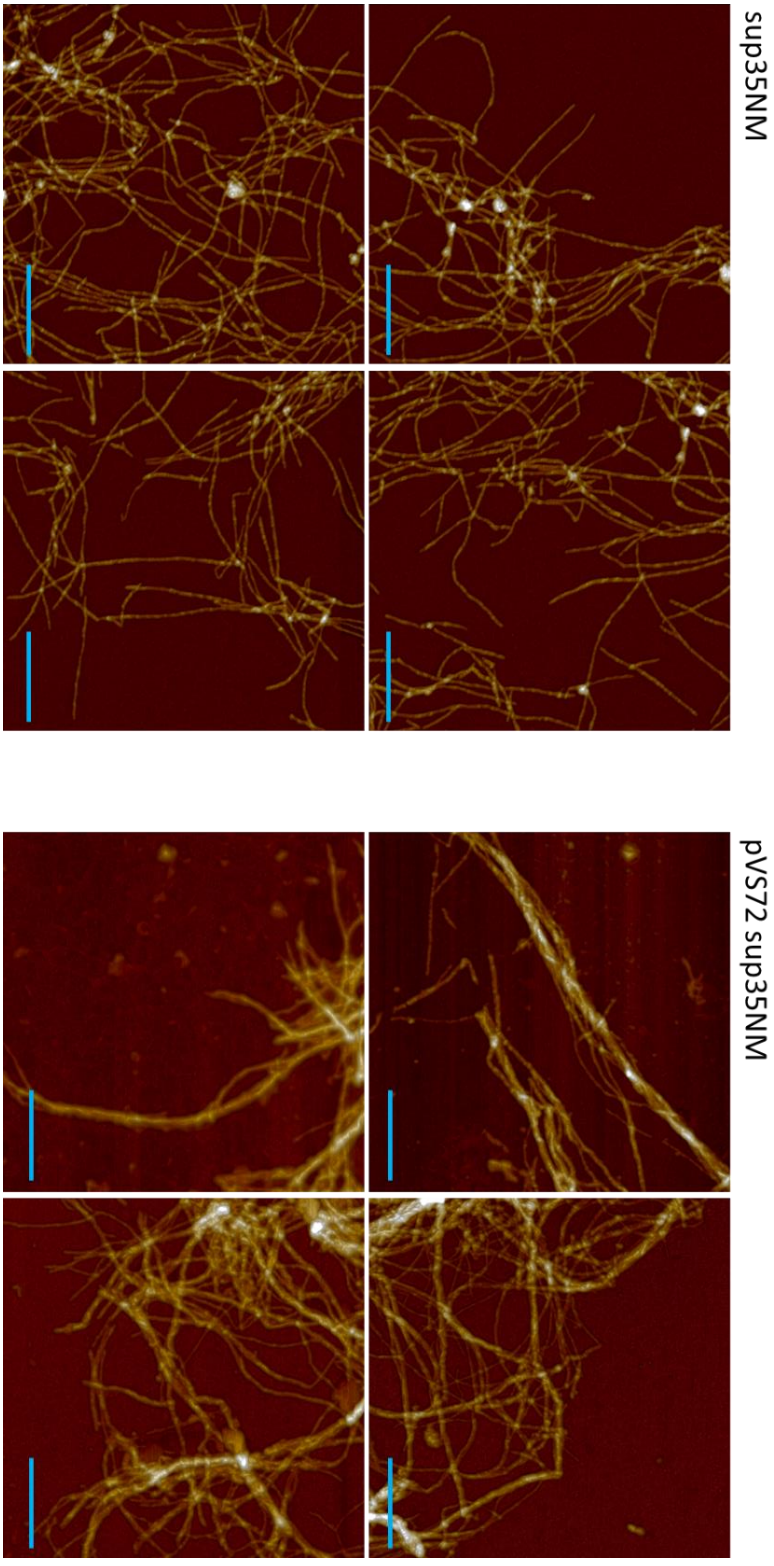
A qualitative comparison of Sup35NM fibrils

For the functional application of the C-DAG system, the resulting fibrils may be desired to display structural similarity to the same amyloid formed *in vitro.* It has been suggested that *in vitro* formed material may be structurally different from material formed *in vivo* either in a mammalian pathogenic environment or in yeast for the case of Sup35. This structural difference is in regards to the fibrils assembly morphology and has been highlighted for Aβ in that *in vitro* formed fibrils display different polymorphism to fibrils extracted from Alzheimer's patients (Annamalai et al., 2016). However, it is important to note that each population of fibrils maintains the ability to form polymorphic fibrils, which was a trait under debate in regards to its prevalence in pathogenic populations of fibrils. Additionally, Sup35 fibrils which form in either strong or weak [*PSI+*] strains are structurally distinct, suggesting the prevalence of polymorphism In yeast prion environments (Tycko, 2014).

To qualitatively compare fibrils formed *in vitro* and using C-DAG, fibril populations were prepared under the respective conditions (chapter 2). Structural characterisation was then performed on AFM generated images to determine morphology. Figure 40 displays Sup35NM fibrils which have been formed *in vitro* at 37°C as well as fibrils that have been formed using the C-DAG system. The C-DAG system, requires incubation at 22°C to allow for fibril formation in an environment that will not encourage bacterial overgrowth. Therefore, each of the fibril populations were also formed at different temperatures.

Figure 40. Comparing fibril structures of Sup35NM

Comparative AFM images for C-DAG formed Sup35NM fibrils and In vitro formed Sup35NM fibrils. Each image represents a 2 x 2 µm cropped sample area with a pixel density of 2.9 pixels/nm and the scale bar represents 500nm. While both fibril populations display entangled networks of fibrils, those formed using the C-DAG system displays large fibril structures; likely the bundling of multiple fibrils.



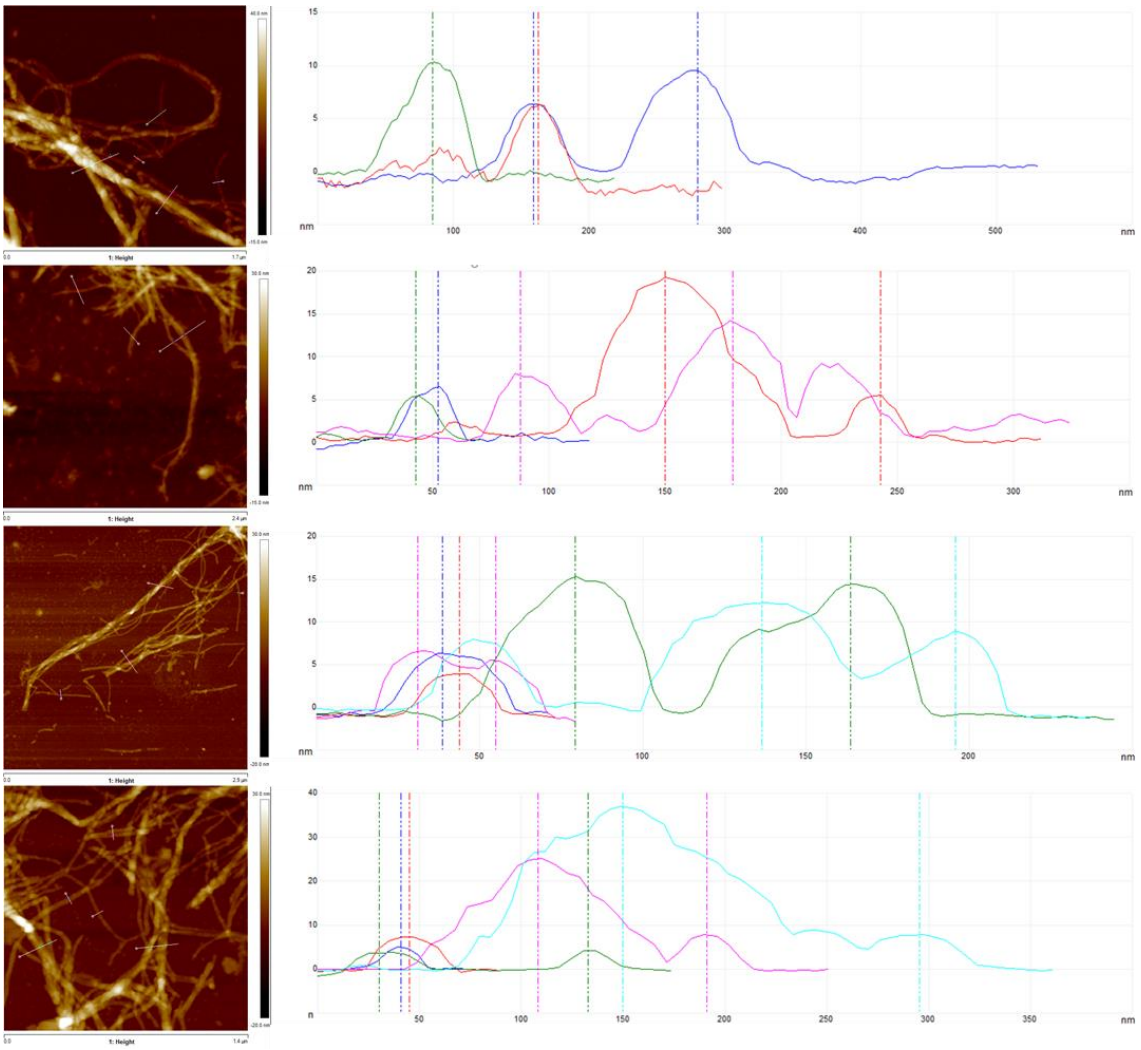


Figure 41. Height analysis of Sup35NM fibrils formed using the C-DAG system

Height analysis of Sup35NM fibrils formed using the C-DAG system, using nanoscope analysis (Bruker). Individual fibril particles are picked out and their height determined in contrast to the flat mica surface. While the fibrils tend to form large bundles structures, individual fibrils can be observed and display heights ranging from 5-10 nm with larger structures displaying heights exceeding 10nm. The observed individual fibrils appear morphologically similar across the observed population. Filament twisting within the fibrils is difficult to detect within the images.

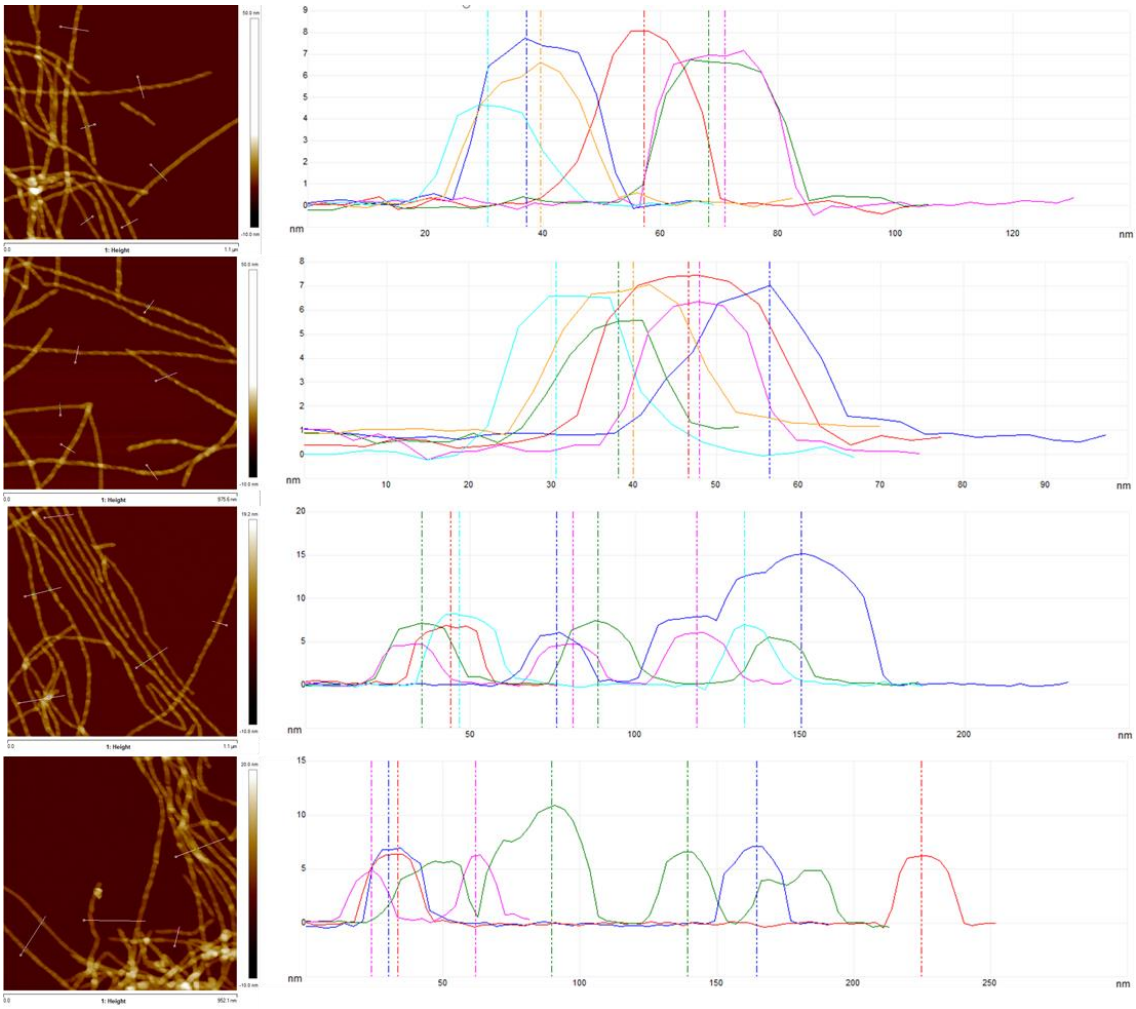


Figure 42. Height analysis of in vitro formed Sup35NM fibrils

Height analysis of in vitro formed Sup35NM fibrils, using nanoscope analysis (Bruker). Individual fibril particles are picked out and their height determined in contrast to the flat mica surface. Fibrils tend to form extensive interactive networks of fibrils; however individual fibrils can be observed and display heights ranging from 5-10 nm. Twisting can be observed along the fibril axis, indicated by height profile fluctuations, something possibly not present in C-DAG formed fibrils.

The fibrils formed *in vitro* display an interactive network of fibrils that display slight variation in height however, display a common height value of around 5-7nm. This is consistent with previous structural analysis of Sup35NM fibrils for fibrils formed at 37°C (Castro et al., 2011). Sup35NM fibrils display strong inter-fibril interactions that result in the large networks observed. This interaction may be due to the charged M domain of the fibrils, and it is possible that the tight clumping of fibrils provides an enhanced resistance to fragmentation. Fibrils formed using the C-DAG system display fibrils that

appear morphologically distinct initially, however upon qualitative analysis conform to a similar range of fibril heights. Figure 41 displays fibril height measurements of individual particles within the AFM images with fibrils formed by C-DAG displaying many individual polymers with a height around 5nm. The initial difference arises from the seemingly enhanced association of fibrils, which form structures that appear as abnormally large fibrils. The large amyloid structures appear to be intertwined fibrils that are a result of the interaction between Sup35NM fibrils. The mechanical perturbation resulting from this may encourage interaction and result in intertwined fibrils that deposit on the mica surface. Various culture debris and cellular material such endogenous protein exports may also enhance the clumping seen within samples imaged in figure 41. Cellular and culture debris are often expected when imaging whole cells. However, regions in which display individual polymers are sufficient for qualitative analysis, albeit a quantitative structural analysis such as with the WALTZ peptides would be more informative as to the differences in structure. Despite this, fibrils formed both via C-DAG and *in vitro* display comparable structures with heights within similar ranges suggesting filament arrangement conservation for C-DAG formed fibrils. Additional structural analysis is required to predict any differences within the core β-sheet structure which has been suggested to differ between [*PSI+*] strain variants, with Sup35p adopting a parallel in register β-sheet or a helical β-sheet (Liebman & Chernoff, 2012). An additional consideration for the potential difference in structure is the attachment of an N-terminal 22 residue signalling sequence. This is the post-cleavage remainder of the 42 residue bipartite signalling sequence (Sivanathan & Hochschild, 2013) to the C-DAG Sup35NM. This non-amyloidgenic sequence is a potential issue for the fibril formation of WALTZ peptides using this system, however for Sup35NM does not seem to hinder the amyloid forming potential. However, it still provides a potential

driving force for alternative filament arrangements or twisting patterns. The sequence itself will repeat along the fibril axis, extending from the N terminal region of the amyloid core while the Sup35 M domain will extend from the C-terminal region of the amyloid core. Therefore, the resulting fibril structures likely display non-amyloidgenic extensions, on two opposing sides of the same lateral plane. Fibrils seen within figure 41 for the C-DAG produced fibrils, display a reduction in height fluctuations along the fibril axis. While *In vitro* formed, fibrils do not display consistent periodic profiles like seen with the WALTZ peptide fibrils, there is a clear fluctuation in height that is indicative of twisting arrangements of the filaments.

In general, polymorphism is a wide spread observation amongst amyloid populations and is manipulated by a plethora of potential factors. There has been extensive studies understanding how different growth conditions affect polymorphism (Meinhardt et al., 2009)(Usov et al., 2013). In addition to this, the concentration of the seeds has been implicated in affecting displayed morphology for recombinant mouse prion protein fragment 89-230 (Sneideris, Milto, & Smirnovas, 2015). Therefore, considering the dynamics of polymorphism it is likely that Sup35NM fibrils formed via the C-DAG machinery present different morphologies to fibrils formed *in vitro.* The fibrils themselves do no not form intracellularly, therefore the fibril formation can be considered *in situ* as it occurs following cellular export of monomer. The environment in which the fibrils form differs from the Sup35NM fibril forming buffer used *in vitro* which consists of 25mM sodium phosphate and 50mM NaCl. The induction plates are the environment for C-DAG fibril formation which is LB media with a NaCl molarity of around 170mM. Therefore, the salt concentration is higher for the C-DAG formed fibrils, with ionic strength being one of the most implicated conditions in affecting morphology (Sidhu, Segers-Nolten, & Subramaniam, 2014). Therefore, it seems that any difference

in polymorphism seen is likely affected by differences in environment conditions such as pH and ionic strength. Additional analysis is required to characterise this fully. However, height fluctuation lacking in C-DAG formed fibrils suggests a reduction in the twisting of filaments within individual fibrils. This may be due to the bi-lateral extensions from the M domain and the CsgA signalling sequence, providing a conformational restriction for tight twisted filament arrangements. How a reduction in twisting would affect the mechanical properties of the fibrils is unclear, however it may reduce the intrafibrillar stability given the stability enhancing effects achieved through filament twisting (Jozef Adamcik et al., 2010). That said, the C-DAG formed fibrils have displayed a level of mechanical stability through their length conservation following a degree of re-suspension mediated mechanical perturbation; albeit not a quantitative measure of stability nor measured exposure of force. Mechanical stability can only be assumed given the temperature of fibril formation, which corresponds to weak [*PSI+*] propagating fibrils that display enhanced rigidity (Castro et al., 2011). Sup35NM has been shown to for strain variants in a temperature dependent manner, with resulting fibrils displaying different mechanical properties potentially due to differences in structure (Castro et al., 2011). The weak [*PSI+*] variant of fibrils formed at 37°C has shown similar strength phenotype to strains propagated by fibrils formed at 23°C, suggesting a structural and mechanical similarity, through resistance to fragmentation (Tanaka et al., 2006). The height of the fibrils is indicative of the filament arrangement, and therefore a characteristic that relates to the fibril thickness and related rigidity. Here, the height similarity between C-DAG (22°C) and *in vitro* formed fibrils (37°C) may suggest similar mechanical properties.

Waltz fusion proteins

Given the implications of the C-DAG system in the nanotechnology field, one potential advantage to using synthetic peptide sequences is the utilization of subsequent amyloid fibrils with desired traits. Traits such as the hydrophobicity, the charge and following our other work on WALTZ; the mechanical properties. By controlling these factors, amyloid fibrils can be produced that are suited to a desired role. As we have shown in our previous work, the WALTZ peptides each produce fibrils displaying different morphological landscapes and they display different stabilities towards fragmentation. Supporting the application of sequence dependent, characteristically tailored fibrils. Therefore, we incorporated the three WALTZ sequences into the C-DAG system to enable their expression and biogenesis within the curli based system. Each of the Waltz peptides was synthesised using (Integrated DNA Technologies (IDT)) however, due to the minimum size requirement of short peptide sequences with this service, the sequence was extended to have overhangs that would usually be unnecessary for subsequent Gibson assembly. Regardless, each sequence was synthesised as the specific WALTZ peptide flanked by an N-terminal and C-terminal 55bp overhang (figure). The N-terminal overhang allows for a short linker sequence between the WALTZ sequence itself and the CsgAss, which is necessary for the conformational ability of the amyloidgenic sequence to adopt its required fold. This linker sequence separates the amyloidgenic sequence from the targeting sequence, which will not be included in the amyloid fold and instead will be exposed and available for proteinase K degradation post fibrillization(Chapman et al., 2010).

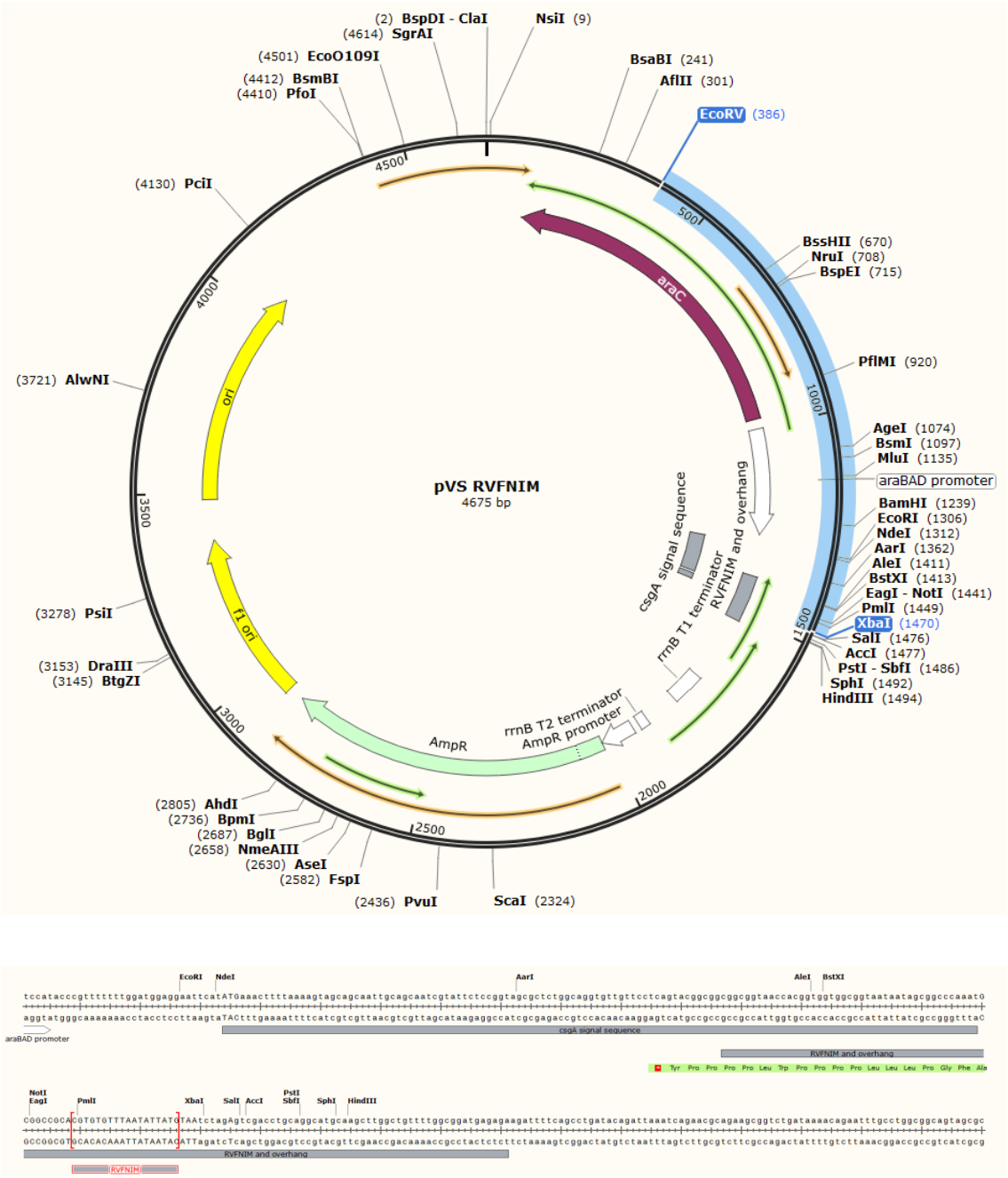


Figure 43. Schematic diagram of the PVS-RVFNIM plasmid

Schematic diagram of the PVS-RVFNIM plasmid used for the attempted generation of RVFNIM fibrils using the C-DAG system. (A) displays the circularized plasmid; consisting of the PVS72 vector with the sup35NM sequence removed and replaced with RVFNIM-CsgAss. The highlighted section dictates the sequence cut out for gel electrophoresis confirmation. (B) the base-pair sequence, which details the RVFNIM sequence flanked by overhangs required for Gibson assembly and the 42-residue bipartite CsgA signal sequence. The Plasmid schematic was presented using Snapgene viewer.

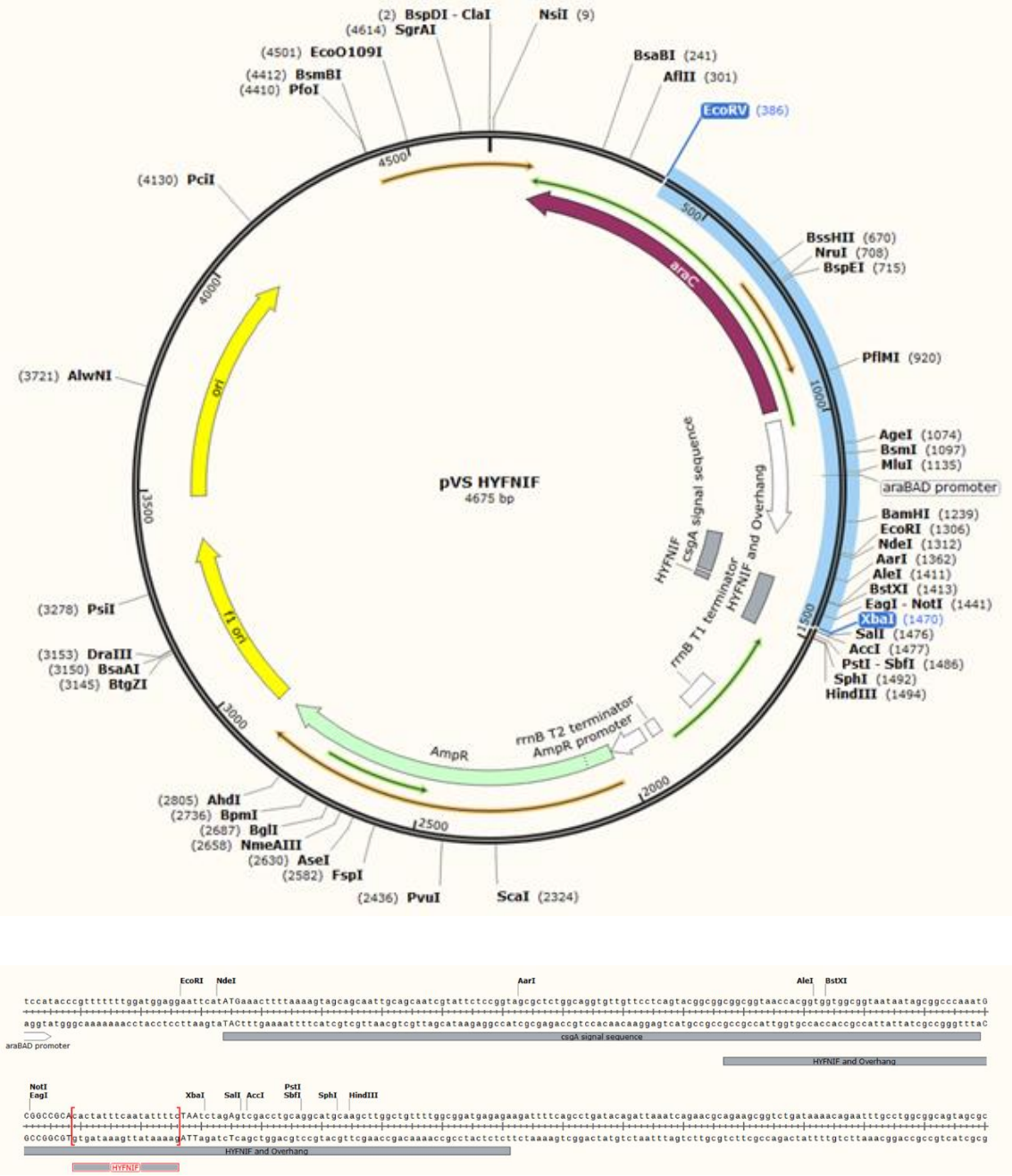


Figure 44. Schematic diagram of the PVS-HYFNIF plasmid

Schematic diagram of the PVS-HYFNIF plasmid used for the attempted generation of RVFNIM fibrils using the C-DAG system. (A) displays the circularized plasmid; consisting of the PVS72 vector with the sup35NM sequence removed and replaced HYFNIF-CsgAss. The highlighted section dictates the sequence cut out for gel electrophoresis confirmation. (B) the base-pair sequence, which details the HYFNIF sequence flanked by overhangs required for Gibson assembly and the 42-residue bipartite CsgA signal sequence. The Plasmid schematic was presented using Snapgene viewer.

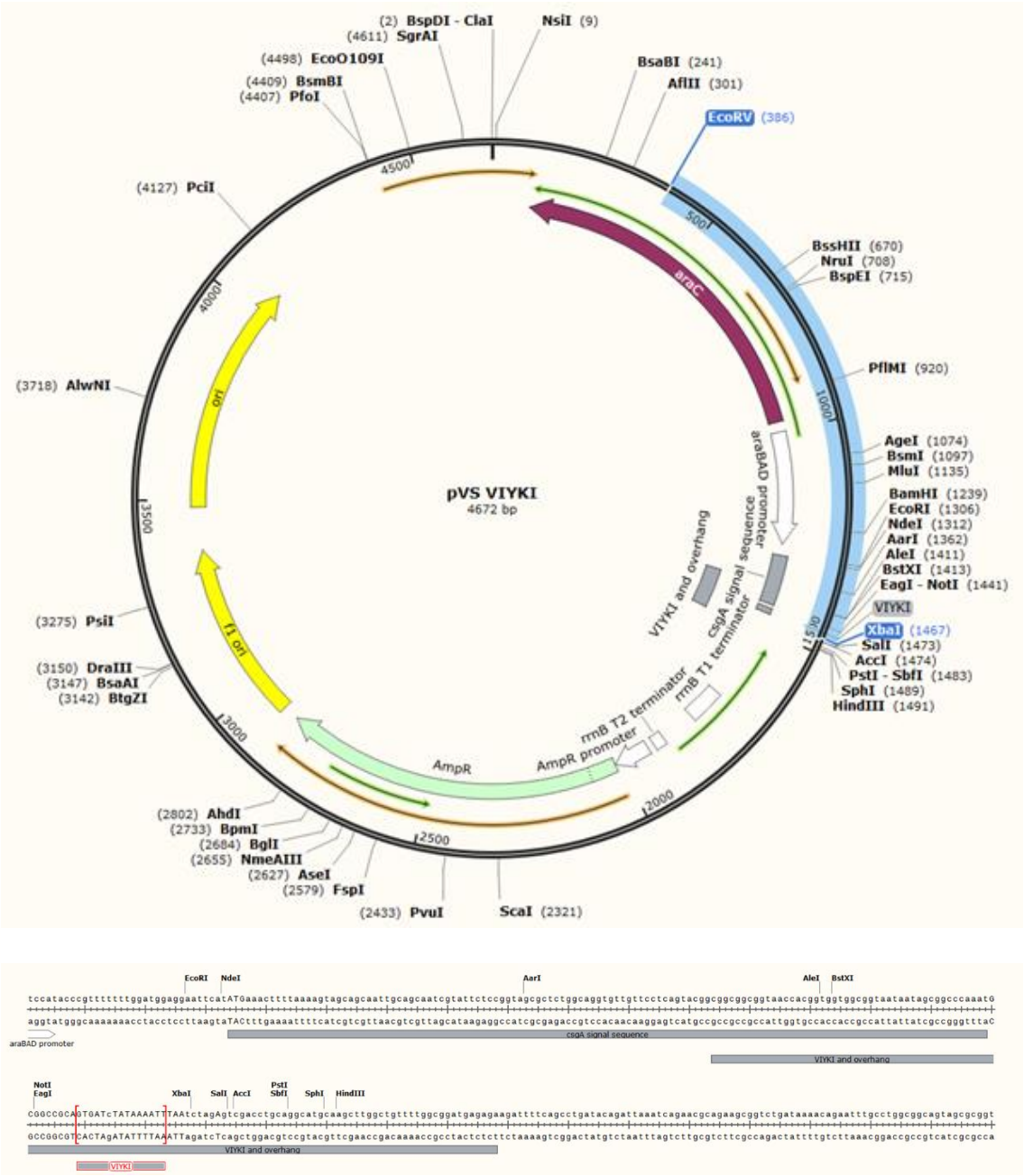


Figure 45. Schematic diagram of the PVS-VIYKI plasmid

Schematic diagram of the PVS-VIYKI plasmid used for the attempted generation of VIYKI fibrils using the C-DAG system. (A) displays the circularized plasmid; consisting of the PVS72 vector with the sup35NM sequence removed and replaced VIYKI -CsgAss. The highlighted section dictates the sequence cut out for gel electrophoresis confirmation. (B) the base-pair sequence, which details the VIYKI sequence flanked by overhangs required for Gibson assembly and the 42-residue bipartite CsgA signal sequence. The Plasmid schematic was presented using Snapgene viewer.

The WALTZ fragments need to be incorporated into the PVS72 backbone which contains the CsgAss and an arabinose inducer for the fusion protein. The pVS72 plasmid consists of the pExport plasmid with Sup35NM as the protein of choice. The pExport plasmid (*bla* P_{BAD} csgAss-(protein)-His_6, pBR322 ori) produces CsgA residues 1 – 42 fused to a protein of choice and a C-terminal His_6-tag. The kit comes with Sup35nm as the inserted protein, forming the PVS72 plasmid. Therefore, the protein or peptide of choice is required to be inserted following the excision of the Sup35nm coding sequence. This was achieved with a simple restriction digest of the PVS72 plasmid followed by a Gibson assembly between the PVS72 backbone (pExport) and the respective WALTZ fragments. PVS72 was digested using Xba1 and Not1, with Not1 cutting within the linker region between the Sup35nm coding sequence and the CsgAss (figure). The PVS72 fragment which contains the CsgAss was then ligated with the respective WALTZ sequences by Gibson assembly. The Gibson assembled products result in the constructs seen in figure, which consist of an upstream arabinose inducible promoter and the fusion protein; which itself consists of the N-terminal 126bp CsgAss, a short linker region then the WALTZ sequence followed by an immediate stop codon. As per the established mechanism of the C-DAG kit, the bipartite signalling sequence will be cleaved following its transition to the periplasm via the SEC translocon. This cleavage will remove 20 N-terminal residues, leaving the 22-residue CsgG targeting sequence attached to the respective WALTZ peptides. The Gibson assembled products were transformed into DH5α *E. coli* and grown on ampicillin selective LB agar plates, which with successful growth confirmed the transformation of the Gibson assembled product. Standard miniprep cloning procedures were carried out to provide sufficient PVSWALTZ plasmid for subsequent experiments. The final WALTZ constructs were confirmed via restriction digest and gel electrophoresis. Each WALTS construct was digested using EcoRV and

Xba1, and subsequently analysed on an agarose gel following electrophoresis (figure). As seen in figure 46 the expected bands for each WALTZ construct is the same, and correspond to a 3591bp large fragment and 1084bp smaller fragment. The UV imaged agarose gel seen in figure shows bands the correspond to the expected size for each of the WALTZ peptides, as well as PVS72 as a control which presents expected band sizes of 4646bp and 788bp following a digestion with Xba1 and Not1. Therefore, it seems that the Gibson assembly construction of the PVSWALTS peptides was successful given the expected plasmid size and successful growth on selective media. These relatively straight forward steps highlight the versatility and ease of use for the C-DAG kit to by modified as intended, to allow for the expression and export of any amyloidgenic sequence.

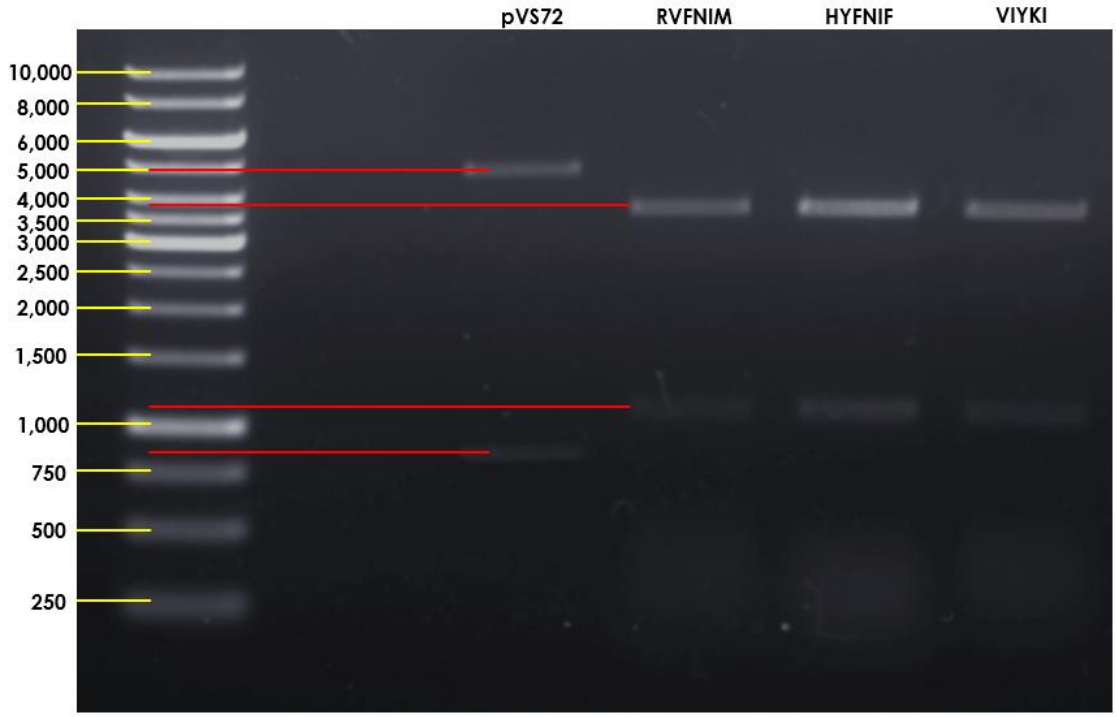


Figure 46. DNA gel electrophoresis of the desired PVS-WALTZ plasmids

DNA gel electrophoresis was used to confirm the Gibson assembly of the desired PVS-WALTZ plasmids. Each sample was digested with the appropriate restriction enzymes and loaded in1X DNA loading dye. The bands are identified using GeneRuler 1 kb DNA Ladder. pVS72 serves as a control to reject the possibility of a failed Gibson assembly. Each band corresponds to the expected band size of 3591bp for the large fragment and 1084bp for the smaller. Each WALTZ construct was designed with the same sized insert; therefore, the band sizes are equal.

To observe the differences in fibril morphology of the respective WALTZ fusion proteins within the C-DAG system, we induced and grew VS45 cultures that had been transformed with one of the 3 WALTZ constructs. This was done per the well characterised and patented induction procedure for facilitating the biogenesis of exogenous protein or peptide sequences (Sivanathan & Hochschild, 2012). As seen in figure 46, we have confirmed the construction of the required plasmid therefore a simple incorporation into the required strain and the subsequent induction on IPTG and arabinose supplemented media is sufficient to enable the production of the WALTZ peptides. Transformation was confirmed following the colony formation on media supplemented with chloramphenicol and ampicillin. Chloramphenicol resistance is conferred to VS45 by the plasmid pVS76. Ampicillin resistance is conferred by the PVS-WALTZ plasmid. The induction on the required media is a 5-day procedure at 22C and the resulting plates can be seen on figure 47. The induction of the WALTZ transformed strains was accompanied by a negative and a positive control in the form of PVS105 and PVS72 transformed VS45, respectively. In addition, the standard induction media, the strains were all grown on equivalent Congo red media. Congo red is an amyloid specific dye that interacts and binds to the amyloid core, and fluoresces with a bright red culture. Therefore, bacterial cultures that have exported amyloidgenic monomer which forms fibrils, such colonies will be stained Congo red resulting in bright red colonies. This can be observed in figure 47 with PVS72 the positive control which produces Sup35nm fibrils. These colonies have a distinctive bright red colour indicative of the presence of amyloid. Unfortunately, the 3 WALTZ transformed strains do not display red colonies and are similar in colour to the PVS105 negative control. This was an unexpected result given the confirmation of the WALTZ fusion plasmids, and suggests that the material has either not been produced, not been exported, or monomeric material has not been

able to adopt the correct and required amyloid fold. However, Congo red cannot be solely relied upon for the confirmation of amyloid material.

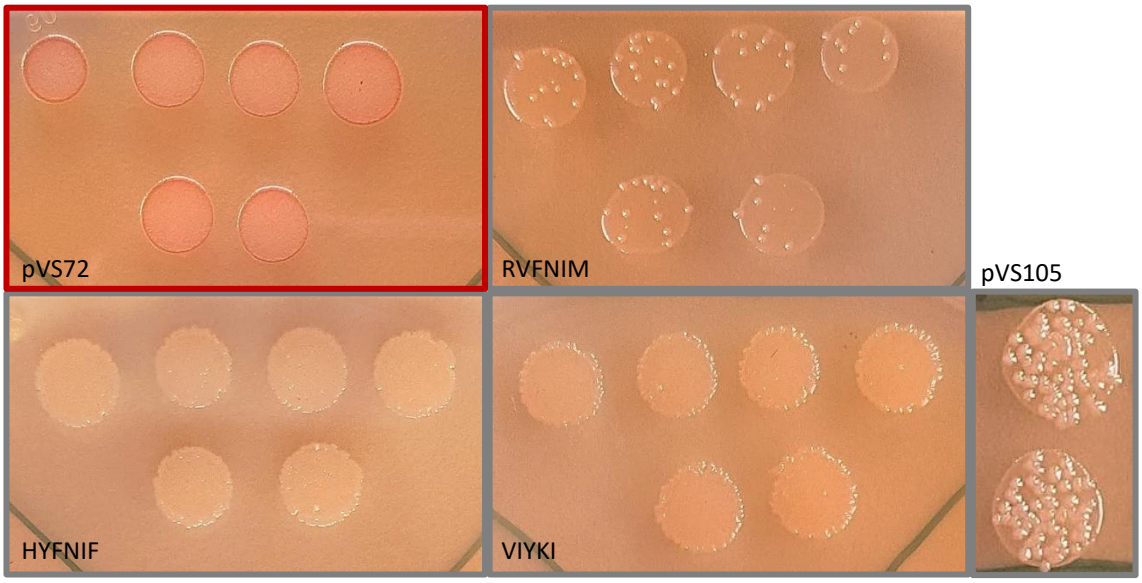


Figure 47. Congo Red validation of amyloid forming colonies

Congo Red detects amyloid structure, resulting in a bright red color expressed by material grown on media containing the dye. In this figure, pVS72 is expressed within VS45 E. coli and grown on Congo red induction media. The resulting colonies express a bright red color due to the generation of extracellular Sup35NM fibrils. The VS45 strains expressing RVFNIM, HYFNIF and VIYKI however, do not present red colonies due to an inability of WALTZ-CsgAss fibrils to form using the C-DAG system. The control pVS105 does not form red colonies as expected.

Imaging of WALTZ expressing VS45 using AFM

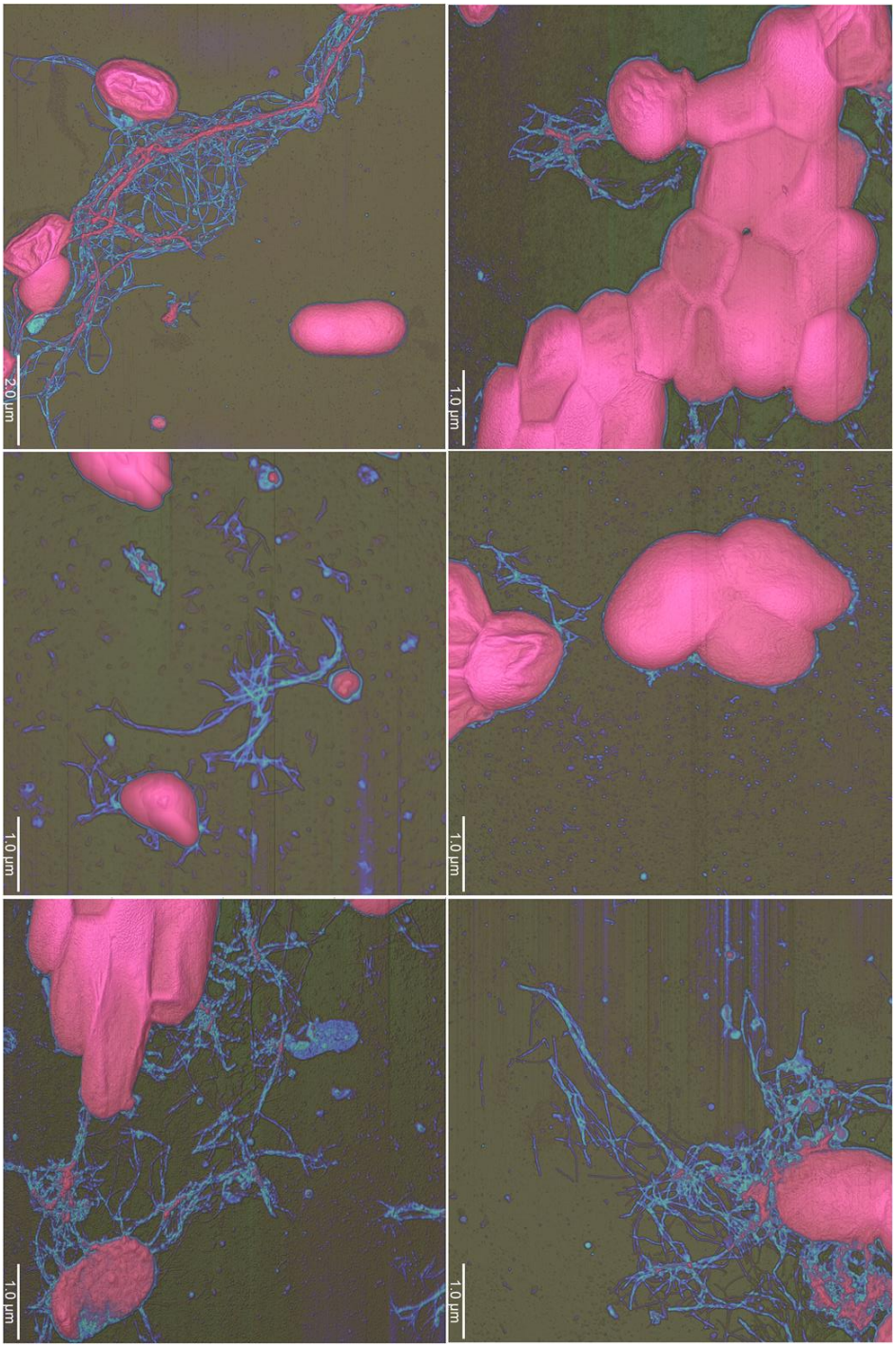


Figure 48. Example AFM images of VS45 expressing pVS72

Example AFM images of VS45 expressing pVS72. The C-DAG system effectively generates Sup35NM fibrils, which are observed as being closely associated with the E. coli cells. Fibril particles form interactive networks of fibrils possibly facilitated by the same interactions observed with in vivo formed Sup35NM fibrils. Isolated fibrils display a height expected of Sup35NM fibrils. The cells themselves are of the expected size for E. coli but appear dehydrated and flattened through the AFM sample preparation which is necessary to ensure adhesion of both cells and fibrils. Images range from 6 x 6-10 x 10 µm surface areas however all maintain a resolution of 2048 x 2048 pixels. Images required a reduced scan rate of 0.203Hz for effective image collection. The height based color contrast provides clear differentiation between cells and amyloid material, however is not scaled for accurate visual determination of height. Scale bars represent either 1 or 2 µm

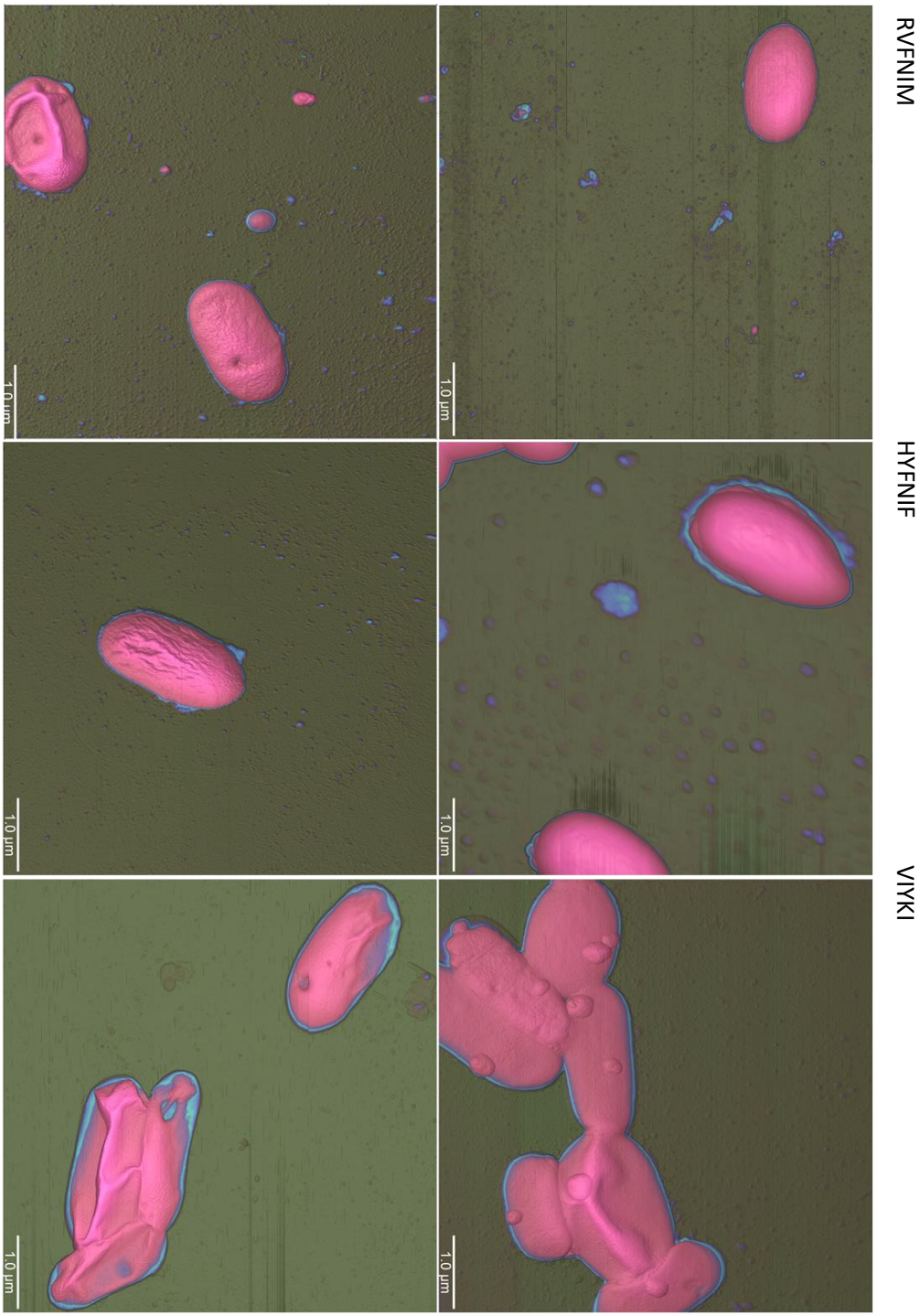


Figure 49. Example AFM images of VS45 expressing RVFNIM, HYFNIF and VIYKI fusion proteins

Example AFM images of VS45 expressing RVFNIM, HYFNIF and VIYKI fusion proteins. While the C-DAG system likely exports protein monomer, there is a failure to generate WALTZ peptide fibrils. The cells themselves are of the expected size for E. coli but appear dehydrated and flattened through the AFM sample preparation which is necessary to ensure adhesion of both cells and fibrils. Images range from 6 x 6 µm surface areas however all maintain a resolution of 2048 x 2048 pixels. Images required a reduced scan rate of 0.203Hz for effective image collection. The height based color contrast provides clear differentiation between cells and amyloid material, however is not scaled for accurate visual determination of height. Scale bars represent 1µm.

Following the results of the amyloid detecting Congo red assay, visual validation was required. AFM was used to image the samples which had been resuspended from the induction media and incubated on mica surfaces. AFM is useful for imaging cells; however appropriate probes are required and are generally ones that are relatively stiff as to not create too much image noise that arises from the strong interactions between the probes tip and the cell surface. However, to image the amyloid fibrils in addition to the cells, stiffer probes will not achieve the required sensitivity to pick up detail qualitative features that are required for a morphological comparison with *in vitro* formed fibrils. In addition to this, the considerable size difference between the cells and the fibrils will mean that tip velocity associated convolution of the cells may result in an inability to detect fibrils in close proximity to the cell. Therefore, to achieve all imaging requirements for a qualitative AFM analysis of the C-DAG produced fibrils and the *E. coli* themselves, we opted for the flexible ScanAsyst probe (silicon nitride triangular tip with tip height = 2.5-8µm, nominal tip radius = 2nm, nominal spring constant 0.4N/m and nominal resonant frequency 70kHz) that is ideal for resolving details within small samples such as the fibrils. However, to account for excessive noise and force induced tip resonance we adjusted the parameters to counter this (as described in chapter 2).

A reduced peak force, gain and a minimal scanning rate allowed for the effective image of these potentially difficult samples. Congo red induction plates confirmed the presence of amyloid within the PVS72 strains (figure 47) which is to be expected given its extensive characterisation (Sivanathan & Hochschild, 2012). However, for the purposes of a control to the WALTS strains and for subsequent morphological qualification, this sample was imaged using AFM. The images seen in figure 48 present the expected morphology of cells and amyloid fibrils. There is varied fibril density, from

extensive to sparse in relation to cellular proximity. However the association of Sup35NM-CsgAss fibrils to the cells themselves is not well characterised, and potentially minimal given the absence of a cell surface linked nucleation protein like CsgB in the endogenous curli system (Hammer et al., 2012). However, there is evidence of fibril association to the cells; considering the relatively harsh technique required for imaging that involves the resuspension of colonies from the induction media. This technique is likely to both separate the fibrils from the cells, as well as potentially fragment them. However, the images display what can be considered as unfragmented fibrils, which are in close proximity to the cells. This interaction may be facilitated by the highly-charged M domain of Sup35NM which interacts with the cell surface, as well as to other fibrils. Unfortunately, the negative Congo red assay for the WALTZ strains was confirmed through imaging. The images seen in Figure 49 display cells deposited on the mica surface, however they lack any evidence of fibril formation. This together with the Congo red result confirm an inability for the system to produce amyloid forming WALTZ peptides, at least in accordance to the method we used here. The VS45 strain had effectively been transformed which was confirmed by its displayed resistance to both ampicillin and chloramphenicol, however the subsequent failure in the system is unknown.

The goal of this C-DAG application was to determine if the expression of different fusion peptide sequences will produce fibrils with different structure and mechanical properties; which are a result of the different morphologies adopted by the respective sequences. By manipulating the resulting fibrils in regards to their mechanical properties, there is a potential to subsequently tailor the properties of the biofilm itself. By facilitating the biogenesis of polymorphic fibrils, or fibrils with differences in hydrophobicity, charge, or aromaticity, you may be able to influence the characteristic

behaviour that results from biofilm production. For example, the colonising potential for different surfaces is influenced by the biofilm itself. Therefore, by manipulating the charge if the fibrillar element which is potentially the most important structural element, the charge based interaction between the bacterial biofilm and the culture seeding surface can be manipulated. In addition to this, the hydrophobic nature of amyloid fibrils can be altered, by changing the surface exposed residues and therefore the net hydrophobicity of the polymerised fibril. Manipulation of this property may be able to affect the propensity for bacterial biofilms to form pellicles which is colonisation at the air water interface. In addition to this the hydrophobicity and the aromaticity may be altered to potentially control the rate of fibril formation following monomer export from the cell. This is on the basis that the conformational changes required to facilitate fibril formation is a process driven by the aromatic or hydrophobic potential. This is a favourable conformation with the most hydrophobic residues folded into the core structure and the stability enhancing alignment of aromatic residues within the steric zipper interface (Nelson et al., 2005). A likely explanation for the inability of the WALTZ fusion peptide to form fibrils is a conformational difficulty given the size of the WALTZ peptide and the attached signalling sequence, CsgAss. RVFNIM and HYFNIF are hexapeptides and VIYKI a pentapeptide, however the signalling sequence once cleaved is 22-residues long.

It is known that amyloid, for example Sup35p, can form fibrils with attached non-amyloidgenic sequences through an aggregation prone propensity of the amyloid core which supersedes any potential restrictions provided by non amyloidgenic regions (Luckgei et al., 2013). Therefore, Sup35p is amyloidgenic as its full-length protein while the WALTZ peptides are not. Each WALTZ peptide is identified and isolated from larger non-amyloidgenic proteins (Maurer-Stroh et al., 2010). The WALTZ sequences represent

the maximum amyloid forming frame for themselves, while also representing a general minimum number of residues required to form amyloid fibrils. Therefore, the amyloid core is both small, and restrictive in that expansion can inhibit the amyloid forming potential (Reumers et al., 2009). The size of the core, may not be able to support the additional attachment of non-amyloidgenic domains which may present a size restriction that does not allow the fusion peptide to adopt a confirmation that favours aggregation. The term gatekeeper in regards to amyloidgenic sequences describes residues that flank the amyloid forming core sequence, that through charge repulsion and inaccessibility to β-sheet, inhibit amyloid formation (Reumers et al., 2009). Arginine, lysine, aspartate, and glutamate are all charged residues, and the presence of these residues as gatekeepers disrupt the ability of the core sequence to form amyloid. In addition to this Proline, disrupts amyloid formation though its inability to conform to β-sheet structure. The WALTZ sequences themselves are characterised in their propensity to form amyloid, using the position sensitive scoring algorithm that identified them (Morris et al., 2013). However, the TANGO algorithm is more effective at detecting the window of amyloidgenicity. It does so by scoring the amyloid propensity from 0-100%; the window is the region in which the residues score >0% and the window itself scores above >50%. Long stretches of incompatible or detrimental residues will reduce an expanded windows score, therefore the window sequence >50% when isolated will form amyloid (Fernandez-Escamilla et al., 2004). With a fixed 6 residue window, sequences can be scanned and scored per the position sensitive WALTZ algorithm which is most effective at differentiating between amyloid forming and amorphous aggregation prone sequences. The full fusion peptide sequence as a TANGO analysed sequence, contains 2 prolines within the N-terminal targeting sequence therefore reducing the amyloid forming potential. In addition to this the WALTZ identified

sequence is no longer isolated and the amyloid forming core may not provide sufficient amyloidgenicity to overcome the propensity reducing residues within the attached sequence, therefore the WALTZ identified core is not able to polymerise with an attached domain. Studies on gatekeepers generally consider immediately flanking residues, which in the case of the fusion peptide used here; is a triple alanine linker sequence. However, the range in which gatekeeper residues can enact electrostatic inhibitory effects is not known, and is likely far reaching with undefined boundaries (Reumers et al., 2009). The N22 sequence only contains proline out of the possible gatekeeper residues (X. Wang & Chapman, 2008) therefore likely cannot provide sufficient disruptive effects to inhibit aggregation of Sup35NM which possesses a much larger amyloid forming core. However, proline displays the strongest inhibitory effects compared to the other gatekeeper residues; therefore, the 2 residues that are present may be sufficient to inhibit a small amyloid forming region such as the WALTZ peptides. Additional experiments may shed light on this, and it may be possible that tandem repeats of the WALTZ peptides can provide a sufficient propensity to overcome any potential gatekeeper or conformational constraints.

Conclusion

The C-DAG system provides the machinery to express, export and facilitate the formation of exogenous amyloid fibrils. The extensively characterised incorporation of Sup35NM(Sivanathan & Hochschild, 2012) results in fibrils that qualitatively appear comparable to fibrils formed via more classical *in vitro* methods. While they appear similar in height derived filament arrangement, the twisting patterns appear reduced; possibly due to the presence of bi-laterally attached non amyloidgenic domains. The fibrils are likely comparable in mechanical properties to the *in vitro* formed fibrils given that the respective temperatures give rise to fibril populations that propagate the [*PSI+*]

phenotype with similar strengths (Tanaka et al., 2006). While Sup35 is considered a functional amyloid through its role in translation termination (Knowles & Buehler, 2011), the rigid fibrils likely resulting from C-DAG production present desirable mechanical traits for nanomaterial applications, such as conductive nanowires. Unfortunately, the application of the C-DAG system to facilitate the biogenesis of any amyloidgenic sequence does not likely extend to the incorporation of short peptide sequences such as the WALTZ sequences attempted here. The exact cause of the inability to form WALTZ-CsgAss fibrils is not known. However, it is possibly a conformational restriction in which the small amyloid core cannot overcome the addition of a non amyloidgenic domain. Or it is a reduction in the amyloid forming potential of the WALTZ sequence due to the addition of flanking residues, including the 'gatekeeping' residue proline within the N-terminal 22 residue CsgA targeting sequence.

Chapter 6

Summary and conclusions

The relationship between protein structure and function is one that has been at the forefront of research when aiming to understand how a protein enacts its purpose. The correct folding of native proteins is vital towards the correct biological function. Amyloid is a class of protein or peptide that has adopted a common but highly ordered structure, characterised by the presence of a β-sheet rich core (Nelson et al., 2005). There are many diseases linked to amyloid, for example Parkinson's disease which involves Alpha-synuclein and Diabetes mellitus type 2 which involves Amylin (IAPP). Alzheimer's disease (AD) is the most common of the fatal neurodegenerative dementia diseases, and is an associated pathology of $A\beta_{40/42}$ and Tau cytotoxicity. Therefore, the structure-function relationship of amyloid is an important research field. Aiming to better understand how structure may influence disease progression, therefore is of high importance.

How amyloid conveys its cytotoxic potential is unclear however. Fibril ends have been proposed as the active sites for membrane interaction and disruption (W. F. Xue, Hellewell, et al., 2009). How fibrils interact with biological membranes may be a mechanism shared amongst all amyloid (Williams & Serpell, 2011) however it may also be due to the specific residues of a given amyloid protein monomer. This in addition to a size dependent mechanism for enhanced cytotoxicity (W. F. Xue, Hellewell, et al., 2009), implicates structure and morphology as a key factor in disease progression. Various studies have eluded to a sequence – structure relationship for amyloid (Campioni et al., 2010; Karen E Marshall et al., 2014) for both the intermediate elongation species that form and the fragmentation or dissociation of fibrils, also producing toxic species (W. Xue et al. 2010). Considering length is an important factor towards disease progression, other structural dimensions are also of interest. The thickness of an amyloid fibril has implications for both size-dependent cell surface

interactions, and mechanical properties. The latter being an important factor that likely affects how a fibril resists fragmentation, essentially increasing the particle population.

Assembly polymorphism describes a difference in the hierarchal arrangement of filaments within a fibril. EM and AFM imaging can differentiate between polymorphs of a single species of fibril, by identification of the number of filaments per fibril, degree of twisting (via periodicity) and the length and diameter of fibrils (Jozef Adamcik & Mezzenga, 2012; Fändrich et al., 2009; Fitzpatrick, Debelouchina, Bayro, et al., 2013; W. F. Xue, Homans, et al., 2009). Such differences in morphology likely infer differences in mechanical properties. Amyloid fibrils have been shown to have impressive mechanical properties, with tensile strength like steel and a bending rigidity comparable to spider silk (J. F. Smith, Knowles, Dobson, Macphee, & Welland, 2006). However, such properties are not shared by all amyloid, suggesting they are influenced by a sequence-structure relationship.

It has been suggested that all proteins can form amyloid under certain conditions and in physiological conditions, the barrier of free energy between the two conformations (aggregated and native) is often overcome (Christopher M Dobson, 2003). Some amyloid is not linked to disease and is deemed “functional” amyloid. For example, Curli fibers of E. coli and other bacteria produced amyloid has been linked to cell signalling via the retention of quorum signalling molecules within biofilms (Seviour et al., 2015). Considering the diverse mechanical properties of amyloid and the efficiency of aggregation, amyloid has strong implications in a biotechnology role. It has been the nanomaterial of choice in the development of nanomechanical projects such as the development of gold labelled conductive nanowires (Scheibel et al., 2003).

The work reported here, enhances the understanding of fibril morphology in addition to the relationship between structure and stability towards breakage.

Expanding the morphology landscape of amyloid fibrils

Amyloid fibrils are known to adopt various morphologies, defined by the number and arrangement of filaments. It is important to understand the structural morphology of amyloid fibrils as different filament arrangements may grant variation in the residue side chain profile along a fibril axis and at the fibril ends; which may be involved in cytotoxicity in a sequence dependent manner (Karen E Marshall, Marchante, Xue, & Serpell, 2014). This concept, is further complicated based on a sequence dependent polymorphism such as that presented here. The polymorphism adopted by a particular amyloid protein is often quite narrow, in that only a few different structures are observed (Claire S Goldsbury, Cooper, Goldie, & Mu, 1997). In addition, it is often reported that the polymorphism present only varies in filament number and not arrangement (Jozef Adamcik & Mezzenga, 2012). This is often a result of direct observation from various imaging techniques, with the acknowledgment of further structures that may be accessible, but simply not observed.

Here in chapter 3, using the WALTZ peptides: HYFNIF, VIYKI and RVFNIM we formed fibril populations known to be polymorphic (Morris et al., 2013). Fibrils were imaged using AFM and the high-resolution images were subsequently analysed in Matlab using hierarchical clustering. Using this clustering technique, a wide range of fibril morphologies were identified and clustered based on their height profile and periodicity with similar fibrils being grouped into clusters. Using the characterised dimensions of the repeating unit for each peptide (Morris et al., 2013) structural models were developed, accounting for all possible arrangements of filaments in a hierarchical order.

The maximum, minimum and average heights of fibrils within identified clusters, were used to correlate with equivalent dimensions of model structures (chapter 3, figure 23). Thereby identifying which morphologies are present within the respective populations.

We found that each fibril population displayed a wide range of morphologies that varied regarding their filament number and twisting archetype. The ranges of morphologies present for each sample differed also, supporting a sequence dependent relationship for structure. We observed considerable variation across all height and periodicity values, suggesting a flexibility in the structure that fibrils adopt. The hierarchical clustering technique enabled the identification of boundaries for fibril morphologies, which otherwise were too unclear to define. In a biological environment polymorphism has been reported amongst clinically relevant amyloid (Annamalai et al., 2016; Liu et al., 2016; Lu et al., 2013). And the manipulation of morphology has been demonstrated by altering the fibril forming conditions, and the conditions of mature fibrils (Usov et al., 2013). Therefore, disease associated fibrils may have access to a wide range of morphologies such as the ones presented here through a condition dependent mechanism. Fibril morphology can be altered through changes in environment pH and ionic strength (Kurouski et al., 2013; Vandenakker et al., 2011). This supports an increased range of adoptable morphologies for all amyloid and it is possible that slight changes in conditions will increase the likelihood of alternate morphologies being adopted. The most common morphology will likely be one that is the lowest in free energy, based on the inter-residue interactions and that with the environment. While the sequence based properties such as aromaticity and net charge will affect morphology, so will the residue interaction with the environment (Chiti & Dobson, 2006). Many morphologies may present similar free energy levels, making them similar in likelihood of adoption. However, others such as nanotube strcutures may be time

dependent, with increased inter-fibril and solution interactions over time (chapter 3;(Lara et al., 2014).

The work presented here, expands on the understanding of the possible structures that an amyloid protein may adopt. Morphologies adopted within the fibril populations characterised here, vary in frequency, suggesting that certain morphologies are less common than others. Through the observation of switching fibril morphologies (figure 26), it is likely that the fibril morphological landscape is dynamic and progressive and not static. Meaning that fibril morphologies observed may not be indicative of the entire range of structures accessible to an amyloidgenic protein or peptide.

Understanding Mechanical stability of amyloid fibrils

Amyloid fibrils have been well characterised regarding their mechanical properties, such as the persistence length (L_p) which has been shown to be affected by molecular structure (Vandenakker et al., 2011). Here in chapter 4 we have generated fibrils from HYFNIF, RVFNIM and VIYKI and a quantitative analysis of fibril fragmentation was conducted, in comparison to fibrils formed from α-synuclein. Each sample was sonicated to achieve a reduction in length over time and images were collected by AFM. Subsequent image analysis allowed us to quantify the population reduction in length, shift in height and population L_p. During the time course of fragmentation, the shift in height distribution (chapter 4, Figure 33) gave us some insight into how different WALTZ formed fibril structures (characterised in chapter 3) are possibly fragmenting at different time points. However, fibrils formed from α-synuclein did not display the same shift in height distribution which is likely attributed to the low level of observed polymorphism. Morphological differences in fragmentation are probably due to differences in the mechanical properties of the respective structures, with each population having different distributions of structures that are more, or less resistant to fragmentation

(chapter 3). Overall, we found that Each of the WALTZ fibril populations, displayed increased resistance to fragmentation compared to α-synuclein (chapter 4). Each WALTZ fibril population displays structures that are considerably larger than those formed by α-synuclein, suggesting a height and morphology relationship towards fragmentation resistance. In addition to this, we found a strong relationship between persistence length and fragmentation stability that suggests enhanced rigidity provides increased fragmentation resistance. Further to this, we propose a relationship between the difference between length and L_p. We suggest that the larger the difference in fibril length compared to persistence length, the less likely the fibril is to buckle and therefore break. The results in chapter 4 support this, meaning that the shorter a particle becomes the more resistant to fragmentation it will be, however will not display equal resistance with fibrils of equal length but increased L_p.

The mechanism for cytotoxicity has been extensively researched, which implicate the mechanical properties of fibrils. The enhancement of cytotoxicity through fragmentation (W. Xue et al., 2010a) suggests a size dependent mechanism for cytotoxicity. In such a mechanism, the generation of smaller particles can be reduced amongst fibril structures with an increased resistance towards breakage. Alternatively, a size-independent mechanism of cytotoxicity would involve cell membrane interaction through lateral association (Monsellier et al., 2016). The mechanical properties of fibrils would affect the ability of the fibrils to effectively bend and interact with a cell membrane due to changes in rigidity. In such a mechanism, short rigid particles would enact a higher number of membrane interactions as an equivalent mass of long rigid particles.

Structural comparison between amyloid fibrils formed either *in vitro* or *in situ*.

The C-DAG system provides the machinery to express, export and facilitate the formation of exogenous amyloid fibrils (Sivanathan & Hochschild, 2013). Properties of the amyloid produced using this system are based on the structural and functional knowledge of *in vitro* formed fibrils. Due to the addition of a signalling sequence and an alternate environment for fibril formation it is likely that fibril properties may change. Understanding the possible discrepancy between *in vitro* formed amyloid fibrils and those formed using C-DAG *in situ* on *E. coli* cells, is important to determine possible issues with using the system. This understanding is required to draw meaningful conclusions that can apply to in vitro and in vivo formed fibrils. Characterising and comparing the resulting fibril structures, will support a synthetic biology approach to generating amyloid for desired functions or as a nanomaterial.

Here in chapter 5, we present a structural comparison between fibrils formed from Sup35NM in vitro and those using the C-DAG system. Sup35NM fibrils formed by *E. coli* using the C-DAG methodology described in chapter 2, were imaged by AFM and subsequently analysed regarding observed height. These observations were made in parallel with Sup35NM fibrils formed *in vitro*. The resulting similarities support the application of this system to produce fibrils based on in vitro formed properties. The extensively characterised incorporation of Sup35NM (Sivanathan & Hochschild, 2012) results in fibrils that qualitatively appear comparable to fibrils formed via more classical in vitro methods. While they appear similar in height derived filament arrangement, the twisting patterns appear reduced; possibly due to the presence of bi-laterally attached non amyloidgenic domains. Sup35 is considered a functional amyloid through its role in translation termination (Knowles & Buehler, 2011), the rigid fibrils likely resulting from

C-DAG production present desirable mechanical traits for nanomaterial applications, such as conductive nanowires. Unfortunately, the application of the C-DAG system to facilitate the biogenesis of any amyloidgenic sequence does not likely extend to the incorporation of short peptide sequences such as the WALTZ sequences attempted here. The exact cause of the inability to form WALTZ-CsgAss fibrils is not known. However, it is possibly a conformational restriction in which the small amyloid core cannot overcome the addition of a non amyloidgenic domain.

Concluding remarks and future work

The relationship between amyloid and disease is a complex one, with the involvement of amyloid being suggested as either the cause of pathology or conversely as a result of the pathology (Herrup, 2015). Regardless, the observed activity of amyloid in their respective disease models has been varied and linked to structure (Annamalai 2016; Liu et al., 2016; Lu et al., 2013).

Amyloid structure is incredibly varied, with polymorphism present at both the molecular level of the β-sheet interface (Nelson et al., 2005; Petkova et al., 2005) and polymorphism in the hierarchical assembly of filaments, reported here. Therefore, given the importance of protein structure towards function, it is likely that amyloid structure has an important involvement in any potential mechanism of cytotoxicity.

The hierarchical clustering method used here was fully developed from the ground up and represents a completely novel approach to characterising fibril morphology from large populations of fibrils. Due to the flexibility of the scripts involved, the technique can be applied to any set of AFM collected fibril data, which would allow for similar structural analysis to be conducted with more clinically relevant fibril populations. The results here, suggest that amyloid species have access to a vast range of morphologies

in a sequence dependent manner. Certain morphologies are more common than others, and some are similar in structure which suggests a flexibility in overall morphology. Therefore, to further our work it will be important to determine which structural motifs, or which property altering residues may drive the adoption of certain structures. This could be achieved by systematic synthesis of modified peptide sequences based on the 3 WALTZ sequences used here. By controlling the residue addition, replacement or removal of these 3 sequences we can determine how both polymorphism changes, and the subsequent stability towards fragmentation. In addition to this, X-ray fibre diffraction (XRFD) could enable modelling of the molecular interface (Morris et al., 2013) therefore determining how the sequence-core structure relationship may affect polymorphism and stability towards fragmentation. Understanding this relationship will enable the tailoring of synthetic sequences that favour certain, desired morphologies.

Using the same polymorphic fibril populations, we have observed distinct differences in both stability towards fragmentation and overall mechanical properties. Again, this likely due to the differences in adopted morphology by each population of fibrils. Therefore, to fully understand how structure can infer mechanical stability it is important to isolate and analyse individual morphologies within the population. This could be achieved through an increase in the size of our quantitative analysis. Due to the nature of certain morphologies being less common, or some possibly time dependent (Lara et al., 2014), increasing our structure database would provide sufficient numbers of each morphology to apply separate analysis. Therefore, we could determine separate L_p values, providing valuable information on the mechanical properties of independent morphologies. Finally, modification of the WALTZ sequences may overcome the limitations of applying our WALTZ peptides to the C-DAG system. This could be achieved be using a repeating sequence, thereby overcoming any

conformational restrains due to an increased size of the repeating amyloidgenic monomer. An overall combination of the understanding and control of fibril morphologies with unique mechanical properties formed from synthetic sequences, with the C-DAG system, provides a powerful platform for the generation of fibrils for a desired purpose. Additionally, understanding the structural motifs that confer properties of fibrils which enhance cytotoxicity, provides support for the development of drugs which may target the amyloid material itself or its respective DNA sequence.

Bibliography

Adamcik, J., Castelletto, V., Bolisetty, S., Hamley, I. W., & Mezzenga, R. (2011). Direct observation of time-resolved polymorphic states in the self-assembly of end-capped heptapeptides. *Angewandte Chemie - International Edition*, *50*(24), 5495–5498. https://doi.org/10.1002/anie.201100807

Adamcik, J., Jung, J.-M., Flakowski, J., De Los Rios, P., Dietler, G., & Mezzenga, R. (2010). Understanding amyloid aggregation by statistical analysis of atomic force microscopy images. *Nature Nanotechnology*, *5*(6), 423–428. https://doi.org/10.1038/nnano.2010.59

Adamcik, J., Lara, C., Usov, I., Jeong, J. S., Ruggeri, F. S., Dietler, G., … Mezzenga, R. (2012). Measurement of intrinsic properties of amyloid fibrils by the peak force QNM method. *Nanoscale*, *4*(15), 4426–9. https://doi.org/10.1039/c2nr30768e

Adamcik, J., & Mezzenga, R. (2011). Adjustable twisting periodic pitch of amyloid fibrils. *Soft Matter*, *7*(11), 5437–5443. https://doi.org/Doi 10.1039/C1sm05382e

Adamcik, J., & Mezzenga, R. (2012). Study of amyloid fibrils via atomic force microscopy. *Current Opinion in Colloid & Interface Science*, *17*(6), 369–376. https://doi.org/10.1016/j.cocis.2012.08.001

Al-Hilaly, Y. K., Biasetti, L., Blakeman, B. J. F., Pollack, S. J., Zibaee, S., Abdul-Sada, A., … Radford, S. E. (2016). The involvement of dityrosine crosslinking in α-synuclein assembly and deposition in Lewy Bodies in Parkinson's disease. *Scientific Reports*, *6*(July), 39171. https://doi.org/10.1038/srep39171

Amylin, H., Tracz, S. M., Abedini, A., Driscoll, M., & Raleigh, D. P. (2004). Role of Aromatic Interactions in Amyloid Formation by Peptides Derived from, 15901–15908.

Andrews, J. M., & Roberts, C. J. (2007). A Lumry-Eyring nucleated polymerization model of protein aggregation kinetics: 1. Aggregation with pre-equilibrated unfolding. *Journal of Physical Chemistry B*, *111*(27), 7897–7913. https://doi.org/10.1021/jp070212j

Annamalai, K., Gührs, K. H., Koehler, R., Schmidt, M., Michel, H., Loos, C., … Fändrich, M. (2016). Polymorphism of Amyloid Fibrils in Vivo. *Angewandte Chemie - International Edition*, *55*(15), 4822–4825. https://doi.org/10.1002/anie.201511524

Askarieh, G., Hedhammar, M., Nordling, K., Saenz, A., Casals, C., Rising, A., … Knight, S. D. (2011). Self-assembly of spider silk proteins is controlled by a pH-sensitive relay. *Nature*, *465*(7295), 236–238. https://doi.org/10.1038/nature08962

Bai, J., Zhang, Z., Liu, M., & Li, C. (2015). A-Synuclein-Lanthanide Metal Ions Interaction: Binding Sites, Conformation and Fibrillation. *BMC Biophysics*, *9*(1), 1. https://doi.org/10.1186/s13628-016-0026-1

Barnhart, M. M., & Chapman, M. R. (2006). Curli Biogenesis and Function, *60*, 131–147. https://doi.org/10.1146/annurev.micro.60.080805.142106.Curli

Baxa, U., Keller, P. W., Cheng, N., Wall, J. S., & Steven, A. C. (2011). In Sup35p filaments (the [PSI+] prion), the globular C-terminal domains are widely offset from the amyloid fibril

backbone. *Molecular Microbiology*, *79*(2), 523–532. https://doi.org/10.1111/j.1365-2958.2010.07466.x

Berryman, J. T., Radford, S. E., & Harris, S. A. (2009). Thermodynamic description of polymorphism in Q- and N-rich peptide aggregates revealed by atomistic simulation. *Biophysical Journal*, *97*(1), 1–11. https://doi.org/10.1016/j.bpj.2009.03.062

Bradley, M. E., Edskes, H. K., Hong, J. Y., Wickner, R. B., & Liebman, S. W. (2002). Interactions among prions and prion "strains" in yeast. *Proceedings of the National Academy of Sciences of the United States of America*, *99 Suppl 4*, 16392–16399. https://doi.org/10.1073/pnas.152330699

Bucciantini, M., Calloni, G., Chiti, F., Formigli, L., Nosi, D., Dobson, C. M., & Stefani, M. (2004a). Prefibrillar amyloid protein aggregates share common features of cytotoxicity. *Journal of Biological Chemistry*, *279*(30), 31374–31382. https://doi.org/10.1074/jbc.M400348200

Bucciantini, M., Calloni, G., Chiti, F., Formigli, L., Nosi, D., Dobson, C. M., & Stefani, M. (2004b). Prefibrillar amyloid protein aggregates share common features of cytotoxicity. *The Journal of Biological Chemistry*, *279*(30), 31374–82. https://doi.org/10.1074/jbc.M400348200

Busciglio, J., Lorenzo, a, Yeh, J., & Yankner, B. a. (1995). Beta-Amyloid Fibrils Induce Tau Phosphorylation and Loss of Microtubule Binding. *Neuron*, *14*(4), 879–88.

Byrne, L. J., Cole, D. J., Cox, B. S., Ridout, M. S., Morgan, B. J. T., & Tuite, M. F. (2009). The number and transmission of [PSI+] prion seeds (propagons) in the yeast Saccharomyces cerevisiae. *PLoS ONE*, *4*(3). https://doi.org/10.1371/journal.pone.0004670

Campioni, S., Mannini, B., Zampagni, M., Pensalfini, A., Parrini, C., Evangelisti, E., … Chiti, F. (2010). A causative link between the structure of aberrant protein oligomers and their toxicity. *Nature Chemical Biology*, *6*(2), 140–147. https://doi.org/10.1038/nchembio.283

Castelletto, V., Hamley, I. W., Cenker, Ç., Olsson, U., Adamcik, J., Mezzenga, R., … Rodríguez-Llansola, F. (2011). Influence of end-capping on the self-assembly of model amyloid peptide fragments. *Journal of Physical Chemistry B*, *115*(9), 2107–2116. https://doi.org/10.1021/jp111168s

Castro, C. E., Dong, J., Boyce, M. C., Lindquist, S., & Lang, M. J. (2011). Physical properties of polymorphic yeast prion amyloid fibers. *Biophysical Journal*, *101*(2), 439–448. https://doi.org/10.1016/j.bpj.2011.06.016

Chapman, M. R., Robinson, L. S., Pinkner, J. S., Roth, R., Hammar, M., Normark, S., & Hultgren, S. J. (2010). Role of Escherichia coli Curli Operons in Directing Amyloid Fiber Formation. *Science*, *295*(5556), 851–855. https://doi.org/10.1126/science.1067484.Role

Chiti, F., & Dobson, C. M. (2006). Protein Misfolding, Functional Amyloid, and Human Disease. *Annual Review of Biochemistry*, *75*(1), 333–366. https://doi.org/10.1146/annurev.biochem.75.101304.123901

Collins, S. R., Douglass, A., Vale, R. D., & Weissman, J. S. (2004). Mechanism of Prion Propagation: Amyloid Growth Occurs by Monomer Addition. *PLoS Biol*, *2*(10), e321. https://doi.org/10.1371/journal.pbio.0020321

Costa, T. R. D., Felisberto-rodrigues, C., Meir, A., Prevost, M. S., Redzej, A., Trokter, M., & Waksman, G. (2015). Secretion systems in Gram-negative insights. *Nature Publishing Group*, *13*(6), 343–359. https://doi.org/10.1038/nrmicro3456

Cox, B. S. (1965). [PSI] A cytoplasmic suppressor of super-supressor in yeast. *Heredity*, *20*, 505–521.

Cox, B. S., Byrne, L. J., & Tuite, M. F. (2007). Prion Stability. *Prion*, *1*(3), 170–178.

https://doi.org/4839 [pii]

Dobson, C. M. (1999). Protein misfolding, evolution and disease, *4*(September), 329–332.

Dobson, C. M. (2001). The structural basis of protein folding and its links with human disease. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *356*(1406), 133–45. https://doi.org/10.1098/rstb.2000.0758

Dobson, C. M. (2003). Protein folding and misfolding, *426*(December).

Dueholm, M. S., Nielsen, S. B., Hein, K. L., Nissen, P., Chapman, M., Christiansen, G., … Otzen, D. E. (2013). Fibrillation of the Major Curli Subunit CsgA under a Wide Range of Conditions Implies a Robust Design of Aggregation. *Biochemistry*, *50*(39), 8281–8290. https://doi.org/10.1021/bi200967c.Fibrillation

Engel, M. F. M., Khemtémourian, L., Kleijer, C. C., Meeldijk, H. J. D., Jacobs, J., Verkleij, A. J., … Höppener, J. W. M. (2008). Membrane damage by human islet amyloid polypeptide through fibril growth at the membrane. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(16), 6033–8. https://doi.org/10.1073/pnas.0708354105

Evans, M. L., Chorell, E., Almqvist, F., Chapman, M. R., Go, A., Li, F., … Chapman, M. R. (2015). The Bacterial Curli System Possesses a Potent and Selective Inhibitor of Amyloid Formation Article The Bacterial Curli System Possesses a Potent and Selective Inhibitor of Amyloid Formation, 445–455. https://doi.org/10.1016/j.molcel.2014.12.025

Fändrich, M., Meinhardt, J., & Grigorieff, N. (2009). Structural polymorphism of Alzheimer Abeta and other amyloid fibrils. *Prion*, *3*(June), 89–93. https://doi.org/8859 [pii]

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. *Nature Biotechnology*, *22*(10), 1302–6. https://doi.org/10.1038/nbt1012

Ferreira, P. C., Ness, F., Edwards, S. R., Cox, B. S., & Tuite, M. F. (2001). The elimination of the yeast [PSI+] prion by guanidine hydrochloride is the result of Hsp104 inactivation. *Molecular Microbiology*, *40*(6), 1357–1369. https://doi.org/10.1046/j.1365-2958.2001.02478.x

Ferrone, F. (1999). Analysis of protein aggregation kinetics. In R. W. B. T.-M. in Enzymology (Ed.), *Amyloid, Prions, and Other Protein Aggregates* (Vol. Volume 309, pp. 256–274). Academic Press. https://doi.org/http://dx.doi.org/10.1016/S0076-6879(99)09019-9

Fitzpatrick, A. W. P., Debelouchina, G. T., Bayro, M. J., Clare, D. K., & Caporini, M. A. (2013). Atomic structure and hierarchical assembly of a cross-β amyloid fibril, *5590*(4). https://doi.org/10.1073/pnas.1219476110/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1219476110

Fitzpatrick, A. W. P., Debelouchina, G. T., Dobson, C. M., & Et.al. (2013). Atomic structure and hierarchical assembly of a cross - β amyloid fibril. *PNAS*, *5590*(4). https://doi.org/10.1073/pnas.1219476110/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1219476110

Fitzpatrick, A. W. P., Vanacore, G. M., & Zewail, A. H. (2015). Nanomechanics and intermolecular forces of amyloid revealed by four-dimensional electron microscopy. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(11), 3380–5. https://doi.org/10.1073/pnas.1502214112

Friedman, L., Kolter, R., & Branda, S. S. (2005). Biofilms : the matrix revisited. *TRENDS in Microbiology*, *13*(1). https://doi.org/10.1016/j.tim.2004.11.006

Giurleo, J. T., He, X., & Talaga, D. S. (2008). ??-Lactoglobulin Assembles into Amyloid through

Sequential Aggregated Intermediates. *Journal of Molecular Biology*, *381*(5), 1332–1348. https://doi.org/10.1016/j.jmb.2008.06.043

Glabe, C. G. (2006). Common mechanisms of amyloid oligomer pathogenesis in degenerative disease. *Neurobiology of Aging*, *27*(4), 570–575. https://doi.org/10.1016/j.neurobiolaging.2005.04.017

Goldgaber D, Lerman MI, McBride OW, Saffiotti U, G. D. (1987). Characterization and chromosomal localization of a cDNA encoding brain amyloid of Alzheimer's disease. *Science*, *235*, 877–80.

Goldsbury, C. S., Cooper, G. J., Goldie, K. N., Müller, S. a, Saafi, E. L., Gruijters, W. T., … Kistler, J. (1997). Polymorphic fibrillar assembly of human amylin. *Journal of Structural Biology*, *119*(1), 17–27. https://doi.org/10.1006/jsbi.1997.3858

Goldsbury, C. S., Cooper, G. J. S., Goldie, K. N., & Mu, S. A. (1997). Polymorphic Fibrillar Assembly of Human Amylin, *27*, 17–27.

Goyal, P., Krasteva, P. V, Gerven, N. Van, Gubellini, F., Broeck, I. Van Den, Pinkner, J. S., … Remaut, H. (2014). Structural and mechanistic insights into the bacterial amyloid secretion channel CsgG. *Nature*, *516*, 250. https://doi.org/10.1038/nature13768

Greenwald, J., & Riek, R. (2010). Biology of amyloid: Structure, function, and regulation. *Structure*, *18*(10), 1244–1260. https://doi.org/10.1016/j.str.2010.08.009

H, O., Strain, E., Weiss-muszkat, M., Shakh, D., Zhou, Y., Pinto, R., … Sela, S. (2010). Biofilm Formation by and Multicellular Behavior of Escherichia coli, *76*(5), 1545–1554. https://doi.org/10.1128/AEM.01395-09

Halfmann, R., Alberti, S., Krishnan, R., Lyle, N., O'Donnell, C. W., King, O. D., … Lindquist, S. (2011). Opposing Effects of Glutamine and Asparagine Govern Prion Formation by Intrinsically Disordered Proteins. *Molecular Cell*, *43*(1), 72–84. https://doi.org/10.1016/j.molcel.2011.05.013

Hall-stoodley, L., & Stoodley, P. (2009). Microreview Evolving concepts in biofilm infections. *Cellular Microbiology*, *11*(April), 1034–1043. https://doi.org/10.1111/j.1462-5822.2009.01323.x

Hammer, N. D., Mcguffie, B. A., Zhou, Y., Badtke, M. P., Reinke, A. A., Brännström, K., … Chapman, M. R. (2012). The C-Terminal Repeating Units of CsgB Direct Bacterial Functional Amyloid Nucleation. *Journal of Molecular Biology*, *422*(3), 376–389. https://doi.org/10.1016/j.jmb.2012.05.043

Hammer, N. D., Schmidt, J. C., & Chapman, M. R. (2007). The curli nucleator protein , CsgB , contains an amyloidogenic domain that directs CsgA polymerization.

Han Zhang, Qilin Ma, Yun-wu Zhang, and H. X. (2012). Proteolytic processing of Alzheimer's β-amyloid precursor protein. *J Neurochem.*, *120*(Suppl 1), 9–21. https://doi.org/10.1111/j.1471-4159.2011.07519.x.Proteolytic

Hardy, J. (2009). The amyloid hypothesis for Alzheimer's disease: A critical reappraisal. *Journal of Neurochemistry*, *110*(4), 1129–1134. https://doi.org/10.1111/j.1471-4159.2009.06181.x

Herrup, K. (2015). The case for rejecting the amyloid cascade hypothesis. *Nat Neurosci*, *18*(6), 794–799. https://doi.org/10.1038/nn.4017

Jahn, T. R., Makin, O. S., Morris, K. L., Marshall, K. E., Tian, P., Sikorski, P., & Serpell, L. C. (2010a). The common architecture of cross-beta amyloid. *Journal of Molecular Biology*, *395*(4), 717–27. https://doi.org/10.1016/j.jmb.2009.09.039

Jahn, T. R., Makin, O. S., Morris, K. L., Marshall, K. E., Tian, P., Sikorski, P., & Serpell, L. C. (2010b). The Common Architecture of Cross-β Amyloid. *Journal of Molecular Biology*, *395*(4), 717–727. https://doi.org/10.1016/j.jmb.2009.09.039

Kad, N. M., Myers, S. L., Smith, D. P., Smith, D. A., Radford, S. E., & Thomson, N. H. (2003). Hierarchical assembly of beta2-microglobulin amyloid in vitro revealed by atomic force microscopy. *Journal of Molecular Biology*, *330*(3), 785–797. https://doi.org/10.1016/S0022-2836(03)00583-7

Karran, E., Mercken, M., & De Strooper, B. (2011). The amyloid cascade hypothesis for Alzheimer's disease: an appraisal for the development of therapeutics. *Nature Reviews. Drug Discovery*, *10*(9), 698–712. https://doi.org/10.1038/nrd3505

Kim, J., Chakrabarty, P., Hanna, A., March, A., Dickson, D. W., Borchelt, D. R., … Janus, C. (2013). Normal cognition in transgenic BRI2-Aβ mice. *Molecular Neurodegeneration*, *8*(1), 15. https://doi.org/10.1186/1750-1326-8-15

Knowles, T. P. J., & Buehler, M. J. (2011). Nanomechanics of functional and pathological amyloid materials. *Nature Nanotechnology*, *6*(8), 469–479. https://doi.org/10.1038/nnano.2011.102

Krishnan, R., & Lindquist, S. L. (2005). Structural insights into a yeast prion illuminate nucleation and strain diversity. *Nature*, *435*(June), 765–772. https://doi.org/nature03679 [pii]\r10.1038/nature03679

Kryndushkin, D. S., Alexandrov, I. M., Ter-Avanesyan, M. D., & Kushnirov, V. V. (2003). Yeast [PSI+] prion aggregates are formed by small Sup35 polymers fragmented by Hsp104. *The Journal of Biological Chemistry*, *278*(49), 49636–49643. https://doi.org/10.1074/jbc.M307996200

Kumar, J., Tsumatori, H., Yuasa, J., Kawai, T., & Nakashima, T. (2015). Self-discriminating termination of chiral supramolecular polymerization: Tuning the length of nanofibers. *Angewandte Chemie - International Edition*, *54*(20), 5943–5947. https://doi.org/10.1002/anie.201500292

Kurland, N. E., Drira, Z., & Yadavalli, V. K. (2012). Measurement of nanomechanical properties of biomolecules using atomic force microscopy. *Micron*, *43*(2–3), 116–128. https://doi.org/10.1016/j.micron.2011.07.017

Kurouski, D., Dukor, R. K., Lu, X., Nafie, L. A., & Lednev, I. K. (2012). Normal and reversed supramolecular chirality of insulin fibrils probed by vibrational circular dichroism at the protofilament level of fibril structure. *Biophysical Journal*, *103*(3), 522–531. https://doi.org/10.1016/j.bpj.2012.04.042

Kurouski, D., Lu, X., Stubbs, G., Dukor, R. K., Lednev, I. K., & Nafie, L. A. (2013). Is Supramolecular Filament Chirality the Underlying Cause of Major Morphology Differences in Amyloid Fibrils? https://doi.org/10.1021/ja407583r

Kushnirov, V. V., & Ter-Avanesyan, M. D. (1998). Structure and replication of yeast prions. *Cell*, *94*(1), 13–16. https://doi.org/10.1016/S0092-8674(00)81216-7

Lara, C., Reynolds, N. P., Berryman, J. T., Xu, A., Zhang, A., & Mezzenga, R. (2014). ILQINS hexapeptide, identified in lysozyme left-handed helical ribbons and nanotubes, forms right-handed helical ribbons and crystals. *Journal of the American Chemical Society*, *136*(12), 4732–4739. https://doi.org/10.1021/ja500445z

Lewis, L., Walden, H., & Hydrophobic, S. L. C. (2014). The relationship between amyloid structure and cytotoxicity, *8*(2), 192–196.

Liebman, S. W., & Chernoff, Y. O. (2012). Prions in yeast. *Genetics*, *191*(4), 1041–1072.

https://doi.org/10.1534/genetics.111.137760

Linse, S., Cabaleiro-Lago, C., Xue, W.-F., Lynch, I., Lindman, S., Thulin, E., … Dawson, K. a. (2007). Nucleation of protein fibrillation by nanoparticles. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(21), 8691–8696. https://doi.org/10.1073/pnas.0701250104

Liu, J., Costantino, I., Venugopalan, N., Fischetti, R. F., Hyman, B. T., Frosch, M. P., … Makowski, L. (2016). Amyloid structure exhibits polymorphism on multiple length scales in human brain tissue. *Nature Publishing Group*, (September), 1–11. https://doi.org/10.1038/srep33079

Lu, J.-X., Qiang, W., Yau, W.-M., Schwieters, C. D., Meredith, S. C., & Tycko, R. (2013). Molecular structure of β-amyloid fibrils in Alzheimer's disease brain tissue. *Cell*, *154*(6), 1257–68. https://doi.org/10.1016/j.cell.2013.08.035

Luckgei, N., Sch??tz, A. K., Bousset, L., Habenstein, B., Sourigues, Y., Gardiennet, C., … B??ckmann, A. (2013). The conformation of the prion domain of Sup35 p in isolation and in the full-length protein. *Angewandte Chemie - International Edition*, *52*(48), 12741–12744. https://doi.org/10.1002/anie.201304699

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*(2), 315–320. https://doi.org/10.1073/pnas.0406847102

Malato, L., Reis, S. Dos, Benkemoun, L., Sabate, R., & Saupe, S. J. (2007). Role of Hsp104 in the Propagation and Inheritance of the [Het-s] Prion. *Molecular Biology of the Cell*, *18*(1), 4803–4812. https://doi.org/10.1091/mbc.E07

Marshall, K. E., Marchante, R., Xue, W.-F., & Serpell, L. C. (2014). The relationship between amyloid structure and cytotoxicity. *Prion*, *8*(2), 192–196. https://doi.org/10.4161/pri.28860

Marshall, K. E., Morris, K. L., Charlton, D., & O'Reilly, N. (2011). Hydrophobic, Aromatic, and Electrostatic Interactions Play a Central Role in Amyloid Fibril Formation and Stability - Biochemistry (ACS Publications). *Biochemistry*, 2061–2071. Retrieved from http://pubs.acs.org/doi/abs/10.1021/bi101936c%5Cnpapers2://publication/uuid/8F74A16B-A7E3-4CBB-9074-28DE6A3710AC

Marshall, K. E., & Serpell, L. C. (2009). Insights into the Structure of Amyloid Fibrils. *The Open Biology Journal*, *2*, 185–192.

Maurer-Stroh, S., Debulpaep, M., Kuemmerer, N., Lopez de la Paz, M., Martins, I. C., Reumers, J., … Rousseau, F. (2010). Exploring the sequence determinants of amyloid structure using position-specific scoring matrices. *Nature Methods*, *7*(3), 237–42. https://doi.org/10.1038/nmeth.1432

Mc Donald, J. M., Savva, G. M., Brayne, C., Welzel, A. T., Forster, G., Shankar, G. M., … Walsh, D. M. (2010). The presence of sodium dodecyl sulphate-stable Aβ dimers is strongly associated with Alzheimer-type dementia. *Brain*, *133*(5), 1328–1341. https://doi.org/10.1093/brain/awq065

Meinhardt, J., Sachse, C., Hortschansky, P., Grigorieff, N., & Fändrich, M. (2009). Aβ(1-40) Fibril Polymorphism Implies Diverse Interaction Patterns in Amyloid Fibrils. *Journal of Molecular Biology*, *386*(3), 869–877. https://doi.org/10.1016/j.jmb.2008.11.005

Milanesi, L., Sheynis, T., Xue, W.-F., Orlova, E. V, Hellewell, A. L., Jelinek, R., … Saibil, H. R. (2012). Direct three-dimensional visualization of membrane disruption by amyloid fibrils. *Proceedings of the National Academy of Sciences*, *109*(50), 20455–20460.

https://doi.org/10.1073/pnas.1206325109

Monsellier, E., Bousset, L., & Melki, R. (2016). α-Synuclein and huntingtin exon 1 amyloid fibrils bind laterally to the cellular membrane. *Scientific Reports*, *6*(January), 19180. https://doi.org/10.1038/srep19180

Morris, K. L., Rodger, A., Hicks, M. R., Debulpaep, M., Schymkowitz, J., Rousseau, F., & Serpell, L. C. (2013). Exploring the sequence-structure relationship for amyloid peptides. *The Biochemical Journal*, *450*(2), 275–83. https://doi.org/10.1042/BJ20121773

Nelson, R., Sawaya, M. R., Balbirnie, M., Madsen, A. Ø., Grothe, R., & Eisenberg, D. (2005). Structure of the cross-β spine of amyloid-like fibrils. *Nature*, *435*(7043), 773–778.

Nenninger, A. A., Robinson, L. S., Hammer, N. D., Epstein, E. A., Badtke, M. P., Hultgren, S. J., & Chapman, M. R. (2011). CsgE is a curli secretion specificity factor that prevents amyloid fibre aggregation. *Molecular Microbiology*, *81*(June), 486–499. https://doi.org/10.1111/j.1365-2958.2011.07706.x

Nenninger, A. A., Robinson, L. S., & Hultgren, S. J. (2009). Localized and efficient curli nucleation requires the chaperone-like amyloid assembly protein CsgF. *PNAS*, *106*(3).

Oliveberg, M. (2010). Waltz, an exciting new move in amyloid prediction. *Nature Methods*, *7*(3), 187–188. https://doi.org/10.1038/nmeth0310-187

Park, J., Kahng, B., & Hwang, W. (2009a). Thermodynamic selection of steric zipper patterns in the amyloid cross-?? spine. *PLoS Computational Biology*, *5*(9). https://doi.org/10.1371/journal.pcbi.1000492

Park, J., Kahng, B., & Hwang, W. (2009b). Thermodynamic Selection of Steric Zipper Patterns in the Amyloid Cross- b Spine, *5*(9). https://doi.org/10.1371/journal.pcbi.1000492

Park, Y. N., Zhao, X., Yim, Y. I., Todor, H., Ellerbrock, R., Reidy, M., … Greene, L. E. (2014). Hsp104 overexpression cures Saccharomyces cerevisiae [PSI+] by causing dissolution of the prion seeds. *Eukaryotic Cell*, *13*(5), 635–647. https://doi.org/10.1128/EC.00300-13

Paushkin, S. V, Kushnirov, V. V, Smirnov, V. N., & Ter-Avanesyan, M. D. (1996). Propagation of the yeast prion-like [psi+] determinant is mediated by oligomerization of the SUP35-encoded polypeptide chain release factor. *The EMBO Journal*, *15*(12), 3127–3134. https://doi.org/citeulike-article-id:10067579

Pellarin, R., & Caflisch, A. (2006). Interpreting the aggregation kinetics of amyloid peptides. *Journal of Molecular Biology*, *360*(4), 882–92. https://doi.org/10.1016/j.jmb.2006.05.033

Petkova, A. T., Leapman, R. D., Guo, Z., Yau, W.-M., Mattson, M. P., & Tycko, R. (2005). Self-Propagating, Molecular-Level Polymorphism in Alzheimers ß-Amyloid Fibrils. *Science*, *307*(5707), 262 LP-265. Retrieved from http://science.sciencemag.org/content/307/5707/262.abstract

Prusiner, S. B. (1982). Novel Proteinaceous Infectious Particles Cause Scrapie Stanley. *Science*, *216*(April).

Prusiner, S. B., McKinley, M. P., Bowman, K. A., Bolton, D. C., Bendheim, P. E., Groth, D. F., & Glenner, G. G. (1983). Scrapie prions aggregate to form amyloid-like birefringent rods. *Cell*, *35*(2), 349–358. https://doi.org/10.1016/0092-8674(83)90168-X

Reitz, C. (2012). Alzheimer's disease and the amyloid cascade hypothesis: A critical review. *International Journal of Alzheimer's Disease*, *2012*. https://doi.org/10.1155/2012/369808

Relini, A., Torrassa, S., Ferrando, R., Rolandi, R., Campioni, S., Chiti, F., & Gliozzi, A. (2010). Detection of populations of amyloid-like protofibrils with different physical properties. *Biophysical Journal*, *98*(7), 1277–1284. https://doi.org/10.1016/j.bpj.2009.11.052

Reumers, J., Maurer-Stroh, S., Schymkowitz, J., & Rousseau, F. (2009). Protein sequences encode safeguards against aggregation. *Human Mutation*, *30*(3), 431–437. https://doi.org/10.1002/humu.20905

Ritter, C., Adrian, M., Riek-loher, D., Bohrmann, B., Do, H., Schubert, D., & Riek, R. (2005). 3D structure of Alzheimer ' s amyloid-beta (1 – 42) fibrils. *PNAS*, *102*(48), 17347.

Rochet, J. C., & Lansbury, P. T. (2000). Amyloid fibrillogenesis: Themes and variations. *Current Opinion in Structural Biology*, *10*(1), 60–68. https://doi.org/10.1016/S0959-440X(99)00049-4

Rochet, J. C., & Lansbury, P. T. (2000). Amyloid fibrillogenesis: themes and variations. *Current Opinion in Structural Biology*, *10*(1), 60–8.

Rohit V. Pappu*, Xiaoling Wang, Andreas Vitalis, S. L. C. (2008). A polymer physics perspective on driving forces and mechanisms for protein aggregation. *Arch Biochem Biophys*, *469*(1), 132–141.

Ross, E. D., Edskes, H. K., Terry, M. J., & Wickner, R. B. (2005). Primary sequence independence for prion formation. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(36), 12825–12830. https://doi.org/10.1073/pnas.0506136102

Ryu, J., & Beuchat, L. R. (2005). Biofilm Formation by Escherichia coli O157 : H7 on Stainless Steel : Effect of Exopolysaccharide and Curli Production on Its Resistance to Chlorine, *71*(1), 247–254. https://doi.org/10.1128/AEM.71.1.247

Scheibel, T., Parthasarathy, R., Sawicki, G., Lin, X.-M., Jaeger, H., & Lindquist, S. L. (2003). Conducting nanowires built by controlled self-assembly of amyloid fibers and selective metal deposition. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(8), 4527–32. https://doi.org/10.1073/pnas.0431081100

Schuler, B. (1999). Formation of Fibrous Aggregates from a Non-native Intermediate: The Isolated P22 Tailspike beta -Helix Domain. *Journal of Biological Chemistry*, *274*(26), 18589–18596. https://doi.org/10.1074/jbc.274.26.18589

Serpell, L. C. (2000). Alzheimer â€™ s amyloid fibrils : structure and assembly. *Biochimica et Biophysica Acta*, *1502*, 16–30.

Serpell, L. C., & Smith, J. M. (2000). Direct visualisation of the beta-sheet structure of synthetic Alzheimer's amyloid. *J Mol Biol*, *299*(1), 225–231. https://doi.org/10.1006/jmbi.2000.3650

Seviour, T., Hansen, S. H., Yang, L., Yau, Y. H., Wang, V. B., Stenvang, M. R., … Dueholm, M. S. (2015). Functional amyloids keep quorum-sensing molecules in check. *Journal of Biological Chemistry*, *290*(10), 6457–6469. https://doi.org/10.1074/jbc.M114.613810

Shorter, J., & Lindquist, S. (2004). Hsp104 catalyzes formation and elimination of self-replicating Sup35 prion conformers. *Science (New York, N.Y.)*, *304*(1995), 1793–7. https://doi.org/10.1126/science.1098007

Sidhu, A., Segers-Nolten, I., & Subramaniam, V. (2014). Solution conditions define morphological homogeneity of α-synuclein fibrils. *Biochimica et Biophysica Acta - Proteins and Proteomics*, *1844*(12), 2127–2134. https://doi.org/10.1016/j.bbapap.2014.09.007

Sivanathan, V., & Hochschild, A. (2012). Generating extracellular amyloid aggregates using E. coli cells. *Genes and Development*, *26*(23), 2659–2667. https://doi.org/10.1101/gad.205310.112

Sivanathan, V., & Hochschild, A. (2013). A bacterial export system for generating extracellular amyloid aggregates. *Nature Protocols*, *8*(7), 1381–1390.

https://doi.org/10.1097/MPG.0b013e3181a15ae8.Screening

Smith, J. F., Knowles, T. P. J., Dobson, C. M., Macphee, C. E., & Welland, M. E. (2006). Characterization of the nanoscale properties of individual amyloid fibrils. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(43), 15806–15811. https://doi.org/10.1073/pnas.0604035103

Smith, R. A. S., Nabok, A., Blakeman, B. J. F., Xue, W. F., Abell, B., & Smith, D. P. (2015). Analysis of toxic amyloid fibril interactions at natively derived membranes by ellipsometry. *PLoS ONE*, *10*(7), 1–15. https://doi.org/10.1371/journal.pone.0132309

Sneideris, T., Milto, K., & Smirnovas, V. (2015). Polymorphism of amyloid-like fibrils can be defined by the concentration of seeds. *PeerJ*, *3*, e1207. https://doi.org/10.7717/peerj.1207

Sunde, M., Serpell, L. C., Bartlam, M., Fraser, P. E., Pepys, M. B., & Blake, C. C. . (1997). Common core structure of amyloid fibrils by synchrotron X-ray diffraction. *Journal of Molecular Biology*, *273*(3), 729–739. https://doi.org/10.1006/jmbi.1997.1348

Supnet, C., & Bezprozvanny, I. (2010). The dysregulation of intracellular calcium in Alzheimer disease. *Cell Calcium*, *47*(2), 183–189. https://doi.org/10.1016/j.ceca.2009.12.014

Suzuki N, Cheung TT, Cai XD, Odaka A, Otvos L Jr, Eckman C, Golde TE, Y. S. (1994). An increased percentage of long amyloid beta protein secreted by familial amyloid beta protein precursor (beta APP717) mutants. *Science*, *264*, 1336–40.

Tanaka, M., Collins, S. R., Toyama, B. H., & Weissman, J. S. (2006). The physical basis of how prion conformations determine strain phenotypes. *Nature*, *442*(7102), 585–9. https://doi.org/10.1038/nature04922

Taran, E., Gentle, I. R., Gobius, K. S., & Dykes, G. A. (2011). CsgA Production by Escherichia coli O157 : H7 Alters Attachment to Abiotic Surfaces in Some Growth Environments □, *77*(20), 7339–7344. https://doi.org/10.1128/AEM.00277-11

Tipping, K. W., van Oosten-Hawle, P., Hewitt, E. W., & Radford, S. E. (2015). Amyloid Fibres: Inert End-Stage Aggregates or Key Players in Disease? *Trends in Biochemical Sciences*, *40*(12), 719–727. https://doi.org/10.1016/j.tibs.2015.10.002

Tseng, B. P., Green, K. N., Chan, J. L., Blurton-Jones, M., & LaFerla, F. M. (2008). Abeta inhibits the proteasome and enhances amyloid and tau accumulation. *Neurobiology of Aging*, *29*(11), 1607–18. https://doi.org/10.1016/j.neurobiolaging.2007.04.014

Tu, L. H., & Raleigh, D. P. (2013). Role of aromatic interactions in amyloid formation by islet amyloid polypeptide. *Biochemistry*, *52*(2), 333–342. https://doi.org/10.1021/bi3014278

Tycko, R. (2014). Physical and structural basis for polymorphism in amyloid fibrils. *Protein Science*, *23*(11), 1528–1539. https://doi.org/10.1002/pro.2544

Tyedmers, J., Treusch, S., Dong, J., McCaffery, J. M., Bevis, B., & Lindquist, S. (2010). Prion induction involves an ancient system for the sequestration of aggregated proteins and heritable changes in prion fragmentation. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(19), 8633–8. https://doi.org/10.1073/pnas.1003895107

Uptain, S. M., Sawicki, G. J., Caughey, B., & Lindquist, S. (2001). Strains of [PSI+] are distinguished by their efficiencies of prion-mediated conformational conversion. *EMBO Journal*, *20*(22), 6236–6245. https://doi.org/10.1093/emboj/20.22.6236

Usov, I., Adamcik, J., & Mezzenga, R. (2013). Polymorphism complexity and handedness inversion in serum albumin amyloid fibrils. *ACS Nano*, *7*(12), 10465–10474. https://doi.org/10.1021/nn404886k

Van Der Wel, P. C. A., Lewandowski, J. R., & Griffin, R. G. (2007). Solid-state NMR study of amyloid nanocrystals and fibrils formed by the peptide GNNQQNY from yeast prion protein Sup35p. *Journal of the American Chemical Society*, *129*(16), 5117–5130. https://doi.org/10.1021/ja068633m

Van Gerven, N., Klein, R. D., Hultgren, S. J., & Remaut, H. (2015). Bacterial amyloid formation: Structural insights into curli biogensis. *Trends in Microbiology*, *23*(11), 693–706. https://doi.org/10.1016/j.tim.2015.07.010

Van Melckebeke, H., Wasmer, C., Lange, A., Ab, E., Loquet, A., Böckmann, A., & Meier, B. H. (2010). Atomic-resolution three-dimensional structure of HET-s(218-289) amyloid fibrils by solid-state NMR spectroscopy. *Journal of the American Chemical Society*, *132*(39), 13765–75. https://doi.org/10.1021/ja104213j

Vandenakker, C. C., Engel, M. F. M., Velikov, K. P., Bonn, M., & Koenderink, G. H. (2011). Morphology and persistence length of amyloid fibrils are correlated to peptide molecular structure. *Journal of the American Chemical Society*, *133*(45), 18030–18033. https://doi.org/10.1021/ja206513r

Verges, K. J., Smith, M. H., Toyama, B. H., & Weissman, J. S. (2011). Strain conformation, primary structure and the propagation of the yeast prion [PSI+]. *Nat Struct Mol Biol*, *18*(4), 493–499. https://doi.org/10.1038/nsmb.2030\rnsmb.2030 [pii]

Villemagne, V. L., Pike, K. E., Chételat, G., Ellis, K. A., Mulligan, R. S., Bourgeat, P., … Rowe, C. C. (2011). Longitudinal assessment of AB and cognition in aging and Alzheimer disease. *Annals of Neurology*, *69*(1), 181–192. https://doi.org/10.1002/ana.22248

Volpatti, L. R., Vendruscolo, M., Dobson, C. M., & Knowles, T. P. J. (2013). PERSPECTIVE A Clear View of Polymorphism , Fibril Formation. *Acs Nano*, *7*(12), 10443–10448. https://doi.org/Doi 10.1021/Nn406121w

Walsh, D. M. (1999). Amyloid beta -Protein Fibrillogenesis. STRUCTURE AND BIOLOGICAL ACTIVITY OF PROTOFIBRILLAR INTERMEDIATES. *Journal of Biological Chemistry*, *274*(36), 25945–25952. https://doi.org/10.1074/jbc.274.36.25945

Wang, L., Schubert, D., Sawaya, M. R., Eisenberg, D., & Riek, R. (2010). Multidimensional structure-activity relationship of a protein in its aggregated states. *Angewandte Chemie - International Edition*, *49*(23), 3904–3908. https://doi.org/10.1002/anie.201000068

Wang, X., & Chapman, M. R. (2008). Curli provide the template for understanding controlled amyloid propagation. *Prion*, *2*(2), 57–60. https://doi.org/10.1006/jmbi.1997.1348

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

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

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

Wasmer, C., Sch??tz, A., Loquet, A., Buhtz, C., Greenwald, J., Riek, R., … Meier, B. H. (2009). The Molecular Organization of the Fungal Prion HET-s in Its Amyloid Form. *Journal of Molecular Biology*, *394*(1), 119–127. https://doi.org/10.1016/j.jmb.2009.09.015

Watanabe-Nakayama, T., Ono, K., Itami, M., Takahashi, R., Teplow, D. B., & Yamada, M. (2016). High-speed atomic force microscopy reveals structural dynamics of amyloid β $_{1–42}$ aggregates. *Proceedings of the National Academy of Sciences*, 201524807. https://doi.org/10.1073/pnas.1524807113

Wickner, R. B. (1994). [URE3] as an altered URE2 protein: evidence for a prion analog in Saccharomyces cerevisiae. *Science*, *264*(5158), 566 LP-569. Retrieved from http://science.sciencemag.org/content/264/5158/566.abstract

Wickner, R. B., Edskes, H. K., Ross, E. D., Pierce, M. M., Baxa, U., Brachmann, A., & Shewmaker, F. (2004). Prion genetics: new rules for a new kind of gene. *Genetics*, *38*(1), 681. https://doi.org/10.1146/annurev.genet.38.072902.092200

Williams, T. L., & Serpell, L. C. (2011). Membrane and surface interactions of Alzheimer's A?? peptide - Insights into the mechanism of cytotoxicity. *FEBS Journal*, *278*(20), 3905–3917. https://doi.org/10.1111/j.1742-4658.2011.08228.x

Xue, W.-F. (2014). *Bio-nanoimaging. Protein Misfolding & Aggregation*. (V. N. Uversky & Y. L. Lyubchenko, Eds.). Academic Press.

Xue, W.-F., Homans, S. W., & Radford, S. E. (2008). Systematic analysis of nucleation-dependent polymerization reveals new insights into the mechanism of amyloid self-assembly. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(26), 8926–31. https://doi.org/10.1073/pnas.0711664105

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*(49), 34272–34282. https://doi.org/10.1074/jbc.M109.049809

Xue, W. F., Homans, S. W., & Radford, S. E. (2009). Amyloid fibril length distribution quantified by atomic force microscopy single-particle image analysis. *Protein Engineering, Design and Selection*, *22*(8), 489–496. https://doi.org/10.1093/protein/gzp026

Xue, W. F., & Radford, S. E. (2013). An imaging and systems modeling approach to fibril breakage enables prediction of amyloid behavior. *Biophysical Journal*, *105*(12), 2811–2819. https://doi.org/10.1016/j.bpj.2013.10.034

Xue, W., Hellewell, A. L., Hewitt, E. W., & Radford, S. E. (2010a). Fibril fragmentation in amyloid assembly and cytotoxicity - When size matters. *Prion*, *4*(1), 20–25.

Xue, W., Hellewell, A. L., Hewitt, E. W., & Radford, S. E. (2010b). When size matters, *4*(1), 20–25.

Yoon, G., Lee, M., Kim, J. I., Na, S., & Eom, K. (2014). Role of sequence and structural polymorphism on the mechanical properties of amyloid fibrils. *PLoS ONE*, *9*(2), 1–13. https://doi.org/10.1371/journal.pone.0088502

Zangi, R., Hagen, M., & Berne, B. J. (2007). Effect of ions on the hydrophobic interaction between two plates. *Journal of the American Chemical Society*, *129*(15), 4678–4686. https://doi.org/10.1021/ja068305m

Appendix

Appendix 1

Hierarchical clustering complete script.

```
%defines distance cut-off
y=1;
%determines average particle height
s=size(fibs);
for n=1:s;
    strucH{n} = mean (fibs(n,1).z);
end
%Fast Fourier transform for periodicity
for n=1:s;
    fibs(n,1).l = fibs(n,1).l-min(fibs(n,1).l);
    ll{n} = fibs(n,1).l;
    Nz{n} = length(fibs(n,1).l);
    zz{n} = fibs(n,1).z;

end
 Nz1=cell2mat(Nz);

for n=1:s;
    Q {n} = max(fibs(n,1).l)/(Nz1(1,n)-1);
end

for n=1:s;
    zft{n} = fft(zz{1,n}, Nz{1,n})./Nz{1,n};
    f{n} = Q{1,n}/2*linspace(0,1,Nz{1,n}/2+1);
    Y1{n}= 2*abs(zft{1,n}(1:Nz{1,n}/2+1));
    X1{n}= f{1,n};
    X1{1,n}(1,1)=0;
    Y1{1,n}(1,1)=0;
    [MaxY1{n},ind{n}] = max(Y1{1,n});
    strucH{2,n}=(X1{1,n}(ind{1,n})*2.9296875);

end
%Determines max and minimum particle height
for n=1:s;
    c=flipud(unique(sort(fibs(n, 1).z)));
    result=c(1:50);
    strucMandM{1,n}= mean(result);
    c1=flip (c);
    result1=c1(1:50);
    strucMandM{2,n}= mean(result1);
end
%Determines periodic frequency
for n=1:s;
    zft{n} = fft(zz{1,n}, Nz{1,n})./Nz{1,n};
    f{n} = Q{1,n}/2*linspace(0,1,Nz{1,n}/2+1);
    Y1{n}= 1./2*abs(zft{1,n}(1:Nz{1,n}/2+1));
    X1{n}= 1./f{1,n};
    X1{1,n}(1,1)=0;
    Y1{1,n}(1,1)=0;
    [MaxY1{n},ind{n}] = max(Y1{1,n});
    strucP{n}=(X1{1,n}(ind{1,n})*2.9296875);

end
```

```
StrucHandP=cell2mat (strucH);
StrucHandP=StrucHandP.';
StrucMandM=cell2mat (strucMandM);
StrucMandM=StrucMandM.';
StrucP=cell2mat (strucP);
StrucP=StrucP.';
%Measures standardised Euclidean distance
Y=pdist(StrucHandP,'seuclidean');
%Performs complete Linkage function
Z = linkage(Y,'complete');
figure(1)
dendrogram(Z)
%Clusters Data based on distance cut off
T=cluster(Z,'cutoff',y,'criterion','distance');
figure(2)
%Returns scatter graph with colour coded clusters
scatter(StrucHandP(:,2),StrucHandP(:,1),50,T,'filled')

n = max(T);
clust(n)=struct('cluster',{{n}});
for s=1:n;
    value = (find(T==s));

    Ss=size(value);
    for N=1:Ss';
        N=N';
        X=value(N,1);
        clust(s).cluster{1,1}(N,1:2)=StrucHandP(X,1:2);
        clust(s).cluster{1,1}(N,3:4)=StrucMandM(X,1:2);
        clust(s).cluster{1,1}(N,5)=StrucP(X,1);
    end
end
clusteraverages={s,2};
for s=1:n;
 clusteraverages{s,1}= mean (clust(s).cluster{1,1}(:,1));
 clusteraverages{s,2}= mean (clust(s).cluster{1,1}(:,2));
 clusteraverages{s,3}= mean (clust(s).cluster{1,1}(:,3));
 clusteraverages{s,4}= mean (clust(s).cluster{1,1}(:,4));
 clusteraverages{s,5}= mean (clust(s).cluster{1,1}(:,5));
 B{s,1}=std((clust(s).cluster{1,1}(:,1)));
 B{s,2}=std((clust(s).cluster{1,1}(:,2)));
 B{s,3}=std((clust(s).cluster{1,1}(:,3)));
 B{s,4}=std((clust(s).cluster{1,1}(:,4)));
 B{s,5}=std((clust(s).cluster{1,1}(:,5)));
end

Clusteraverages=cell2mat (clusteraverages);
StandardD=cell2mat(B);
figure(3)
Cc=(1:n);
scatter(Clusteraverages(:,2),Clusteraverages(:,1),50,Cc,'filled')
zz=smooth(Z(:,3));
differences = diff(zz);
figure(4)
plot(differences)
figure(5)
plot(zz)

X=0.1:0.1:6;

T=cluster(Z,'cutoff',(X(1,:)),'criterion','distance');

T1=max(T(:,:));
figure(6)
```

```
plot(T1,X)

T=cluster(Z,'cutoff',y,'criterion','distance');
```

Appendix 2

Polprop script with Height and length filtering pre-script.

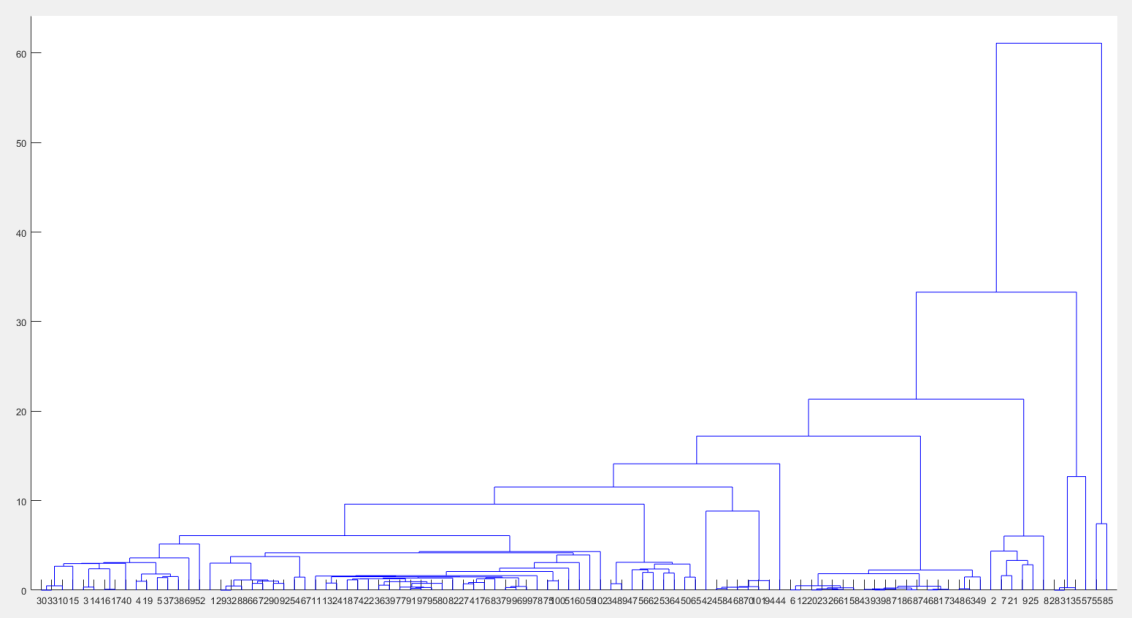
```
s=size([img.features.polymer;img.features.polymerSegment],1);
img2=[img.features.polymer;img.features.polymerSegment];
img1 = [];
for Xx=1:s;
    if (mean (img2(Xx,1).z) <20)
        img1=[img1; img2(Xx)];
    end
end

s=size(img1);
img2=img1;
img3 = [];
for Xx=1:s;
    if ((img2(Xx,1).lContour) >2)
        img3=[img3; img2(Xx)];
    end
end

%Extracts polymer information from the ans file
fib=img3;[img.features.polymer;img.features.polymerSegment];
%generates a value for the no. of polymers in the file
s=size(fib,1);
%A loop which extracts the Contour length, Height, Mode Height and
Mean
%Height from the ans file.
for n=1:s;
    l{n}=fib(n,1).lContour;
    h{n}=fib(n,1).z;
    hparmean{n}=mean(fib(n,1).z)';
    hparmode{n}=mode(fib(n,1).z)';
end
%Converts Mean Particle Height into Matrix from cell
MeanPartHeight=cell2mat(hparmean);
%Calculates the Mean of the mode particle heights
ParticleMeanHeightMean=mean2(MeanPartHeight);
%Converts Mode Particle Height into matrix from cell
ModePartHeight=cell2mat(hparmode);
%Calculates the Mean of the mode particle heights
ParticleModeHeightMean=mean2(ModePartHeight);
%Extracts the max no. of data points in each cell.
maxSize = max(cellfun(@numel,h));
%Expands the cells into a matrix padding any values up to the maximum
with
%NaN
H=cell2mat(cellfun(@(x)cat(1,x,NaN(maxSize-
length(x),1)),h,'UniformOutput',false));
%Concatenates the columns into one column
Height=vertcat(H(:),1);
%All NaNs removed
Height(any(isnan(Height),2),:) = [];
%The total no. of height data from traced fibrils
TracedHeight=size(Height);
```

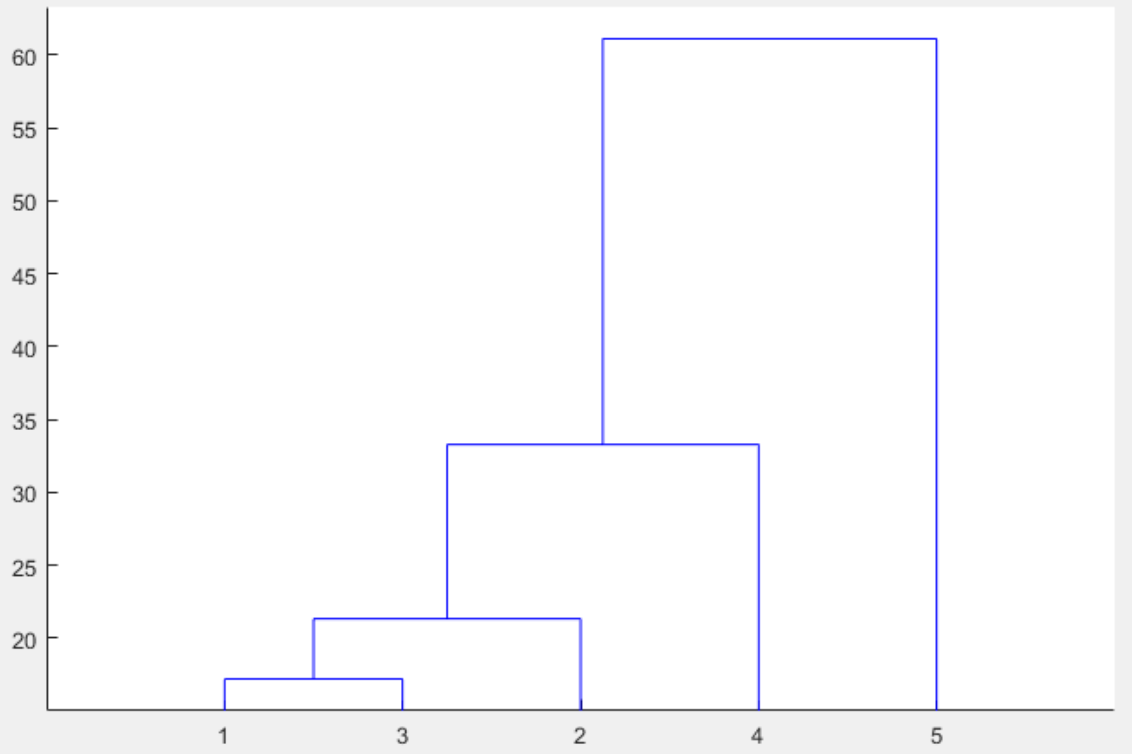
```
%Selects the first value of the matrix, removes the ,1.
TracedHeightPixels=TracedHeight(1,1);
%Finds any pixel with a value greater than 2nm, 1nm and 0.5nm
TotalHeightover2=find(img.z>2);
TotalHeightover1=find(img.z>1);
TotalHeightover0_5=find(img.z>0.5);
%Calculates the no. pixels over 2nm
NumofPixelsover2=length(TotalHeightover2);
NumofPixelsover1=length(TotalHeightover1);
NumofPixelsover0_5=length(TotalHeightover0_5);
%Converts the l cell array into a matrix Len
Len=cell2mat(l);
%Flips the horizontal orientation to vertical
Len=Len';
%Converts the no. of pixels to the length in nm (9.7704 dependent on
%resolution of AFM scanning 20um2 at 2048 x 2048 and 10uM2 at 1024x
1024 =
%9.7704
CL=Len*14.65;
%The contour length values squared
CL2=CL.^2;
%The sum of the squared contour lengths - single as data is double
SumCL2=sum(single(CL2));
%The sum of the contour lengths - single as data is double
SumCL=sum(single(CL));
%Generation of the length of the longest fibril in nm
LongestFib=max(CL);
%loop which finds particles which are of length less than or equal to
the
%value p from 1 up to the longest fibril length
for p=1:LongestFib
    CDF1{p}=find(CL<=p);
end
%Extracts the no. of particles in the cells generated from the loop
CDF2=cellfun('prodofsize', CDF1);
%Calculation of the Cumulative Distribution function - each value of
CDF2
%divided my the total no. of fibrils
CDFy=CDF2/n;
%Generation of Cumulative Distribution Function x axis in nm - upto
the
%maximum fibril length
CDFx=1:p;
%Calculation of Contour Length weighted mean
ContourLengthWeightMean=SumCL2/SumCL;
ContourLengthWeightMeanSEM=(SumCL2/SumCL)/sqrt(s);
%Calculation of Mean Contour Length
ContourLengthMean=mean2(CL);
%Calculation of standard error of mean for the Contour Lengths
ContourLengthSEM=std(CL)/sqrt(s);
%Calculation of the mean Height from all height values
HeightMean=mean2(Height)
%Calculation of the Standard Deviation for Heights (all height values)
HeightSD=std(Height);

Excel=[s,0,ContourLengthMean,ContourLengthWeightMean,0,HeightMean,Heig
htSD];
```



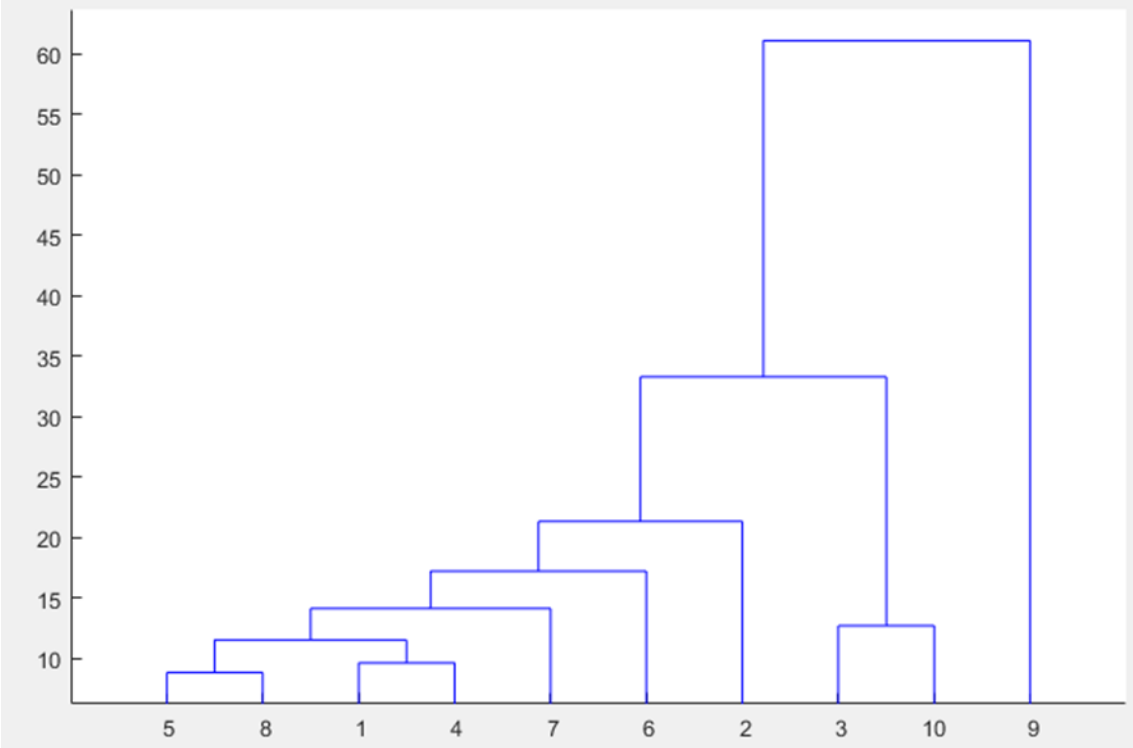
Appendix 3.

A dendrogram displaying the similarity distance (x axis) between each individual fibril particle (y axis) contained in the VIYKI characterisation dataset. Dendrogram was constructed using Matlab. This is an undesired representation of the data using a dendrogram due to the high number of links; clusters are difficult in their determination.



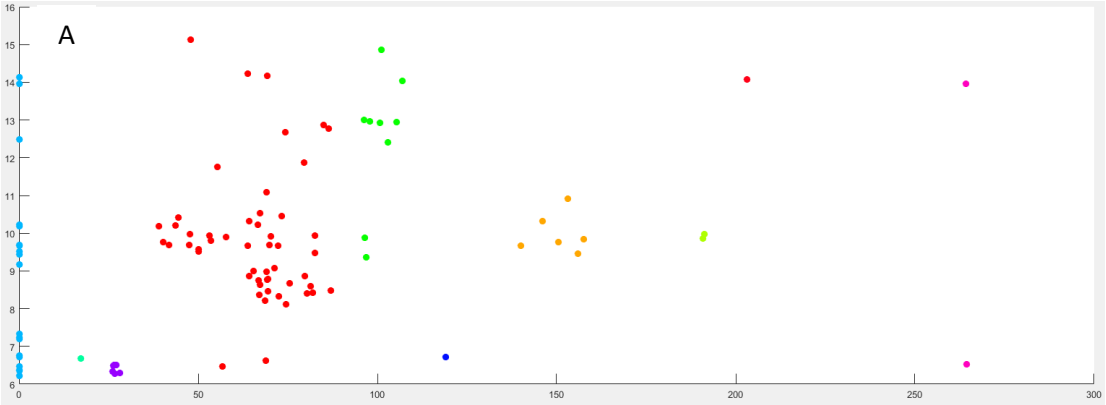
Appendix 4.

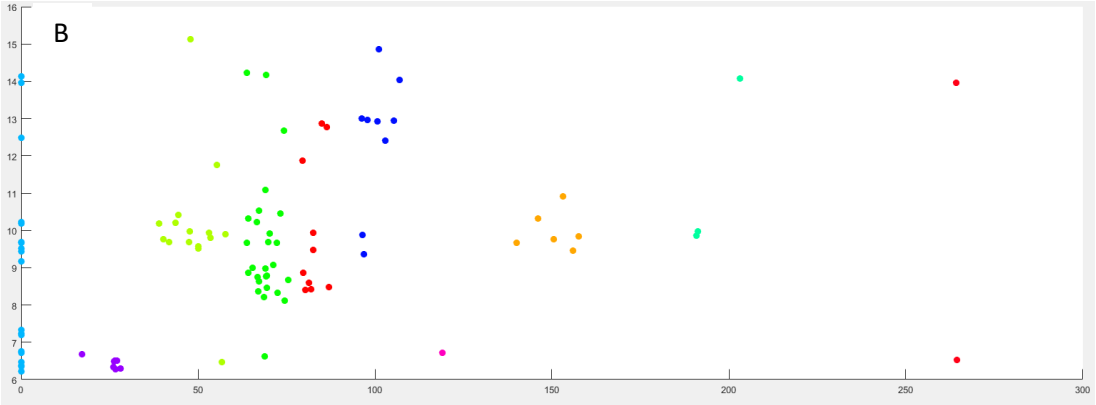
A dendrogram constructed in Matlab, representing the data from Appendix 3; however, the dendrogram has been limited in leaf node number to 10. This represents the similarity distance (x axis) between 10 relatively related clusters (y axis).



Appendix 5.

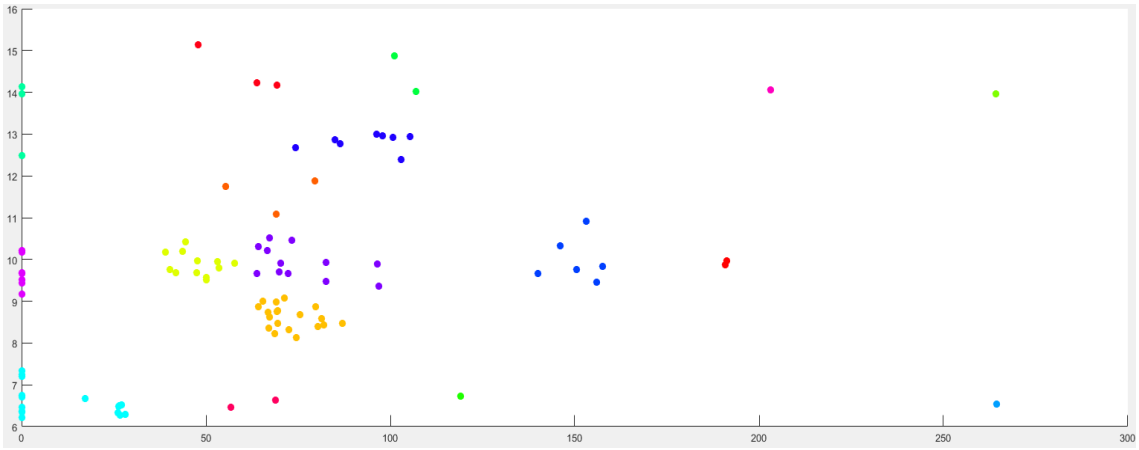
A dendrogram constructed in Matlab, representing the data from Appendix 4; however, the dendrogram has been limited in leaf node number to 5. This represents the similarity distance (x axis) between 10 relatively related clusters (y axis). This shows an even tighter clustering, with an expected increase in similarity distance between the defined clusters.





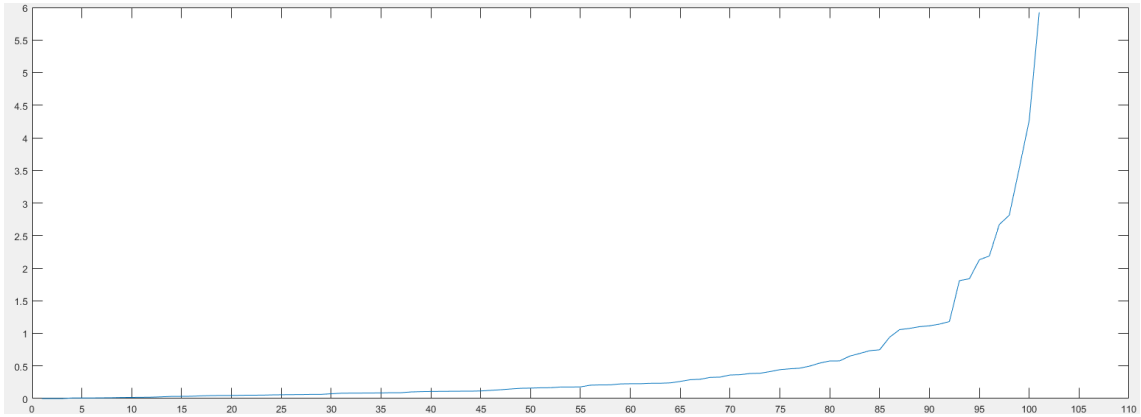
Appendix 6.

Matlab produced scatter graphs, of VIYKI characterisation data displaying the average fibril height (y axis) and the fibril periodicity (x axis). Each particle is colour coded per its respective identified cluster (10 clusters). A) was produced using single linkage and B) using complete linkage; the latter represents natural divisions in the data more accurately.



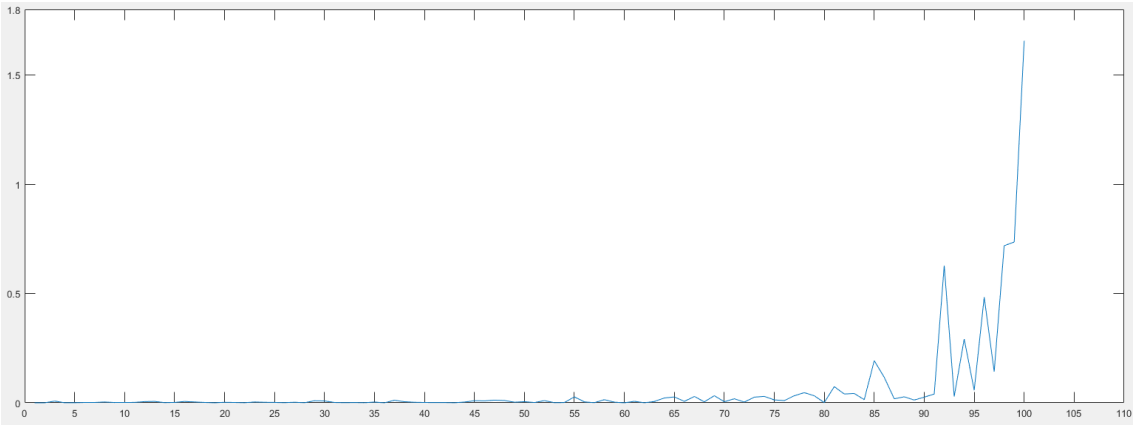
Appendix 7.

Matlab produced scatter graph showing the cluster identification of VIYKI fibrils using an IC derived cut-off. The mean IC value for VIYKI is 0.89; the graph represents the culmination of binary clusters below this similarity distance into grouped clusters. Periodicity is on the x-axis and average fibril height on the y-axis.



Appendix 8.

A Matlab produced graph displaying the increase in similarity distance across the originally allocated binary clusters of the VIYKI dataset used for characterisation. Binary cluster number is on the x-axis and similarity distance on the y axis; binary clusters are created at the start of the hierarchical clustering process and ascend in increasing similarity distance with the final binary cluster linking all particles within the dataset.



Appendix 9.

A Matlab produced graph representing the difference in similarity distance between binary cluster N and N+1. The x-axis displays the relationship between binary cluster N and N+1 and the y-axis shows the corresponding difference in similarity distance. While Appendix 8 displays the increase in similarity distance for VIYKI amongst the binary clusters, this graph aids in identifying the regions of rapidly increasing distance as the difference between 2 binary clusters becomes increasingly larger; which identifies regions of natural division in the data.

```
T1 =HeightMean;
SD=HeightSD;

figure(1)
histogram(CL,50)

     h =
vline(ContourLengthMean,'r',['',num2str(ContourLengthMean),'SD',num2st
r(ContourLengthSEM)]);

figure(2)
     histogram(Height,50,'normalization','pdf')
     hold on
     y = 0:0.1:20;
     mu = T1;
     sigma = SD;
     f = exp(-(y
-mu).^2./(2*sigma^2))./(sigma*sqrt(2*pi));
     plot(y,f,'LineWidth',1.5)
```

Appendix 10

Matlab script producing fragmentation distribution graphs