UNIVERSITY OF LEEDS

Pulling apart the intermolecular interactions of the Parkinson's disease linked protein alpha synuclein

Ciaran Patrick Anthony Doherty

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

The University of Leeds

Astbury Centre for Structural Molecular Biology

September 2017

The candidate confirms that the submitted work is his own, except where work which has formed part of jointly-authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated overleaf. The candidate confirms the appropriate credit has been given within the thesis where reference has been made to the work of others. This copy has been supplied on the understanding that it is copyright material and that no quotation from this thesis may be published without proper acknowledgement.

© 2017 The University of Leeds and Ciaran P. A. Doherty

Acknowledgements

I would like to express my gratitude to my supervisors, Dr David Brockwell and Prof Sheena Radford for making this PhD project possible. Their combined knowledge and infectious enthusiasm have helped shape a successful and interesting project and I am indebted to their unwavering guidance and support throughout my PhD, particularly in difficult periods. I would also like to thank UCB Pharma for funding this PhD (even if it is a million miles from the original proposed project) and in particular Dr Oliver Durrant, whose continued interest and support in this project has been invaluable.

I am very grateful to the many colleagues in the Radford and Brockwell labs past and present who have made a huge contribution to this project and to my time in Leeds. Special thanks goes to Lydia Young, who performed all of the mass spectrometry in this thesis and enriched this project massively. I would also like to thank Sam Lenton who provided excellent knowledge of SAXS, helping with the data collection and processing. NMR experiments were kindly performed by Theo Karamanos; I am extremely grateful to the time and effort put aside for me. Matthew Jackson has also made vital contributions to this work, providing proteins that were used in key experiments, performing the original condition scouting work for αSyn aggregation and providing important insights on this work, for which I am extremely appreciative. I would also like to express my gratitude to Nasir Khan who has dropped everything in order to offer help on many occasions. The endless supply of food has kept me fuelled during the 4 years of this project and I have really enjoyed your company in the office over the course of this PhD.

Outside the lab, I would like to thank my Mam and Dad who have given me the strong foundations and confidence in everything I do. I would also thank many friends who have offered their support (and distraction). I would like to thank Karen for indirectly encouraging me to stay in further education for as long as possible.

My final and absolute deepest gratitude goes to Alice who has always been there for me throughout the course of this PhD. Without her endless love, patience and support especially in the difficult periods, I would have found getting to this point impossible. The home that she has made for us (complete with fury family members) has made me immensely happy and allowed me to achieve so much more than I could have.

Table of Contents

List of Figures

List of Tables

List of Symbols and Abbreviations

Å	Angstrom
AD	Alzheimer's disease
AEX	Anion Exchange
AFM	Atomic force microscopy
ALS	Amyotrophic lateral sclerosis
ANS	8-Anilinonaphthalene-1-sulfonic acid
Aβ	Amyloid beta peptide
Aβ40	Amyloid-beta residues 1-40
Aβ42	Amyloid-beta residues 1-42
CD	Circular Dichroism
CID	Collision induced dissociation
CMA	Chaperone mediated autophagy
Cryo–EM	Cryo electron microscopy
DFS	Dynamic force spectroscopy
E. coli	Escherichia coli
EM	Electron microscopy
EOM	Ensemble optimisation method
ESI-IMS-MS	Native electrospray ionisation- ion mobility spectrometry – mass spectrometry
$\boldsymbol{F}$	Force
FAP	Familial amyloid polyneuropathy
far-UV CD	Far-ultra violet circular dichroism
FCS	Fluorescence correlation spectroscopy
F_R	Rupture Force
FRET	Fluorescence resonance energy transfer
FTIR	Fourier transform infrared
FWHM	Full width at half maximum
GWAS	Genome wide association studies
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HFIP	Hexoflouroisopropanol
HSQC	Heteronuclear single quantum coherence
$\boldsymbol{I}$	Intensity
IAPP	Islet amyloid polypeptide

IDP	Intrinsically disordered protein
$\boldsymbol{k}$	Spring constant
$\boldsymbol{k_B}$	Boltzmann constant
K_d	Dissociation constant
kDa	Kilodalton
kJ mol^{-1}	Kilojoule per mole
KLD	Kinase, ligase and DpnI
KO	Knock out
$\boldsymbol{k_{off}^{0F}}$	Off rate at zero force
LB	Lewy bodies
L_C	Contour length
ms	Millisecond
NAC	Non-amyloid β component
nm	Nanometer
ns	Nanosecond
OD_{600}	Optical density at 600 nm
$\boldsymbol{p}$	Persistence length
$\boldsymbol{P}$	Area of thermal fluctuations
PD	Parkinson's disease
PDD	Parkinson's disease with dementia
PEG	Polyethylene glycol
pI	Isoelectric point
pKA	Acid dissociation constant
pL	Protein L
pN	PicoNewton
poly-Q	Poly glutamine
PRE	Paramagnetic relaxation enhancement
PrP	Prion protein
$\boldsymbol{q}$	Scattering angle
RAGE	Receptor for advanced glycation end products
$\boldsymbol{r_f}$	loading rate
R_g	Radius of gyration
R_H	Hydrodynamic radius
SAXS	Small angle X-ray scattering

SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SEM	Scanning electron microscopy
SILAC	Stable isotope labelling by amino acids in cell culture
SMFS	Single molecule force spectroscopy
SNARE	N-ethylmaleimide–sensitive factor attachment protein receptor
SNc	Substansia nigra pars compacta
SNP	Single nucleotide polymorphism
SOD1	superoxide dismutase 1
ssNMR	Solid state nuclear magnetic resonance
$\boldsymbol{T}$	Absolute temperature
TEM	Transmission electron microscopy
ThT	Thioflavin T
TIRFM	Total internal reflection fluorescence microscopy
Tm	Melting temperature
Tris	Tris(hydroxymethyl)aminomethane
t-SNARE	Target SNARE
TTR	Transthyretin
VAMP2	Vesicle associated membrane protein 2
v-SNARE	Vesicle associated SNARE proteins
WLC	Worm-like chain
$\boldsymbol{WLC_{slope}}$	Worm-like chain slope/ differential worm-like chain model
WT	Wild type
$\boldsymbol{x}$	Extension of a polypeptide chain
$\boldsymbol{x_u}$	Distance along the reaction coordinate from the native to transition state
αSyn	Alpha synuclein
β2m	Beta-2-microglobulin
βSyn	Beta synuclein
γSyn	Gamma synuclein
λmax	Wavelength of maximum intensity

List of amino acids

A	Ala	alanine
C	Cys	cysteine
D	Asp	aspartate
E	Glu	glutamate
F	Phe	phenylalanine
G	Gly	glycine
H	His	histidine
I	Ile	isoleucine
K	Lys	lysine
L	Leu	leucine
M	Met	methionine
N	Asn	asparagine
P	Pro	proline
Q	Gln	glutamine
R	Arg	arginine
S	Ser	serine
T	Thr	threonine
V	Val	valine
W	Trp	tryptophan
Y	Tyr	tyrosine

Abstract

Amyloidoses are a group of protein misfolding diseases that are characterised by the abnormal accumulation of highly ordered filamentous assemblies known as amyloid. This phenomenon is associated with more than 50 human diseases, some of which are the most debilitating disorders that threaten human health today. Many of these disorders have age as the main contributing risk factor and, therefore, pose an ever-increasing risk in the developed world with aging societies. Despite intense research, much remains unknown about the fundamental processes driving protein aggregation in these diseases and there are few disease modifying treatments available.

A protein that undergoes amyloid formation and causes disease is the intrinsically disordered neuronal protein α-synuclein (αSyn), the aggregation of which leads to several diseases including Parkinson's disease (PD) which is the second-most common neurodegenerative disorder that affects 2–3% of the population ≥65 years of age. Importantly, the toxic species on the aggregation pathway are difficult to identify and determine in molecular detail. This thesis was motivated by this fact and aimed to study the initial intermolecular events in αSyn self-assembly (dimerisation) on a single molecule scale. Single molecule force spectroscopy (SMFS) methodologies were therefore utilised in order to study these early protein-protein interaction events.

A display system was firstly designed and validated in which small regions of highly aggregation-prone sequences can be presented in a protein scaffold in a robust and reproducible manner for SMFS studies. It was demonstrated that intermolecular interactions of these sequences could be analysed by implementing this system. A novel heterodimeric interaction between the central aggregation-prone regions of αSyn (residues 71-82) and the same region of its human homologue γSynuclein (γSyn), were revealed by using this system. Further study led to the finding that this interaction played a role in the inhibiting the aggregation of αSyn.

The dimerisation interaction of full length αSyn has also been analysed in this thesis and several important findings have been demonstrated. The SMFS experiments show that force-resistant structure forms in the dimeric species of αSyn and that this structure is dependent on the environmental conditions. SMFS utilising different immobilisation regimes of αSyn have also allowed the location of a novel interaction interface involving the N-terminal region of the protein. Further SMFS experiments investigating the effects of salt and hydrophobicity have on dimerisation, alongside bioinformatics analyses of the protein sequence led to the hypotheses

that the dimeric interaction is driven by hydrophobic stretches in the N-terminal region, but modulated by local electrostatics. In vitro aggregation assays and SMFS on non-aggregation-prone synuclein homologues (β- and γSyn) indicated that that this interaction is protective against aggregation, considering these finding with existing literature prompted speculation that the interactions observed in SMFS may indeed be physiologically relevant. This may therefore be an important finding in regards to targeting the aggregation process with disease modifying agents.

1 Introduction

1.1 Introduction to protein folding

Proteins are the second most abundant molecules in biology after water[1]. They act as the "workhorse" of life, performing a plethora of functions which make the existence of all life on Earth possible. The functions that proteins perform are dictated by their native tertiary and often quaternary structure. In the early 1960's, Anfinsen and colleagues' classic work on the renaturation of Ribonuclease A, led to the hypothesis that all the information required for a protein to fold into its native conformation is encoded in the primary amino acid sequence[2]. This mechanism depends on searching for states with the lowest free energy and therefore highest stability. However, this leads to the protein folding paradox in which it becomes impossible for a protein to perform an exhaustive search of all conformations and achieve a global free energy minimum on a biologically relevant timescale[3,4]. In fact, if a protein of average length (300 amino acids) was to undergo this mechanism of folding, it would have explored more conformations than the total number of atoms in the universe before it would arrive at its native conformation[1]. This paradox was first considered by Levinthal[3], who proposed that folding is directed down specific, kinetically controlled pathways, guided by the rapid formation of local interactions[5]. These interactions determine the further folding of the polypeptide chain which, by the formation of local native-like structures with enhanced stability, results in fewer conformations being sampled en route to the native conformation, allowing protein folding on a biologically relevant timescale.

The nucleation-growth model was the first proposed mechanism of protein folding[6]. This theory postulated that tertiary structure is propagated rapidly from a nucleus of secondary structure in a hierarchical manner. However, proteins were observed to fold via folding intermediates[7,8] which this model can not account for. This led to the emergence of the framework model[9,10] and the analogous diffusion-collision model of folding[11,12] to better account for folding intermediates (Figure 1-1). These models proposed the packing together of secondary structural elements and their collision and coalescence into tertiary structures. The hydrophobic collapse model[13,14] was subsequently proposed. This model proposed the initial collapse of a polypeptide chain followed by secondary structure formation, in contrast to the hierarchical progression of structure in the previous models. This process occurs via the initial sequestration of hydrophobic sidechains from the surrounding aqueous environment to produce a molten globule intermediate, constraining the volume in which the conformational search for the native state can occur, guiding native state topology. This process occurs before

secondary structure formation. However, studies on the folding of chymotrypsin inhibitor-2 did not reveal the presence of intermediates, but showed a 2-state transition[15]. ϕ-value analysis carried out on the same protein revealed that secondary and tertiary structures form in parallel[16]. These findings, along with the discovery of two-state folding kinetics for many other proteins[17,18], resulted in the postulation of the nucleation-condensation model as an amalgamation of features from the above models[19]. In this model, a small nucleus of weak structure forms, which is stabilised by long range tertiary interactions. The nucleus represents the transition state from which the remainder of the secondary and tertiary structures are formed. In reality, a continuum of models probably exists, with the extremes of the framework and hydrophobic collapse models. The formation of structural elements in sequence or in parallel depends on the relative strengths of secondary and tertiary interactions[20,21].

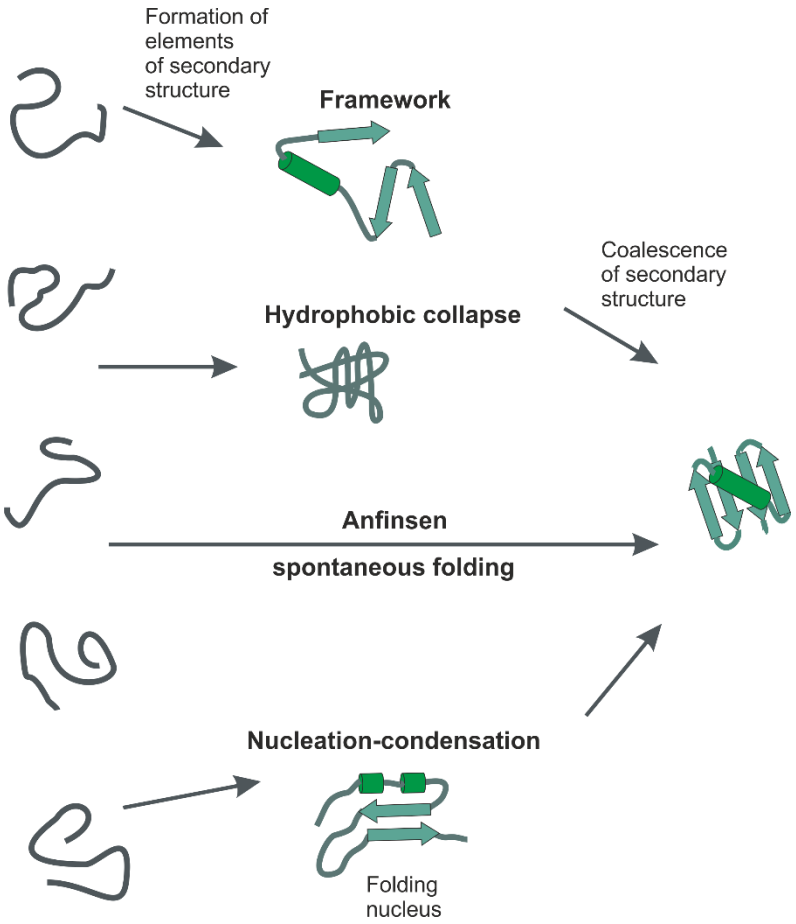


Figure 1-1. Schematic of some of the proposed folding mechanisms. Illustrations of the framework[9,10], hydrophobic collapse[13,14] and nucleation-condensation hierarchical assembly models[19] are shown. Anfinsen's original protein folding hypothesis is depicted as spontaneous folding from unstructured ensemble to the native state[2]. The conformational entropy decreases from the unstructured ensemble on the left to the native state on the right as the structural complexity increases. Figure redrawn and adapted from Morris and Searle 2012[22].

The process of protein folding is often conceptualised as a “downhill” pathway from an unstructured high energy conformation to a more stable, lower energy native state. The energy landscape of this process describes the intermediate ensemble of conformations sampled en route during folding and is expressed as the internal free energy of the protein as it explores conformational space. This process can be illustrated by a folding funnel[23,24]. An idealised folding funnel is depicted in Figure 1-2 in which the intrinsic free energy decreases concurrently with the conformational entropy until an energy minimum is reached corresponding to the native state[23]. This idealistic scenario is likely to differ drastically in reality. Indeed proteins that fold via a two state transition will have a relatively smooth energy landscape. However, the majority of proteins tend to proceed to their native states via a population of folding intermediates. These “on pathway” intermediates give rise to a roughness of the energy landscape with high kinetic barriers and low kinetic traps as shown in Figure 1-3.

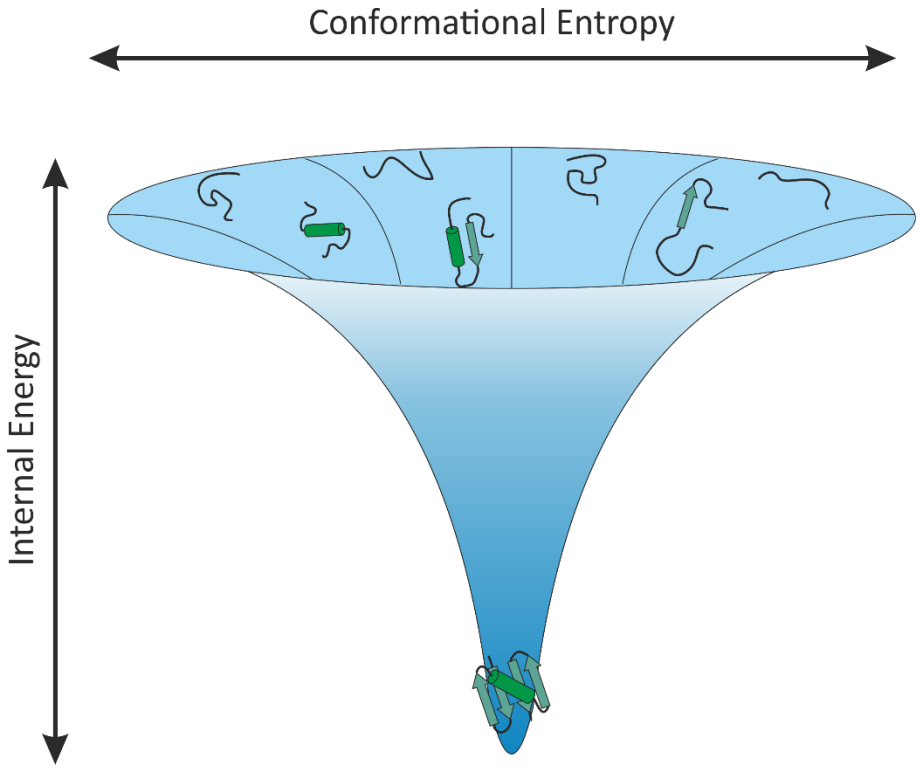


Figure 1-2. An Idealised energy landscape of protein folding. The vertical axis represents the internal free energy of the system. The horizontal axis represents the conformational entropy explored by the polypeptide chain. Folding starts at the rim of the funnel depicted by schematic polypeptide chains that do not contain structural elements. As the number of intramolecular contacts increases, the free energy lowers and the conformational freedom is reduced until the native state is formed at the base of the funnel.

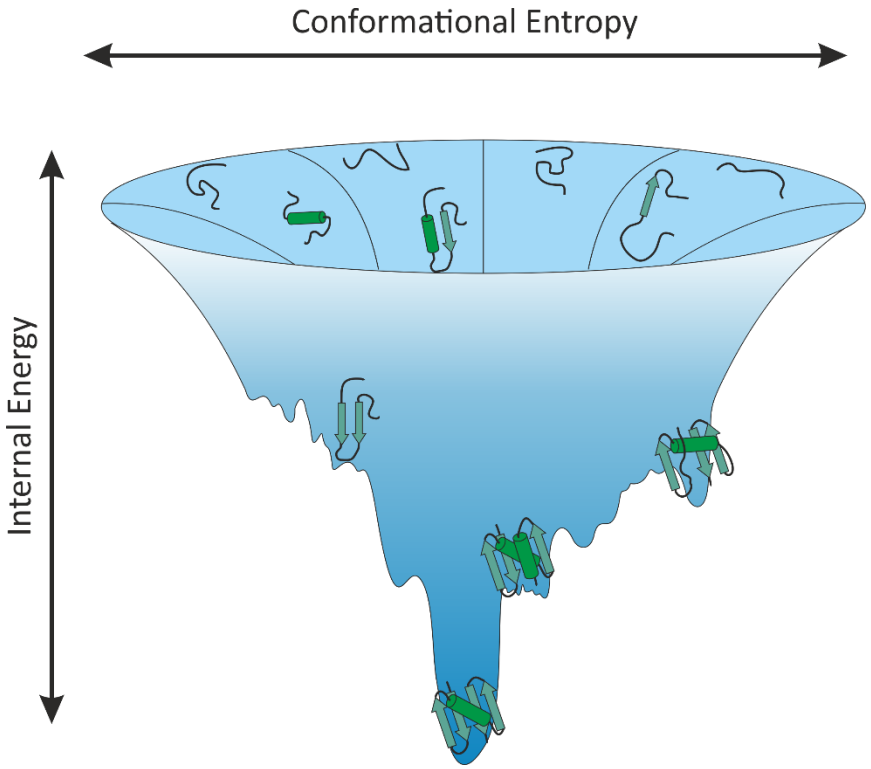


Figure 1-3. A rough protein folding landscape with folding intermediates. The unfolded protein depicted by chains lacking structural components funnel down to a low energy native state at the base of the funnel. The surface of the funnel is rough and contains low energy troughs in which folding intermediates can be populated.

Protein folding pathways are further complicated by "off pathway" structures which are not directly formed in the normal folding pathway of a protein progressing to the native state. These off pathway misfolded species may be kinetically trapped and require substantial reorganisation before the native state can be reached[14].

1.2 Intrinsically disordered proteins

Protein stabilities have been traditionally reported in the range of -10 to -50 kJ mol^{-1}, a fraction of the strength of a covalent bond. Thus folded proteins are only marginally stable and still contain some degree of conformational dynamics or disorder[25]. They exhibit motions that span a range of timescales (from ns to ms or slower)[26,27]. The dynamic properties of proteins have been shown to be key to various aspects of protein function, such as enzymatic catalysis[28], regulation of protein synthesis[29], and many others[30]. Some proteins may have an energy landscape that is drastically different from those shown in both Figure 1-2 and Figure 1-3. The energy landscape for these proteins may be much flatter and lack an obvious folded native state as shown in Figure 1-4. Some proteins are so unstable that they lack structure in extended regions or in the entirety of the protein. Proteins that possess at least one region

that does not have a fixed secondary or tertiary structure are termed intrinsically disordered proteins (IDPs) and these lie at the extreme of the spectrum for molecular dynamism as they can sample a very large space of heterogeneous conformations[31,32]. Regular secondary structures only occupy a small fraction of ϕ and ψ angles. Proteins occupying regions of conformational space that are not in these narrow distributions, defined by Ramachandran[33], can therefore be described as intrinsically disordered. However, IDPs are not completely random; they contain some conformational preferences that are sequence specific[34,35].

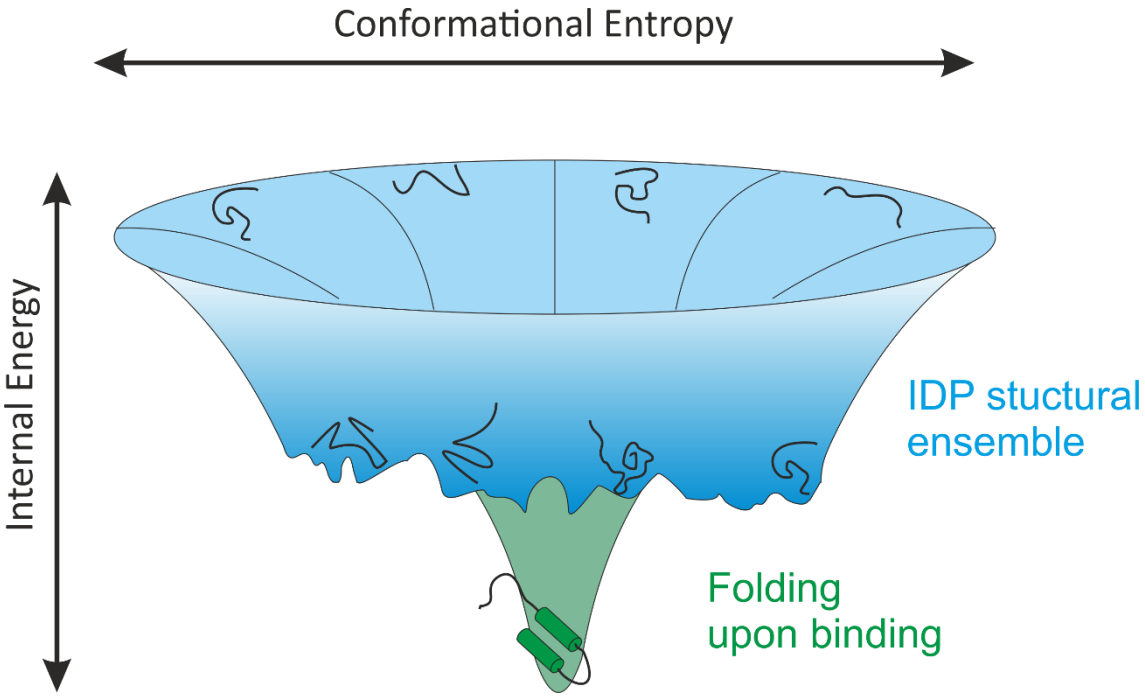


Figure 1-4. Energy landscape of an IDP and its folding upon binding. The energy landscape of an IDP is shown in blue as a relatively flat but rough landscape. The folding funnel lacks any obvious troughs populated by more structured polypeptides. A lower energy state exists that forms under IDP binding to partners is shown in green with a chain with some structure shown at lower free energy.

IDPs play key roles in the cell[31], the prevalence of these proteins in the proteome is testament of this: around 50% of human proteins contain stretches of at least 30 disordered residues and around 25% of human proteins are disordered[32]. They are utilised in the cell as transcription factors, in protein signalling pathways and in regulatory pathways[31]. IDPs involved in some of these pathways are often referred to as 'hub' proteins due to their ability to bind many different partners, playing a major role in protein-protein interaction networks[32,36]. IDPs are weak but specific binders, and even when bound to a protein partner may still possess some disorder or 'fuzziness'[37]. They represent good examples of conformational selection theory[38], as in many cases, IDPs undergo coupled folding and binding due to the stability that binding

confers on one of the many conformations of an IDP[39]. This property of simultaneous folding and binding has been proposed to be the origin of the transient and weak interactions IDPs often undertake, as there is an entropic penalty for folding a disordered region that results in a higher K_d[36,39]. This is because for an IDP, the binding energy is the sum of the favourable binding interaction and the unfavourable folding energy of a disordered protein or region, increasing off rates and therefore dictating a weaker interaction. In this manner, the affinity and the off rates can be adjusted to almost any desired value by varying the free energy of folding, hence some IDPs appear "almost" folded whereas others maintain greater disorder. IDPs therefore, are advantageous in systems where they need to act as a 'volume switch', rather than a binary on/ off control, a property ideal for the modulation of signalling as fast off-rates are required. The fact that IDP's are prevalent in eukaryotes but rare in prokaryotes suggests that IDPs are evolutionary recent molecules and therefore play a role in more complex eukaryotic signalling and modulation. Such functions only arise as modifications to a pathway after the main function has evolved[40].

The transient nature of their intramolecular interactions may have roots in the primary sequence of IDPs which often are of low complexity[31]. Weak binding is often carried out by small repeated regions of a poor version of a consensus binding sequence, an evolved property that enables rapid binding and easy displacement[41]. These sequences are easily lost or gained in the course of evolution by random mutations. An example of this is in the presynaptic IDP α-synuclein (αSyn), which will be the protein of focus in this thesis. αSyn contains at least six but as many as nine imperfect repeats with the consensus sequence KTKEGV present in a region of the protein shown to bind lipid membranes[42-44]. The amino acid composition of these repeats varies considerably in the protein as highlighted in Figure 1-5. Additionally, evolutionary closely related homologue proteins: β-synuclein (βSyn) and γ-synuclein (γSyn), show different numbers of these repeat regions demonstrating how evolution can easily tune the number of these types of regions and thereby modulate the protein's function over a short period of evolutionary time. On the other hand, due to the degenerate nature of these proteins, they are unlikely to lose binding capacity or functionality upon mutation[36,45].

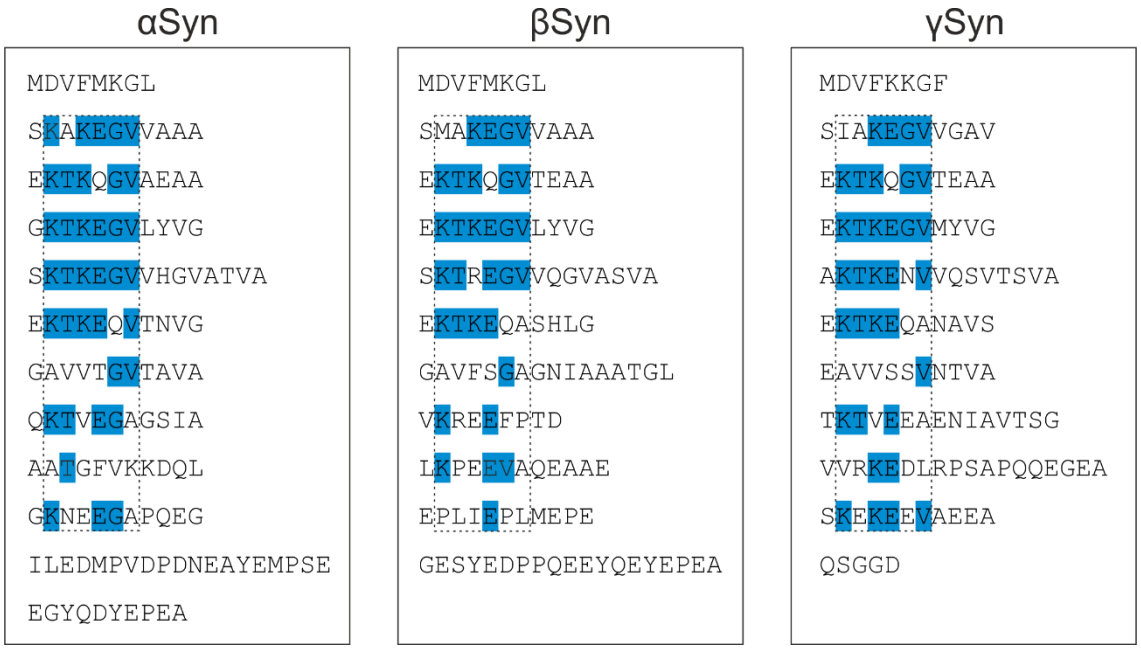


Figure 1-5. Sequence of human synucleins aligned by their KTKEGV repeats. Amino acids that match the KTKEGV repeat motive are shaded in blue. The repeats are imperfect in the manner in they align and their sequence identity. Adapted and redrawn from Dettmer *et al.* 2015[46].

IDPs also have other advantages in regards to their folded counterparts. Their flexibility is a key property in that an IDP has a large solvent-accessible surface area, that can be used to 'grab on' or bind to their more rigid partners, a property that has been termed 'sticky arm'[31]. This is especially the case in IDPs which contain a high proportion of Pro residues. These residues impart rigidity and confer conformations close to a polyproline II helix, such conformations thus provide a 'sticky arm' but pay a lower entropic penalty upon binding[47]. IDPs can also utilise their intrinsic flexibility to adopt conformations that are sterically hindered for structured proteins, such wrapping around a protein partner in order to maximise the buried surface area[36,48]. Flexibility also imparts promiscuity, IDPs are able to bind partners in different ways. For example p21 binds to different cyclin dependent kinase proteins in different conformations[49].

1.3 Protein misfolding and aggregation

Under physiological conditions, the native state of a protein is believed to be the locally most stable, lowest energy structure[1]. However, as discussed in the previous section, proteins are only marginally stable and in a system of constant structural flux. Native proteins experience fluctuations in their minimal energy states and a protein may undergo multiple localised unfolding events throughout the protein's structure[24]. These partially unfolded forms of proteins may represent a functional, on-pathway version of the protein, but may also

represent off-pathway states. Intermediate states are particularly vulnerable to misfolding and aggregation (Figure 1-6). Rare or unfavourable conformations of proteins may be produced by a change in the thermodynamic stability or interconversion kinetics of native or non-native states of a protein. This can occur by destabilising factors such as a change in pH, temperature, absence of a ligand, or mutations in the protein[50]. The misfolded state of a protein may in reality be more stable than that of the native state, as the native state may only be the free energy minimum under physiological conditions as depicted in Figure 1-6[51-53]. It is important to note that intermolecular aggregation as depicted in Figure 1-6 is concentration dependent as it is governed by the thermodynamics of the process.

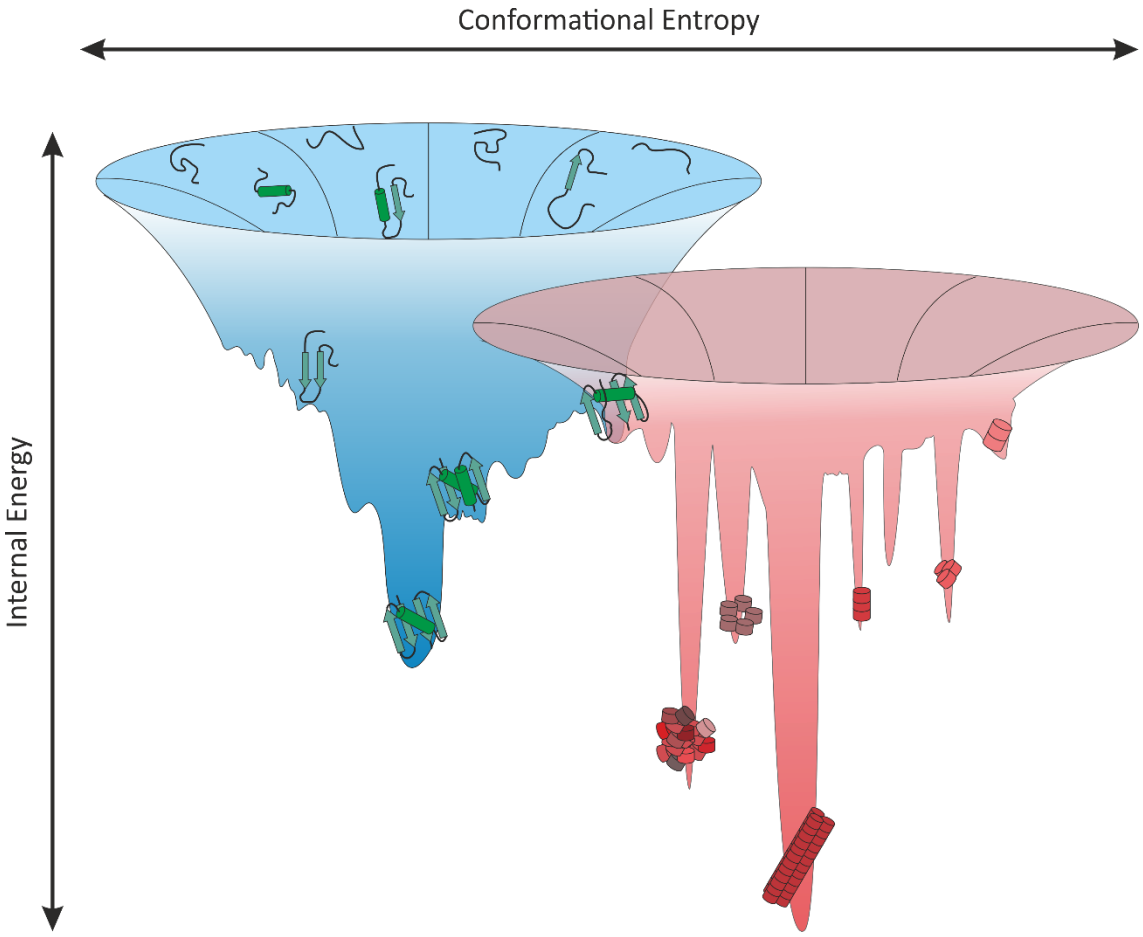


Figure 1-6. Energy landscape of folding and aggregation. An intramolecular folding energy landscape is shown in blue. The surface of the funnel is rough and contains various troughs populated by folding intermediates. An energy landscape of intermolecular aggregation is shown in red. Aggregated species extend to lower energy states than that of the folding landscape with cross-β amyloid fibrils as a lowest energy state[52]. The overlap of the landscapes illustrates the relationship between folding and aggregation where non-native states may be able to traverse from folding to aggregation under certain conditions. Adapted and redrawn from Karamanos *et al.* 2016[54].

Protein species that undergo transient unfolding often have hydrophobic and other residues exposed to solvent that would otherwise be buried in the protein structure. *In vivo*, this can result in undesirable interactions with other molecules, especially given that proteins function under conditions where there is a 300-400 mg/ml concentration of cytosolic proteins and other macromolecules[55].

Much research has been directed towards determining the propensity of proteins to aggregate[1,50,51,56,57]. Folding of a protein into its native state sequesters hydrophobic residues from solution and decreases aggregation propensity[56]. Interestingly, however, IDPs (see Section 1.2) tend to have a reduced aggregation propensity[58] which is probably an evolved characteristic. The higher net charge and lower hydrophobicity of IDPs in relation to globular proteins make aggregation thermodynamically less favourable[59]. The charge patterning in these proteins may also be important, as charged "gatekeeper" residues flanking hydrophobic stretches control against aggregation. Moreover, partially folded intermediates can be more likely to aggregate than fully unfolded proteins[1], as partially folded proteins can expose a high local concentration of hydrophobic residues that would otherwise be sequestered from solution in the core of a protein. This creates an aggregation prone, hydrophobic interface.

There are many other factors that destabilise proteins and make aggregation more probable. Mutations in a protein sequence, either engineered or those that occur spontaneously in nature, have been shown to increase the aggregation propensity of proteins by acting in a variety of ways. They can increase hydrophobicity, increase β-sheet propensity, or even reduce the net charge of the polypeptide. For example, variants of transthyretin (TTR) which are significantly less stable than the wild-type (WT) protein are involved in familial amyloidogenic neuropathy[60]. Similarly, mutated lysozyme variants are associated with aggregation and cause hereditary amyloidosis[61]. Changes in the environment or protein concentration can also play a key role in the aggregation of WT proteins. For example, the accumulation of the protein β_2-microglobulin (β_2m) in patients undergoing long term haemodialysis can raise serum concentrations of the protein to 60-times normal levels[62]. This causes an increase in the total concentration of a transiently populated, partially folded intermediate, which has a high propensity to aggregate[63], a situation illustrated in Figure 1-6. This subsequently causes the protein to deposit in amyloid plaques[24]. A high concentration of misfolded intermediates also risks the exhaustion of the cellular chaperone machinery, increasing aggregation propensity further[64,65]. Impairment of any part of the protein processing machinery can also have

detrimental impacts on the proteostasis networks in a cell, increasing the risks of aggregation[65].

1.4 Amyloid

A common generic misfolded conformation of proteins is amyloid: an aggregated, low energy structure proteins can adopt that is high in β-sheet conformation forming a generic cross-β structure[1,50,51,56,57]. The conversion of soluble proteins and peptides into insoluble aggregates, specifically amyloid aggregates, has emerged as a subject of intense research. This reflects the fact that amyloid deposition is associated with some of the most debilitating disorders that threaten human health in the world at present. The recognition that many disorders associated with amyloid formation are no longer rare has led to a surge of interest in the area. Many of these conditions were discovered only a generation or two ago, but have rapidly become some of the most common diseases in the developed world. It has been estimated that the number of people suffering from dementia (an umbrella term for diseases associated with amyloid deposition in the brain) worldwide in 2016 was 46.8 million, and by 2050 this figure could reach 131.5 million[66]. It is important to note that these figures are not extensive for all amyloid-associated diseases, but just those relating to dementia. The numbers associated with all amyloid-associated diseases is much higher.

The process of amyloid formation is not only relevant in the sense of disease, but it also challenges our understanding of the nature, structure and evolution of functional states of proteins[67-71]. The amyloid state may indeed be a structure that many, if not all proteins may be able to access, and that it reflects a well-defined structural form which is an alternative to the native state due to the intrinsic and generic properties of the polypeptide chain itself[50,56,57,67,72]. The fact that amyloid as an exquisitely organised, self-assembled structure is generically accessible to relatively small, noncomplex polypeptides, and that functional amyloids are present throughout domains of life also raises interesting questions on life's origins[68-71].

1.4.1 Historical perspective

Despite the relatively recent explosion of interest in the amyloid field, the amyloid state was first observed more than 160 years ago[73]. The term amyloid is in fact a misnomer, derived from the Latin "amylum" for starch. It was coined by Rudolph Virchow in 1854[73] as reviewed by Sipe and Cohen[74], based on the observation that the application of iodine to abnormal macroscopic structures in brains stained them blue, indicative of the presence of starch. Amyloid was thus considered to be carbohydrate based. However, it was first recognised in 1859 by Friedrich

and Kekule[75] that amyloid deposits were in fact proteinaceous. Attention therefore shifted to the study of amyloid as a protein, and later, as a class of proteins with generic properties but which are unrelated in amino acid sequence.

In the years since these discoveries, much has been revealed about the structure of the amyloid state. In 1922, it was observed by Bennhold[76] that Congo red, a common industrial dye, is able to bind to amyloid deposits revealing the tinctorial properties of amyloid. It was subsequently shown that the characteristic red-green birefringence of Congo red under polarised light in the presence of amyloidogenic material is enhanced, suggesting the presence of an ordered microscopic structure[77]. Electron microscopy (EM) studies were carried out, which revealed a general common fibrillary ultrastructure to amyloid material[78-81]. X-ray diffraction experiments showed that amyloid fibrils are made up of a characteristic cross-β architecture, and that β-strands are lined up perpendicular to the fibrillar axis[82-84], a property considered to be a defining characteristic of amyloid[51,57].

Long believed to derive from a single origin, it was not until the 1970's that the heterogeneity of amyloid started to be appreciated[85-89]. Over the following decade, around 20 normally soluble proteins were found to form amyloid associated with disease, with deposits localised in organs or distributed systemically. These diseases are collectively termed as amyloidoses.

Amyloidoses are now known to be a clinically and biochemically heterogeneous group of disorders of protein folding. There are currently 40 discovered human proteins[90] (including both extracellular and intracellular aggregates) which give rise to diseases with diverse clinical pathologies and are associated with the aberrant folding and aggregation of normally soluble, functional proteins into an aggregated amyloid state. The vast majority of these diseases are associated with increasing age such as Alzheimer's disease (AD) and Parkinson's disease (PD), hence amyloidosis accounts for some of the most prevalent diseases in the modern world and represents one of the greatest socio-economic burdens of our time[66].

1.4.2 The structure of amyloid

As mentioned previously, amyloid structures are defined by a cross-β pattern in X-ray fibre diffraction studies that is indicative of component β-strands orientated perpendicular to the long axis of a fibril[91-93] as shown in Figure 1-7 A and B. The fibre diffraction patterns contain two main reflections: one equatorial at ~10 Å, thought to originate from the packing of β-sheets perpendicular to the fibril axis; and one meridional at 4.7 Å which arises from the packing of adjacent β-strands along the fibril axis[51]. Similar to natively folded proteins, amyloid

structures are closely packed and highly ordered. However, amyloid structures have fundamentally different features than that of their constituent native proteins, conferring a very high level of kinetic and thermodynamic stability. This is illustrated in Figure 1-6 in which the aggregation landscape contains deep, low energy troughs separated by steep kinetic barriers in comparison with the folding energy landscape. Amyloid states are generically rich in β-sheet structure and have relatively simple secondary and tertiary structures, in contrast to the folds of globular proteins which are highly diverse and can range from all α-helical to all β-sheet generated by evolution to obtain a vast variety of different functions.

At the nanometer length scale, amyloid fibrils from different proteins are similar. Visualised using negative stain transmission electron microscopy (TEM) or atomic force microscopy (AFM), amyloid fibrils appear as unbranched filamentous structures that can extend to micrometers in length but only a few nanometers in diameter[94,95]. They display a hierarchical organisation as shown in Figure 1-8, typically involving multiple protofilaments that twist around each other to form mature fibrils. Advances in cryo-electron microscopy (cryo-EM)[93,96,97] and solid state NMR (ssNMR)[93,96,98-100] of amyloid fibrils have led to an increasingly detailed knowledge of their structure. These studies have confirmed the generic nature of amyloid fibrils, attributed to the properties of a polypeptide backbone to support a continuous hydrogen bonding pattern in the fibril core. They have also revealed that structural origin of the heterogeneity of fibrils is driven by the side chain packing that are incorporated into the fibril architecture[72].

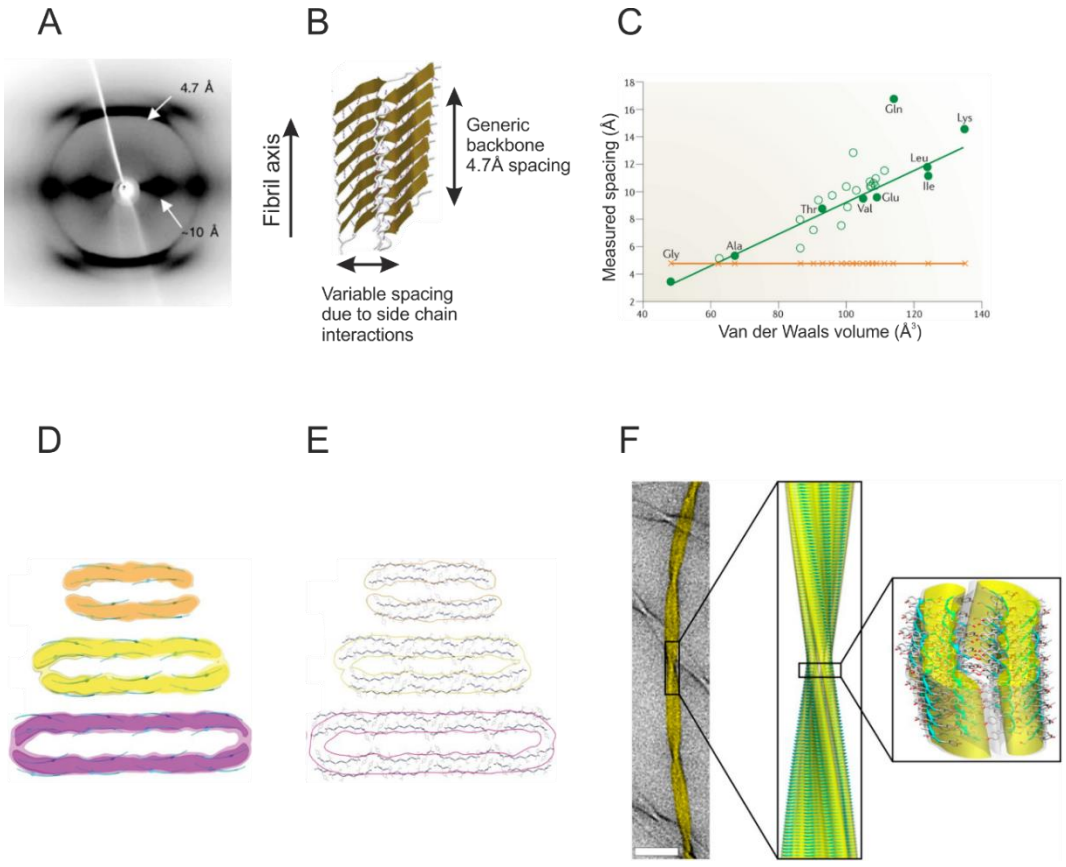


Figure 1-7. Structural characteristics of amyloid fibrils and different biophysical techniques to study them. (A) X-ray diffraction pattern of fibrils formed from islet amyloid polypeptide (IAPP) with characteristic reflections of 4.7 Å and 10 Å[94]. (B) Schematic representation of the cross-β structure common to amyloid fibrils, the spacing between polypeptide chains along the fibril axis is constant at around 4.7 Å due to inter-main chain hydrogen bonding. Also shown is the variable distance between the β-sheets in a perpendicular direction to the axis of the fibril, dependent on the nature of the side chains[51]. (C) Correlation of the main chain spacing (orange crosses and line) and the side chain spacing (green circles and line). Filled circles show fibrils made from homopolymers[72], open circles show fibrils derived from heterogeneous sequences. (D and E) Cryo–EM reconstructions of fibrils of TTR fragment 105-115 determined at atomic level resolution in combination with ssNMR data showing cross-sections through the long fibril axis[93]. Secondary structures of the protofilaments determined by ssNMR are shown in a cyan ribbon representation superimposed on fibril cross-sections. Orange yellow and purple electron densities and envelopes show doublet, triplet and quadruplet fibril assemblies of protofilaments depicting the variation in protofilament assembly leading to polymorphisms of amyloid fibrils. (F) High resolution triplet fibril structure from the same study showing hierarchical assembly of amyloid fibrils visualised by TEM (left), cryo-EM fibrillar structure (middle) with atomic level resolution from ssNMR data. Images adapted from references as indicated.

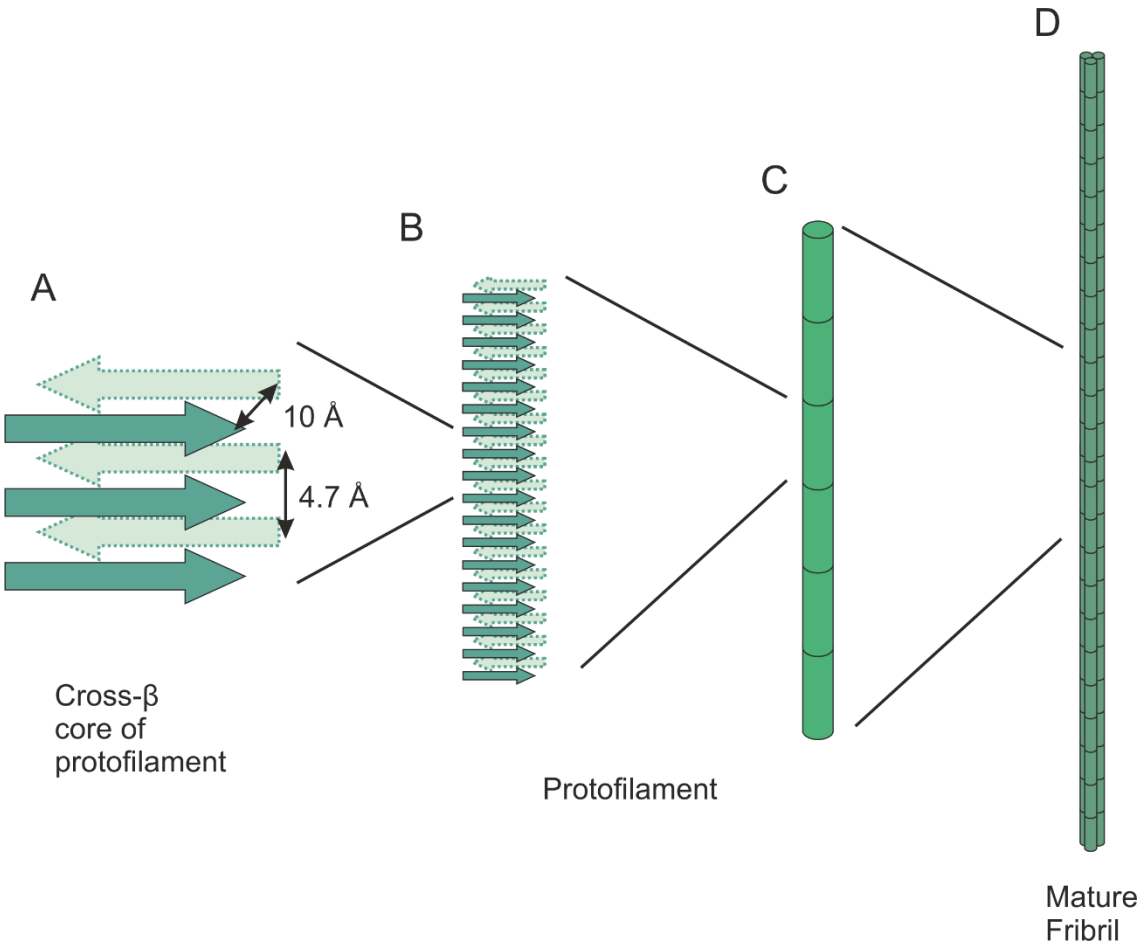


Figure 1-8. Hierarchical assembly of mature amyloid fibrils. (A) and (B) the cross-β core of a protofilament with characteristic spacings of main chains and side chains. (C) Protofilament structure that self-associates with other protofilaments to form mature fibrils (D). Redrawn and adapted from Serpell, 2014.[101].

A recent study utilised cryo-EM and ssNMR to interrogate the fibrillar structure of amyloid fibrils of TTR 105-115 at an atomic level (Figure 1-7 E-F)[93]. This study confirmed the general structure of the amyloid fibrils which are indeed attributed to cross-β hydrogen bonding pattern of the fibril core. This study demonstrated how the packing interactions facilitate the hierarchical assembly of mature amyloid fibrils. It also confirmed a basis of structural polymorphism between the fibrils, which partially results from the manner in which protofilaments assemble into mature fibrils.

Several recent studies have utilised various techniques as discussed above to elucidate fibril structures in atomic detail of full length amyloidogenic proteins. αSyn, Aβ42 and tau are examples of these structures and are highlighted in Figure 1-9.

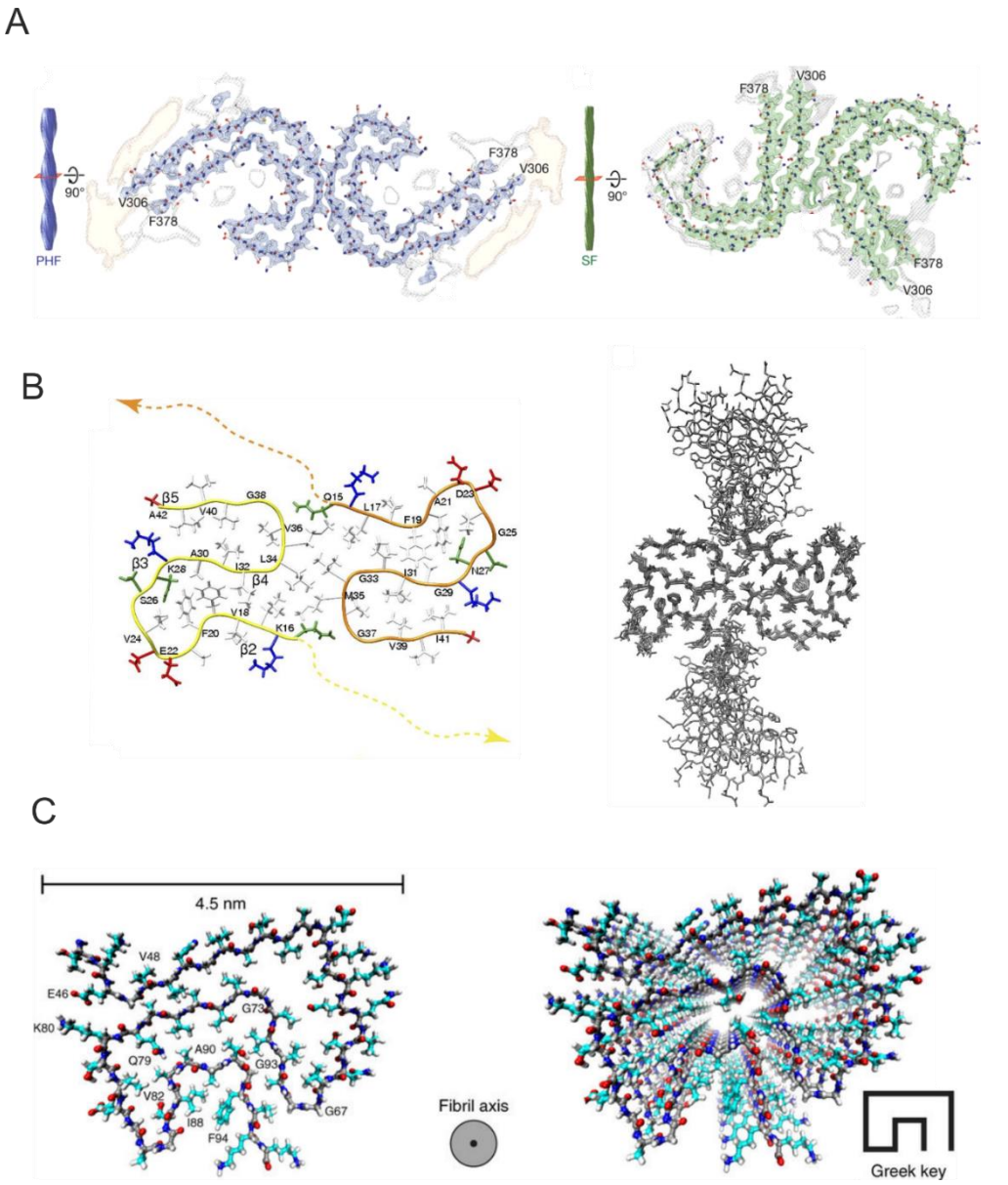


Figure 1-9. Examples of recent fibril structures of full length proteins solved in atomic detail. A) Fibrillar core structure of tau protein from AD patient brains[102]. The structures of two different polymorphs are shown, termed paired helical fibril (PHF, left in purple) and straight fibril (SF, right in green). These polymorphs have identical protofilaments but differ in the inter-protofilament packing. Cryo-EM density maps are shown in grey with high-resolution maps shown in blue (PHFs) and green (SFs). Orange density in the PHF structure corresponds to the "fuzzy coat" of disordered Tau. B) ssNMR-solved structure of fibrils of Aβ42 showing a "double-horseshoe" structure[103]. The backbone of the two point symmetric molecules are shown as yellow and orange spines (left). The 3D structure of the N-terminal residues 1–14 is indicated by dotted lines. The side chains of the positively charged residues are shown in red, the negatively charged in blue, the hydrophobic residues in white, and polar residues in green. Every second residue is labelled. 10 conformers with the lowest energy 3D structures are shown (right). C) ssNMR-solved fibril structure of αSyn revealing a "Greek-key" structure[100]. A view of a central monomer from residues 44 to 97 is shown on the left, as viewed down the fibril axis, showing the Greek-key motif of the fibril core. A view of the stacked monomers is shown on the right, showing the side chain alignment between each monomer down the fibril axis. Images adapted from references as indicated.

1.4.3 Mechanisms of amyloid formation

The mechanism of amyloid formation is a complex process, involving the interplay between a plethora of intermediate species of varied morphology, together with different elementary processes[51]. Soluble protein monomers are converted into fibrillar structures via a transiently populated aggregated nucleus, around which further deposition of monomers occurs[24] (Figure 1-1). This mechanism is an example of a nucleated growth reaction in which the formation of an aggregation-competent nucleus is kinetically disfavoured and is the rate-limiting phase of amyloid formation termed the lag phase. This is followed by a rapid elongation phase in which fibrils form. Protein-protein interactions play a key role in the lag phase of the growth mechanism, interacting in a transient manner to form species that may have, or lack, the propensity to elongate, thus creating a heterogeneous pool of on- and off-pathway intermediates. A variety of other species are formed during amyloid formation such as amorphous aggregates and other prefibrillar species[104-108]. They are often formed faster than fibrils via nucleation independent processes[104].

The nucleated-growth mechanism is supported by the fact that the addition of pre-formed, aggregation-competent seeds greatly increases the rate of aggregation and decreases the rate-limiting lag phase *in vitro* (Figure 1-10)[109]. The interactions of monomers with existing aggregates is favoured by the thermodynamic stability of the aggregate, causing the elongation phase to proceed rapidly[110]. The nucleation-dependent mechanism of amyloid growth can be followed experimentally by the amyloid-specific, aromatic dye thioflavin T (ThT) which binds predominantly to cross-β structures leading to enhanced fluorescence and resulting in a characteristic sigmoidal curve (Figure 1-10). At the molecular level, the connections between nucleation events and the lag time or elongation rates[111] have proved to be complicated[112,113]. This complexity arises as there are other contributing factors to the aggregation process other than just a simple primary nucleated growth mechanism. The nucleation process can also involve various secondary processes (Figure 1-11) which are governed by the behaviour of the aggregates formed in the polymerisation reaction[113,114]. Fibril fragmentation and surface-catalysed processes are important examples of secondary events (Figure 1-11). Fragmentation of pre-formed aggregates multiplies the number of polymerisation-competent fibril ends and so exponentially increases the proliferation of aggregates[113,114]. A similar process takes place during surface-catalysed processes which are dependent on the concentration of pre-existing fibrillar aggregates which facilitate nucleation and aggregation[114].

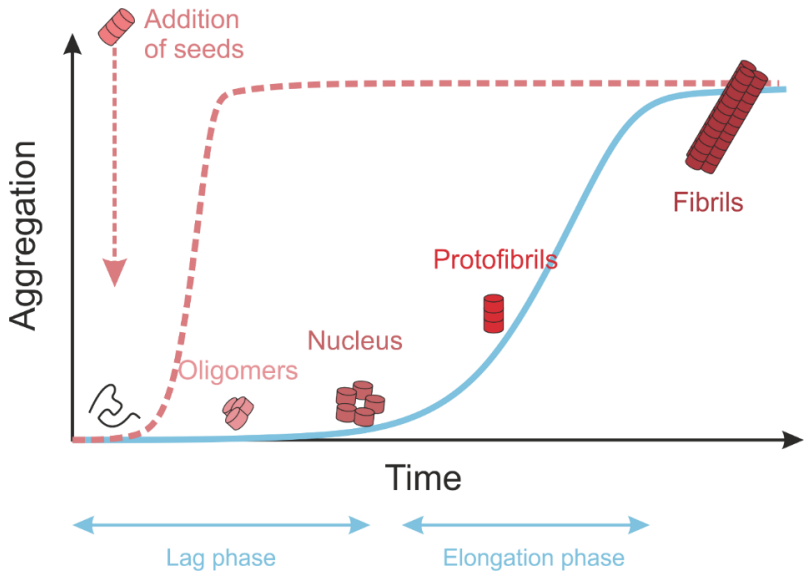


Figure 1-10. Schematic of the nucleated growth mechanism of amyloid formation. The growth of fibrils proceeds via two distinct stages: the lag phase or nucleation stage, and the elongation phase. The lag phase corresponds to the thermodynamically unfavourable formation of aggregation-competent nuclei from monomers that may occur via various metastable states of oligomers. This is the rate limiting step in the aggregation reaction. The elongation phase corresponds to the thermodynamically favourable step in which monomers are added to nuclei until the eventual formation of large fibrillar structures. Addition of preformed, aggregation-competent seeds to the start of the reaction decreases the lag time by overcoming the rate limiting step. Schematic species are examples of the dynamic structures that may form in the pathway. Adapted and redrawn form Eichner and Radford 2011[115].

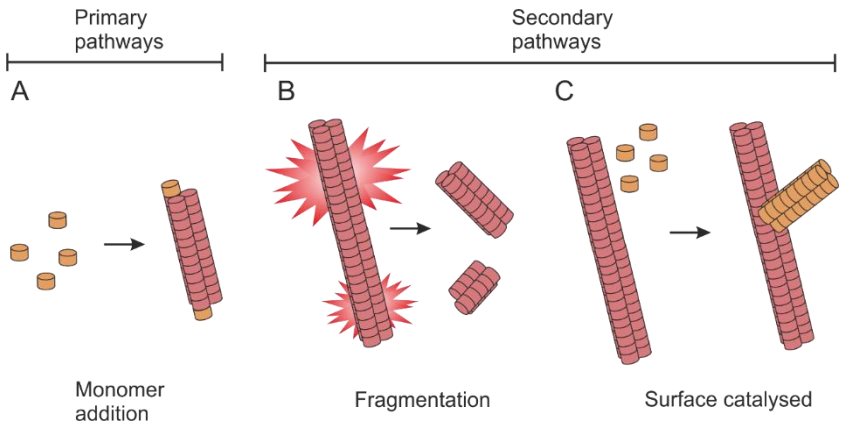


Figure 1-11. Processes that contribute to fibril formation. (A) The primary pathways by which amyloid fibrils can be propagated include monomer addition to aggregates or primary nucleation in which stochastic interactions of monomers results in new aggregates. (B) and (C) The secondary pathways enhance the rate of fibrillation kinetics in a number of ways and depend on the pre-existence of aggregates. (B) Shows the process of fragmentation of preformed aggregates in which mature aggregates fragment to reveal more fibril ends for the polymerisation process to take place. (C) Shows surface-catalysed events in which aggregate formation is dependent on the concentration of preformed fibrils. Figure adapted and redrawn from Cohen *et al*. 2012[112].

The complex mechanisms of aggregation make the application of quantitative chemical kinetics to analyse these systems difficult, despite the established roles of chemical kinetics in many biological systems such as enzymology or protein folding. A main reason for this complexity originates from the non-linear nature of the aggregation processes due to different competing mechanisms, which complicate the mathematics needed for integrated rate laws. However, recent progress has been made to overcome these obstacles and it has been demonstrated that chemical kinetics can be used to elucidate the molecular mechanisms of aggregation[114,116-119]. A recent *in vitro* study on amyloid-β 42 (Aβ42) utilised these techniques to interrogate the microscopic processes from macroscopic experimental measurements[114]. This study showed that the dominating process in the aggregation of Aβ42 was secondary nucleation catalysed by the fibril surface. This observation could be important with regards to disease progression as in a recent study, point mutations of Aβ42 were shown to dramatically alter these microscopic processes of aggregation[120].

1.4.3.1 Other structures in the amyloid pathway and the true nature of the pathogenic state

Despite the fact that it has been clear for many years that the appearance of amyloid deposits is associated with the onset of pathological events in protein misfolding diseases[121], the molecular mechanisms underlying amyloid disease pathways remains unclear. In protein folding diseases which result in systemic amyloidosis, the likely cause of disease is the presence of large quantities of aggregated proteins in vital organs[50]. However, in neurodegenerative disorders, no such correlation between the amount of fibrillary aggregates and disease progression has been detected such as in AD[122-124]. Indeed, post-mortem studies have shown a poor correlation between plaque load and disease severity in AD[125]. This suggests the possibility that the disease process is associated with misfolding events which result in cellular damage but do not necessarily lead to detectable aggregates[126].

It is the general consensus in the field that the oligomeric intermediates in amyloid fibril formation act as the main toxic species in amyloid disorders[126-131], rather than mature fibrillar structures (although this remains controversial[132]). Increasing evidence suggests that oligomeric species are universally observed during aggregation processes and are generically cytotoxic[104,122,133-135]. Their relatively disordered nature involves the display on their surface of chemical groups, especially hydrophobic residues[135-137] that would be sequestered in the core of folded proteins under normal physiological conditions. Thus, these groups are accessible to

the cellular environment and are able to form aberrant interactions with functional cellular components, ranging from other proteins to nucleic acids and lipid membranes[138,139].

During the process of protein aggregation, a heterogeneous array of oligomeric and intermediate species can form. These species can persist throughout the aggregation process and can populate an array of sizes and morphologies such as oligomers, toroids or protofibrils, it is unclear, however, whether these species are able to interconvert or progress to amyloid[140-143] (Figure 1-6). Such oligomeric species may be on-pathway precursors to aggregation or off-pathway, requiring conformational reorganisation to become aggregation-competent[104,144]. These species can either be toxic or non-toxic, and have varying dynamics and half-lives[104]. These properties of heterogeneity and the fact that oligomeric species are often transiently populated and metastable, has led to difficulties in their detection and characterisation. This remains a key challenge as they have been identified in various studies as membrane disrupting/ cytotoxic entities[145-147].

On-pathway amyloid assembly intermediates have been shown to contain β-sheet structure and become increasingly more stable as assembly progresses[104,106,148,149]. It was shown that the αSyn oligomers that form early in the amyloid formation pathway are relatively disordered species which undergo a conformational change resulting in more stable oligomers with a rudimentary amyloid-like core. The latter species are capable of causing pathogenic effects probably due to their large hydrophobic surface area[104]. Interestingly, however, on-pathway oligomeric ensembles that contain structural disorder have been identified for Aβ42 and, furthermore, were shown to be the culprits of cytotoxicity[150]. The high affinity of various oligomeric species for binding to membranes provides a common concept of cytotoxicity in which membrane integrity is disrupted during the aggregation process and so causing subsequent cell impairment[151]. Greater toxicity of prefibrillar and oligomeric species relative to larger amyloid fibrils is not unexpected since the smaller species have a higher surface area to volume ratio and thus have more surface groups exposed per volume that are available to interact aberrantly with cellular components as discussed above. Mature amyloid fibrils may, in fact, be relatively protective to cells (lower cytoxicity) and act as a reservoir sequestering the toxic species[136,152].

The high degree of heterogeneity in relation to the species formed in the aggregation pathway, the uncertainty relating to the toxic species imparts a high importance on the study of the initial interactions that initiate the aggregation cascade.

1.4.4 Amyloid and disease

The healthy state of a cell is characterised by a complex balance between protein expression, degradation and quality control[153]. Perturbations in this fine balance can lead to deleterious effects and disease unless kept under strict control. There are now 40 different proteins, many of which may cause several diseases with a multitude of different symptoms that are associated with the misfolding of normally functional proteins and peptides (Table 1-1)[90], and their subsequent aggregation into amyloid fibrils. Several of these diseases are neurodegenerative in which aggregates and their intermediates accumulate in the brain and exert their toxic effects. These include Alzheimer's disease (AD) and Parkinson's disease (PD), which are the two most frequent causes of dementia[154]. There are also many examples of non-neuropathic amyloidosis, whereby the disease is localised to specific organs or can affect several areas of the body systemically. In many cases, these diseases manifest sporadically in a population. There are, however, hereditary causes in amyloidosis. Furthermore, cases of transmissible aggregation-based diseases are known along with iatrogenic amyloidosis triggered by medical intervention. Examples of these different types of amyloidosis are given in Table 1-1.

The mechanism of toxicity for amyloidosis has not fully been elucidated, but is thought to occur via a gain of toxic function[50]. There is yet no single consensus and cellular toxicity is most likely to be mediated through a complex network of dysfunctions as illustrated in Figure 1-12. Evidence points to the fact that this may be a generic effect, as non-disease related proteins such as the SH3 domain from bovine phosphatidyl-inositol-3-kinase and the N-terminal domain of the *Escherichia coli* (*E. coli*) HypF protein, have been shown to form amyloid structures and exert toxicity when added exogenously to mammalian cells[145].

The toxicity of aggregates depends on both their morphology and location. Several amyloidoses are associated with extracellular protein deposits such as Alzheimer's disease[155] and type II diabetes[156], and have been shown to activate signal transduction pathways leading to apoptosis by interacting with cell surface receptors (Figure 1-12 i). For example, RAGE (receptor for advanced glycation end products) is a surface receptor which has been shown to interact with amyloid fibrils composed of amyloid beta (Aβ), islet amyloid polypeptide (IAPP) and prion protein (PrP), leading to the activation of cellular and immune stress responses[157,158]. There is also evidence that large extracellular fibrillar structures can disrupt membrane structure and hinder normal cellular activity (Figure 1-12 ii)[139,159-161]. As discussed previously, there is extensive evidence that oligomeric and prefibrillar structures also perturb cellular

membranes[139]. Their ability to form 'pore-like' morphologies in cellular membranes, alters ion homeostasis and dysregulates signal transduction before eventually leading to cell death[126,162,163] (Figure 1-12 iii).

Intracellular amyloid fibrils have also been reported and can be localised in various cellular compartments. These can damage cells via sequestration of factors essential for cell viability[138]. The example shown in Figure 1-12 iv is that of recruitment of components of the proteasome and numerous chaperones to an aggregate, preventing these essential proteins from carrying out their normal cellular functions[164,165]. This competition and burden on the proteostasis machinery can apply to all amyloid diseases and especially in age-related diseases where mechanisms of proteostasis are already in decline[166-168]. Indeed this is illustrated in a study of a yeast model of Huntington's disease, where cytosolic inclusions were shown to sequester the cellular chaperone Sis1p, inhibiting the clearance of non-amyloid inclusions of other, unrelated proteins[164].

Many of the amyloid disorders originate from natively unstructured precursor proteins or peptides as shown in Table 1-1. A number of amyloidoses, however, have folded precursor proteins, with native folds that are essential for function and are lost upon misfolding and aggregation (Figure 1-12 v). The loss of biological function has been suggested as the mechanism of toxicity in Huntington's disease. The huntingtin protein has been proposed to play a protective role against apoptosis[169], its aggregation can thus lead to neuronal death. Furthermore, in amyotrophic lateral sclerosis (ALS) also known as motor neurone disease, the protein superoxide dismutase 1 (SOD1) aggregates, preventing the processing of superoxide free radicals, leading to cell damage[125]. However, it must be noted that the loss of function mechanism in these two examples is still unclear[125], as evidence against this mechanism has been reported[169,170]. Although mice homozygous for Huntingtin die early in embryonic development[169], patients that are heterozygous for Huntingtin mutations have similar clinical features to patients that are homozygous[125]. As for ALS, SOD1 knockout studies do not show an increase in neuronal degeneration[170]. Additionally, aggregating proteins may form aberrant interactions with other cellular proteins, thereby leading to loss of function of the co-aggregate[138] (Figure 1-12 vi).

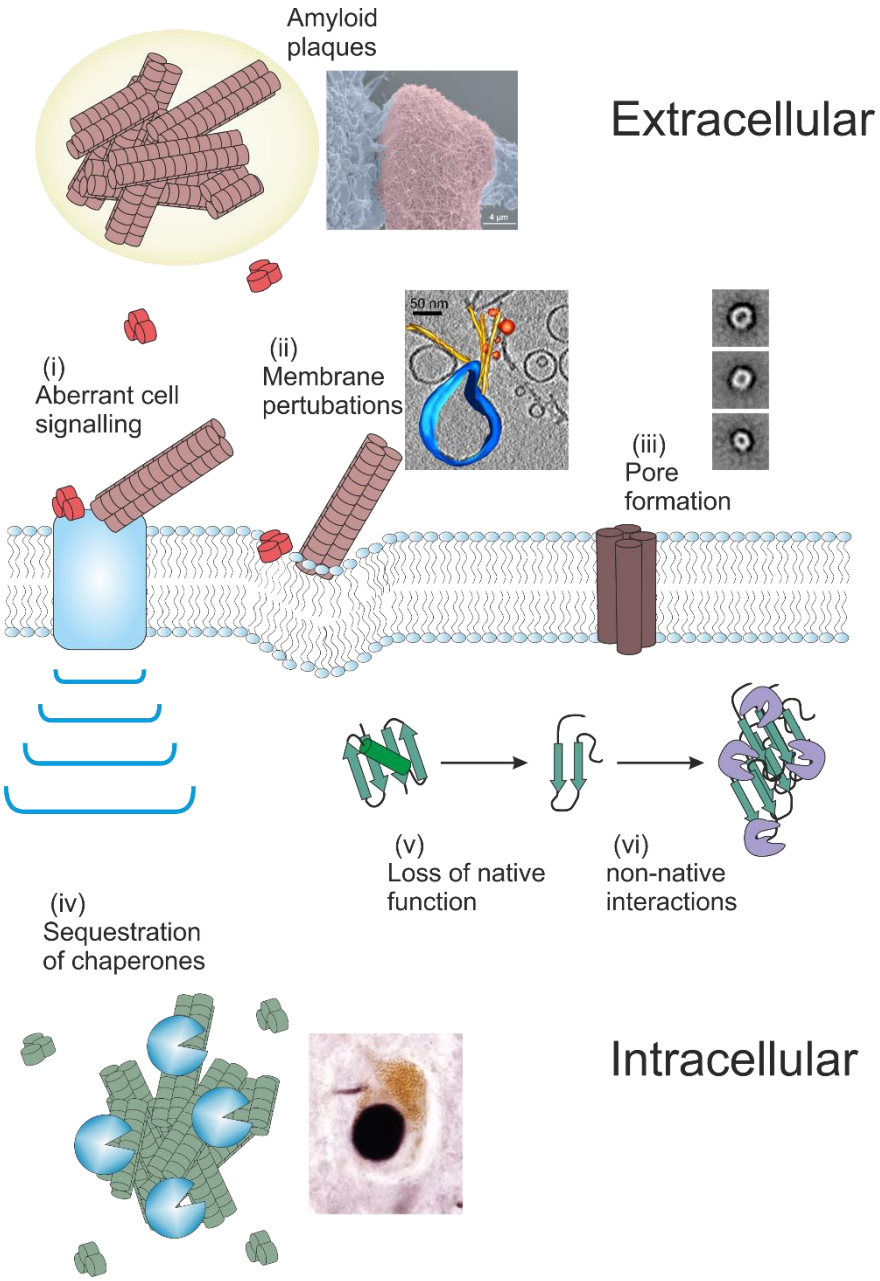


Figure 1-12. Mechanisms of amyloid cytotoxicity. Extracellular fibrillary aggregates are known as amyloid plaques, have a dense amyloid core, with a halo of oligomeric material often present. A scanning electron microscopy (SEM) image of an amyloid plaque from Aβ40[171] is shown inset (blue: cell bodies, red: extracellular plaques). Extracellular amyloid can cause cell toxicity by i) interactions with membrane receptors causing aberrant cell signalling; ii) interacting directly with cell membranes causing perturbations (cryo-electron tomography of liposome–fibril interactions is shown inset[172]); and iii) formation of pore like structures causing aberrant transmembrane transport (TEM projection averages of pore like structures from αSyn A53T[173]). Intracellular amyloid can iv) interact with cellular machinery such as chaperones and can sequester them away from their required function; misfolded proteins can be cytotoxic due to v) a loss of native function but also vi) via aberrant non-native interactions with various cellular components causing them to co-aggregate. A Lewy body positively stained for αSyn is shown inset as an example of an intracellular aggregate[174].

Disease	Aggregating protein or peptide	Number of residues	Native structure of protein or peptide
Neurodegenerative diseases			
Alzheimer's disease	Amyloid β peptide	40 or 42	Natively unfolded
Familial encephalopathy with neuroserpin inclusion bodies	Neuroserpin	410	α+β
Various neurodegenerative disorders	Actin	~400	Mostly α, some β
Neuroferritinopathy	Ferritin	175 or 183	All α
Spongiform encephalopathies	Prion protein or fragments thereof	253	Natively unfolded (residues 1‑120) and α-helical (residues 121‑230)
Parkinson's disease	α-Synuclein	140	Natively unfolded
Dementia with Lewy bodies	α-Synuclein	140	Natively unfolded
Frontotemporal dementia with Parkinsonism	Tau	352‑441	Natively unfolded
Amyotrophic lateral sclerosis	Superoxide dismutase	153	All-β, Ig like
Huntington's diseased	Huntingtin with polyQ expansion	3144	Largely natively unfolded
Spinocerebellar ataxias	Ataxins with polyQ expansion	816	All-β, AXH domain (residues 562‑694); the rest are unknown
Spinocerebellar ataxia 17	TATA box-binding protein with polyQ expansion	339	α+β, TBP like (residues 159‑339); unknown (residues 1‑158)
Spinal and bulbar muscular atrophy	Androgen receptor with polyQ expansion	919	All-α, nuclear receptor ligand-binding domain (residues 669‑919); the rest are unknown
Hereditary dentatorubral-pallidoluysian atrophy	Atrophin-1 with polyQ expansion	1185	Unknown
Familial British dementia	ABri	23	Natively unfolded
Familial Danish dementia	ADan	23	Natively unfolded
Non-neuropathic systemic amyloidoses			
AL amyloidosis	Immunoglobulin light chains or fragments	~90	All-β, Ig like
AH amyloidosis	Immunoglobulin heavy chains or fragments	~220	All-β, Ig like
AA amyloidosis	Fragments of serum amyloid A protein	76‑104	All-α, unknown fold
Familial Mediterranean	Fragments of serum	76‑104	All-α, unknown fold

fever	amyloid A protein		
Senile systemic amyloidosis	Wild-type transthyretin	127	All-β, prealbumin like
Familial amyloidotic polyneuropathy	Mutants of transthyretin	127	All-β, prealbumin like
Hemodialysis-related amyloidosis	β2-microglobulin	99	All-β, Ig like
ApoAI amyloidosis	N-terminal fragments of apolipoprotein AI	80‑93	Natively unfolded
ApoAII amyloidosis	N-terminal fragment of apolipoprotein AII	98	Unknown
ApoAIV amyloidosis	N-terminal fragment of apolipoprotein AIV	~70	Unknown
ApoCII amyloidosis	ApoCII	79	α + unstructured
ApoCIII amyloidosis	ApoCIII	79	α + unstructured
Finnish hereditary amyloidosis	Fragments of gelsolin mutants	71	Natively unfolded
Lysozyme amyloidosis	Mutants of lysozyme	130	α+β, lysozyme fold
Fibrinogen amyloidosis	Variants of fibrinogen α-chain	27‑81	Unknown
Icelandic hereditary cerebral amyloid angiopathy	Mutant of cystatin C	120	α+β, cystatin like
Non-neuropathic localised diseases			
Type II diabetes	Amylin, also called islet amyloid polypeptide (IAPP)	37	Natively unfolded
Aortic media amyloidosis	Lactadherin C2-like domain	50	Unfolded
LECT2 amyloidosis	Leukocyte chemotactic factor-2	151	Unknown
Localised cutaneous amyloidosis	Gelactin-7	136	All-β
Hypotrichosis simplex of the scalp	Corneodesmosin	529 (truncations cause amyloid)	Unknown
Calcifying epithelial odontogenic tumours	Odontogenic ameoblast-associated protein	153	Unknown
Senile seminal vesicle amyloidosis	Semenogelin 1	462	Unknown
Medullary carcinoma of the thyroid	Calcitonin	32	Natively unfolded
Atrial amyloidosis	Atrial natriuretic factor	28	Natively unfolded

Hereditary cerebral haemorrhage with amyloidosis	Mutants of amyloid β peptide	40 or 42	Natively unfolded
Pituitary prolactinom	Prolactin	199	All-α, 4-helical cytokines
Injection-localized amyloidosis	Insulin	21 + 30	All-α, insulin like
Injection-localized amyloidosis	Enfuvirtide	36	Unstructured
Aortic medial amyloidosis	Medin	50	Unknown
Hereditary lattice corneal dystrophy	Mainly C-terminal fragments of kerato-epithelin	50‑200	Unknown
Corneal amyloidosis associated with trichiasis	Lactoferrin	692	α+β, periplasmic-binding protein like II
Cataract	γ-Crystallins	Variable	All-β, γ-crystallin like
Calcifying epithelial odontogenic tumours	Unknown	~46	Unknown
Pulmonary alveolar proteinosis	Lung surfactant protein C	35	Unknown
Inclusion-body myositis	Amyloid β peptide	40 or 42	Natively unfolded
Cutaneous lichen amyloidosis	Keratins	Variable	Unknown

Table 1-1. Human disorders associated with amyloid deposition. Adapted from Chiti and Dobson 2006[50] and Sipe *et al* 2016[90].

1.5 The physiology and pathophysiology of α-synuclein aggregation

α-synuclein (αSyn) is an intrinsically disordered protein (IDP)[175] expressed predominantly in the neurons of the central nervous system (CNS) and localised at pre-synaptic termini[176-178]. Its aggregation into amyloid has been shown to cause various, prevalent and debilitating human diseases[179]. This thesis will primarily focus on αSyn and the mechanisms by which this protein self-associates.

αSyn is a 140-amino acid protein that was first described in 1988 when isolated from the Pacific electric ray *Torpedo californica* as a neuron-specific protein localised to presynaptic nerve termini and nuclei, hence the name *synuclein*[180]. It was not until 1997 that αSyn was associated with the neurodegenerative amyloid disorder Parkinson's disease (PD) after a mutation in the SNCA gene encoding αSyn (causing the missense mutation A53T) was found to be associated with familial cases of early-onset PD[181]. Furthermore, αSyn aggregates were

found as the major components of Lewy bodies (LB), the hallmarks of PD[182]. Although extensive research has been carried out on the protein in the subsequent decades due to its association with significant human diseases, it is not yet fully clear what the physiological functions of αSyn are[183,184].

1.5.1 Physiological functions of αSyn

As of yet, there is not complete clarity as to what the physiological functions of αSyn are[183,184]. This is likely due to the subtle and complex effects that modulating the expression and structure of αSyn have on cells[185,186]. Even in (knock out) KO mice models, the effect is not fatal and there are very few alterations, such as slight differences in the populations of pre-synaptic vesicles, observed[185,186] indicating that there may be some degeneracy in the functions of αSyn. These complexities in understanding the functions of αSyn may also result from the structure of the protein. αSyn is an IDP (Section 1.2) and these proteins by their nature have evolved to be disordered for various reasons as outlined previously. One of these is that intrinsic disorder allows a protein to interact with various partners, IDPs are therefore utilised as 'hub' proteins in a cell[32,36], controlling pathways such as cell signalling where fine 'volume' control rather than a binary on/ off switch is needed. This may indeed be the case for αSyn. In a proteomic study of αSyn using a SILAC technique (stable isotope labelling by amino acids in cell culture), 587 proteins were identified to form complexes with αSyn in dopaminergic cells, 141 of these proteins displayed significant changes in their relative abundance when the cell were treated with rotenone, a chemical that induces PD like toxicity[187]. The promiscuity of αSyn physiological function is certain to contribute to the complicated picture of both function and dysfunction.

Despite this lack of consensus, research has shown that αSyn carries out its functions primarily in the neurons of the CNS where it is estimated to constitute 1% of total soluble cytosolic brain protein[176]. αSyn has been shown to be localised at the presynaptic termini of neurons where it is proposed to regulate synaptic transmission (Figure 1-13) by binding lipid membranes. In presynaptic termini, monomeric αSyn exists in equilibrium between membrane bound and free states[188]. The equilibrium is tightly regulated, and it has been estimated that 15% of αSyn is membrane bound within synaptic termini[44]. More specifically, αSyn has been postulated to be involved in the packing of neurotransmitter vesicles in the distal reserve pool of synaptic vesicles[189-191] and subsequent trafficking to the site of synaptic vesicle release[185,186] (Figure 1-13). This has been shown in various studies in which αSyn expression is reduced via KO mice

or antisense oligonucleotides, resulting in depletion of the docked and reserve pool vesicles, along with impairments in the replenishment of docked pools from the reserve pool[185,186,192].

Over-expression of αSyn has been shown to impair synaptic vesicle exocytosis and neurotransmitter release[193-195]. It also leads to a decrease in readily releasable vesicles[196] and reduces the size of the recycling pool of synaptic vesicles[195]. Moreover, αSyn has been shown to inhibit the inter-synaptic trafficking of vesicles, maintaining the recycling pool homeostasis[197]. The role αSyn plays in synaptic homeostasis is not limited to its interactions with synaptic vesicles: as mentioned previously, αSyn interacts with a plethora of different proteins involved in the regulation of synaptic function and plasticity. These include interactions with various protein members of the soluble N-ethylmaleimide–sensitive factor attachment protein receptor (SNARE) complex which is directly involved in the process of neurotransmitter release[198], such as the vesicular SNARE (v-SNARE) protein vesicle associated membrane protein 2 (VAMP2)[199] (Figure 1-13). αSyn also acts as a chaperone, promoting SNARE complex assembly both *in vivo* and *in vitro* through its membrane binding properties[199].

A recent study suggested a physiological 'double anchor' mechanism in which αSyn can bind the membranes of synaptic vesicles in two regions of its sequence, involving residues 1-25 and residues 65-97. The membrane-binding affinities of these regions are finely tuned so that αSyn binds two synaptic vesicles in a double anchor mechanism, causing them to cluster and assemble. PD mutations (A30P and E46K) disrupt the fine balance of αSyn membrane affinity and may cause increased exposure of aggregation-prone regions leading to pathological self-self-interactions[200].

It seems that αSyn function is not essential under normal conditions[201], but that it subtly regulates synaptic vesicle homeostasis. Under conditions of stress, however, its neuroprotective properties become more prominent and essential[184,202].

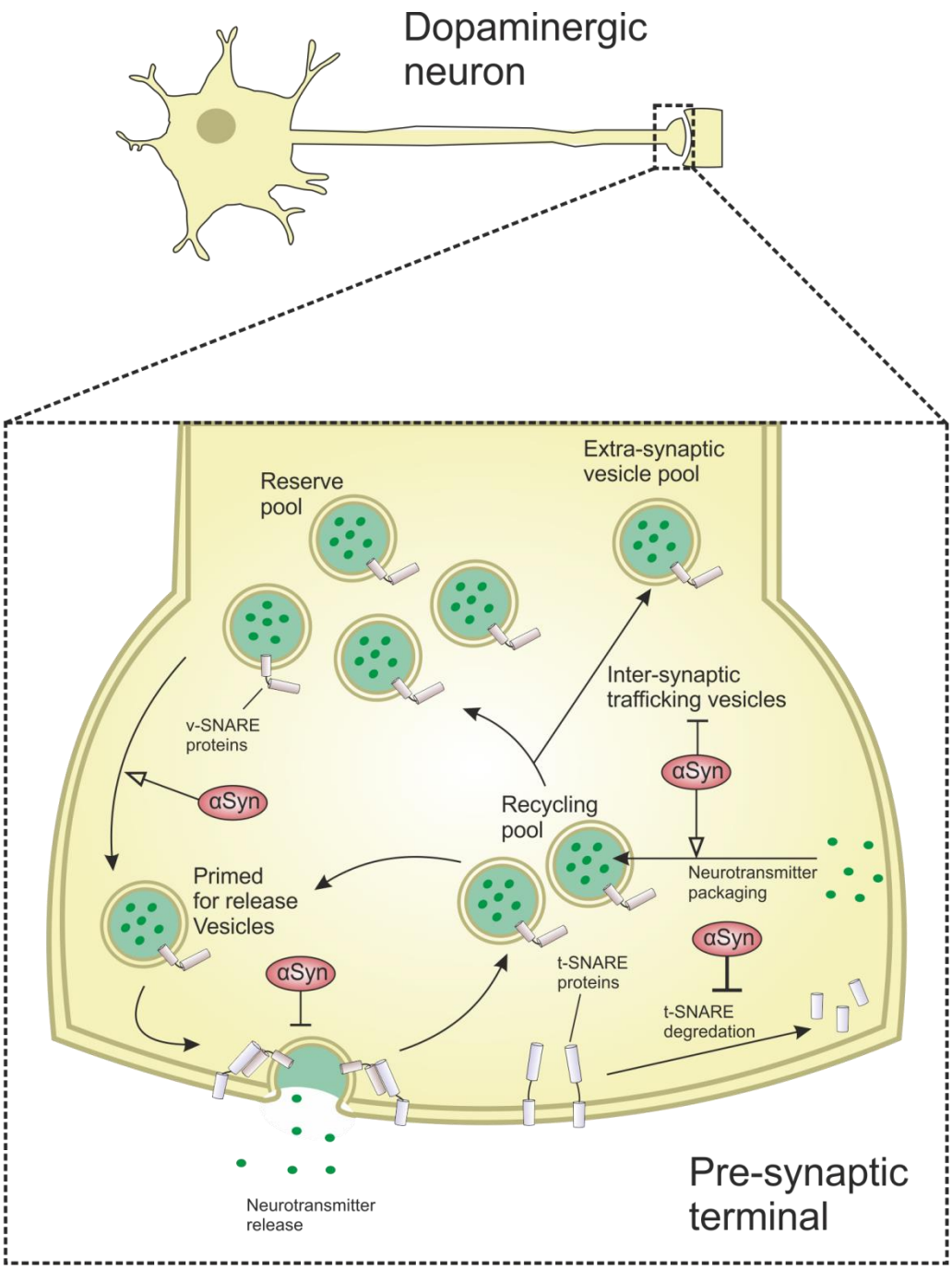


Figure 1-13. Physiological role of αSyn at pre-synaptic termini shows the various regulatory functions αSyn exerts over synaptic vesicles. αSyn is shown in red ellipses, αSyn mediated inhibitory regulatory activity is shown by a flat arrow head and αSyn mediated regulation is shown by an open arrow head. αSyn promotes neurotransmitter packaging, maintains the reserve pool of synaptic vesicles and promotes subsequent trafficking to the site of vesicle release. Chaperone activity has been observed with members of the SNARE complex where it increases the stability of target SNARE (t-SNARE) proteins. Adapted and redraw from Lashuel *et al.* 2013[184].

1.5.2 Structural properties of αSyn and their context dependence

At the primary structural level, the sequence of αSyn can be divided into three major domains as shown in Figure 1-14. The N-terminal domain (residues 1-60) is made up of a series of imperfect repeats with the consensus sequence KTKEGV separated by inter-repeat regions, this region has amphipathic properties. The central domain (residues 60-95) is highly hydrophobic and is known as the non-amyloid β component (NAC), named so as it was originally discovered through its co-purification with Aβ from amyloid plaques in AD patients[203]. The NAC region has been shown to be key in the misfolding and aggregation of αSyn[203,204] and forms the core of amyloid fibril architecture[100]. This central domain also contains additional KTKEGV imperfect repeat motifs. There are varying numbers of the imperfect repeats quoted in the literature due to the difficulty in identifying them, however there is anything from 6-9 repeats in the N-terminal region and NAC domain of αSyn depending on how one counts (Figure 1-5). The first two domains have been shown to be involved in lipid binding thought to be mediated by the imperfect repeats as they are similar to those found in apolipoproteins[205]. Consequently, it has been proposed that the lipid interaction of αSyn could be similar to that of apolipoproteins[205]. Lipid binding induces α-helical structure formation in these regions (Figure 1-14 A)[42,206,207]. The C-terminal domain (96-140) is natively unstructured and is enriched with acidic and proline residues, providing flexibility in the protein and providing some protection against aggregation[208,209].

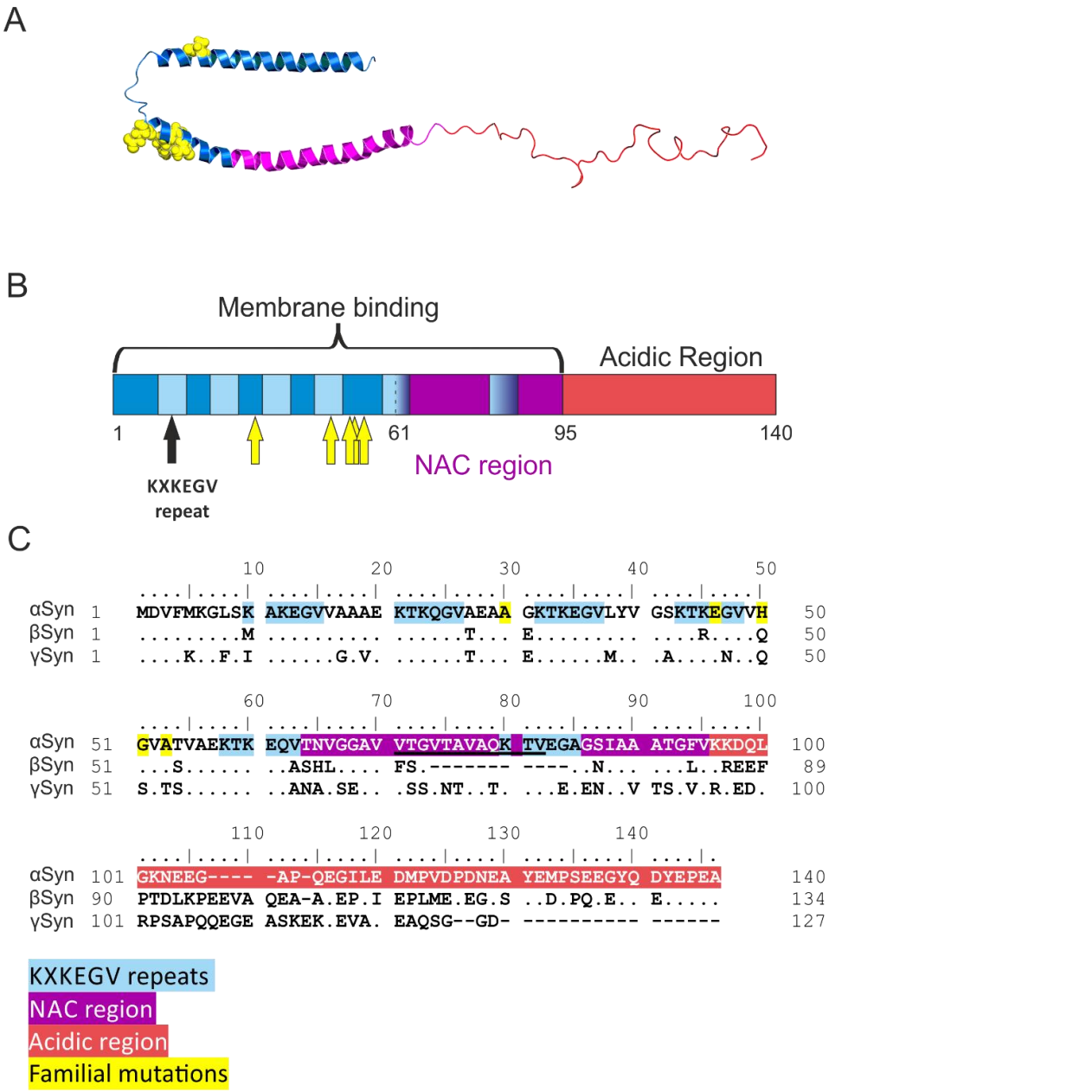


Figure 1-14. SDS micelle induced structure, primary structural elements and a sequence alignment of αSyn. (A) NMR structure of αSyn when bound to SDS detergent micelles[206] (PDB 1XQ8). Lipid or detergent induces α-helical structural formation in the N-terminus and the central NAC region in the form of a broken helix (residues 3-37 and 45-92)[206]. Amino acid substitutions found in familial mutants are highlighted as space filling spheres. The structure is coloured as in B and C. (B) The primary structure of αSyn can be divided into three domains, the aliphatic N-terminal region in blue with KXKEGV consensus repeat motifs (light blue), the central highly hydrophobic NAC region (magenta), and an unstructured C-terminal domain (red) that is highly acidic and contains a high proportion of proline residues. Yellow arrows show sites of familial PD mutations. (C) Sequence alignment of αSyn and the human homologues β- and γSyn. The central NAC region 71-82 that is both necessary and sufficient for aggregation is underlined. This region is not present in βSyn. Dots show residues that are identical; dashes show residues absent in the sequence. Numbers at either side of the alignment the residues number of each protein at the start and end of each row. A key of the colour coding is at the bottom of the figure and is consistent in A, B and C.

There is much evidence showing that αSyn exists as an IDP lacking extensive structure[210,211] (Figure 1-15 A-C). Initial studies showed that αSyn had a larger hydrodynamic radius (R_H), (determined by size exclusion chromatography (SEC)) and sediments more slowly (determined by sucrose gradient ultracentrifugation) than globular proteins of a similar mass[175]. Circular dichroism (CD) and Fourier transform infrared (FTIR) spectroscopy studies showed that the majority of the molecule is unstructured at neutral pH[210]. However, small angle X-ray scattering (SAXS) data showed that αSyn has a smaller radius of gyration (R_g) than that expected of a random coil polypeptide of the same length (40 Å compared to a theoretical R_g of approximately 52 Å for a 140 residue disordered polypeptide chain)[210]. Therefore, αSyn is essentially disordered, but is more compact than a random coil and it may contain some transient intramolecular structure[210]. This relative compactness of αSyn has been observed in various other studies, and it has been proposed that the origin of this structure is primarily due the clustering of hydrophobic residues, as well as showing that αSyn contains transient long range contacts[212-218].

1.5.2.1 Transient conformations

αSyn can be described as a highly diverse ensemble of preferred conformations and not just simply a random structure, similar to the behaviour of other IDPs [34,35]. These preferred conformations involve distil transient interactions of the proteins, the regulation of which is most likely key in the continuum of function and pathological misfolding and aggregation. Transient long range interactions have been proposed between the C-terminus and the NAC region of the protein[215,219] (Figure 1-15 D). These transient interactions have been proposed to form a capping interaction of the central hydrophobic NAC region with the disordered C-terminal region and in doing so, protects it from solvent exposure[215,216,220], reinforcing the hypothesis that the C-terminal domain is a solubilising region in αSyn[208,209].

In a similar manner to other IDPs (as discussed in Section 1.2), the unstructured nature of αSyn is a consequence of its relatively low hydrophobicity and high net charge. The protein, therefore, is sensitive to the environment: an increase in the relative hydrophobicity or change in the net charge can induce partial folding of the protein[210] (Figure 1-15 A-C). The high net negative charge of αSyn (pI 4.7) at physiological pH can be neutralised upon acidification. Hence, at pH 3.0 αSyn exhibits a more ordered secondary structure, and becomes more compact, developing a rudimentary nucleus with a tightly packed core and high affinity for ANS (8-anilinonaphthalene-1-sulfonic acid)[210] (Figure 1-15 A-C). Indeed, αSyn was shown to be more aggregation prone at lower pH[210].

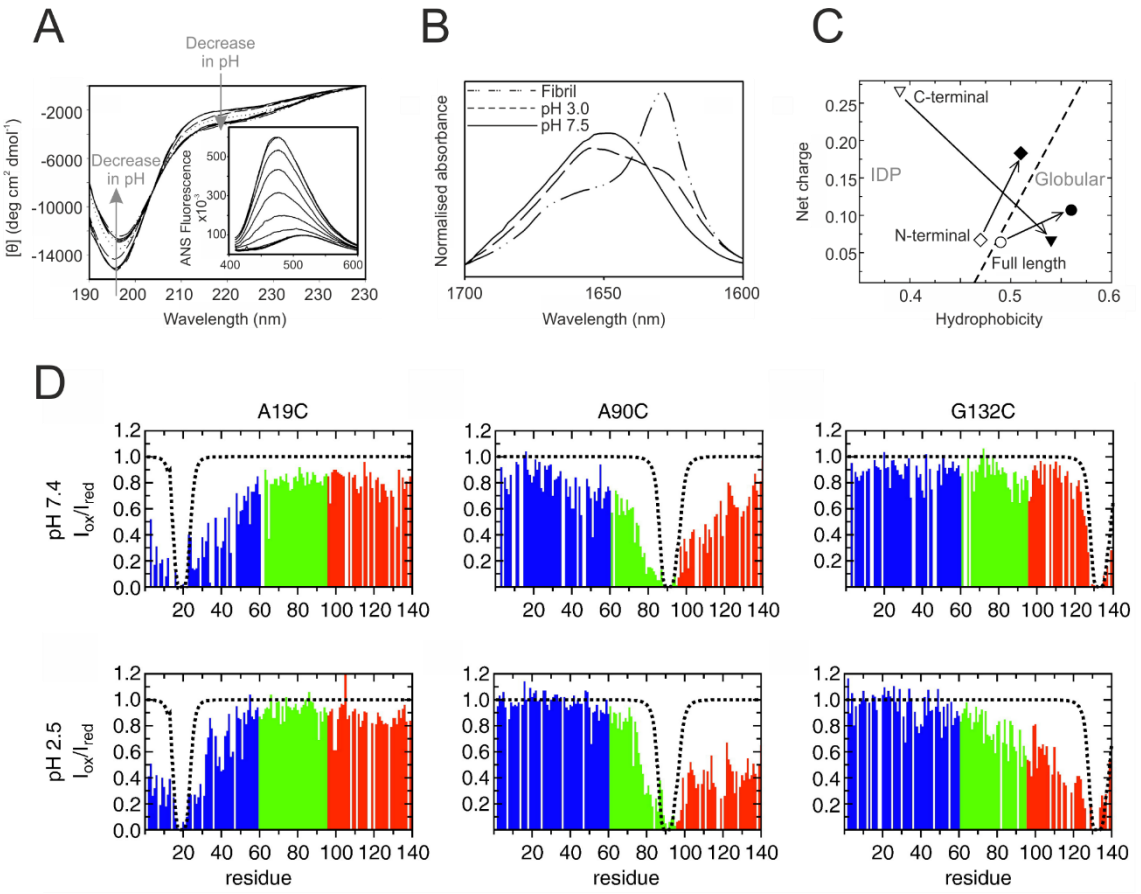


Figure 1-15. αSyn conformational changes driven by changes in pH. (A) Far-UV CD spectra of αSyn as a function of pH[210]. pH values were 8.9 (solid line) 7.3, 6.4, 5.3, 4.3, 3.5, 2.7, 1.7, and 0.9 in order of the increase in negative $[\theta]_{222}$ value. At neutral pH αSyn is essentially a random coil. With decreasing pH, the spectra changes and depicts an induced partial formation of secondary structure. Inset shows ANS fluorescence spectra measured at pH 8.2, 7.5, 6.6, 5.4, 4.6, 4.0, 3.7, 3.1, 2.8, and 2.5 (in order of increasing intensity). A decrease in pH leads to a characteristic blue shift in λ_{max} indicating a partially folded structure with solvent exposed hydrophobic residues. (B) FTIR of the amide I region measured at pH 7.5 (solid line) showing a typically unstructured polypeptide, pH 3.0 (dotted line) showing spectral changes, indicative of increased ordered structure, and of fibrils (dot-dashed line) for comparison[210]. (C) Charge-hydrophobicity plot of αSyn at neutral (open symbols) and at pH 2.5 (filled symbols)[221]. The dashed line represents the boundary between IDP and globular proteins. This plot shows that αSyn should have more structure at acidic pH, driven primarily by a gain in structure of the C-terminus. (D) Intramolecular paramagnetic relaxation enhancement (PRE) NMR intensity ratios of amide protons in spin labelled αSyn[221]. Positions of spin labels are shown at the top of each column. The top panel shows data at pH 7.4, the bottom panel shows data at pH 2.5. Blue green and red represent the N-terminal, NAC and C-terminal domains respectively. Broken lines are the theoretical PRE values of αSyn without any long-range contacts. This data shows a pH induced collapse of the C-terminus. Images adapted from references as indicated.

The extra compaction of αSyn at acidic pH has been shown to primarily originate from the compaction of the C-terminal region[220,221] (Figure 1-15 D) (the C-terminal of αSyn has a net charge of -8, 8 negative and 0 positive residues (Figure 1-14 B)).

1.5.3 Protein aggregation in Parkinson's disease and other synucleopaphies

Several observations have established the involvement of αSyn in the pathogenesis of PD and other synucleopathies. As mentioned previously (Section 1.5), the association between PD and αSyn was first recognised when a mutation (A53T) in the protein encoding gene was found in familial cases of early-onset PD[181], as well the aggregates of αSyn being found to be the main component of LBs, the hallmarks of PD[182]. Other autosomal dominant missense mutations have been shown to result in other cases of familial early-onset PD and are all localised in the N-terminal region (A30P, E46K, H50Q, G51D, A53T, A53E)[181,222-227]. Genomic duplications or triplications that contain the αSyn locus also result in autosomal dominant forms of familial PD[228,229]. Moreover genome wide association studies (GWAS) have identified single nucleotide polymorphisms (SNPs) in αSyn as risk factors that increase the susceptibility to sporadic PD[230,231]. Among the SNPs is a risk variant in a non-coding distal enhancer element that leads to increased αSyn expression[232]. These pieces of genetic evidence highlight the crucial link between αSyn and PD pathology.

In PD, dopaminergic neurons of the substansia nigra pars compacta (SNc) are lost in the basel ganglia, an area of the brain responsible for co-ordinating fine motor control, which ultimately leads to the onset of Parkinsonism symptoms such as bradykinesia, muscle rigidity, resting tremors and postural instability[233]. Dopaminergic neurons in PD also show selective vulnerability to the aggregation and fibrillation of αSyn. αSyn is also implicated in other disorders and is not simply linked to one disease pathology[179], hence the complexity in determining the mechanism of toxicity during the misfolding and aggregation of this protein. αSyn aggregates in other synucleopathies including dementia with Lewy bodies (DLB)[234], multiple system atrophy (MSA)[235] and various lysosomal storage disorders including Gaucher's disease[236]. Importantly, αSyn plays a role in the aggregation of Aβ and tau in AD pathology[237-241].

Disease	Symptoms	Pathology
Parkinson's disease (PD)	– Parkinsonism (bradykinesia, muscular rigidity, resting tremors and postural instability) – Nonmotor symptoms (constipation, impaired olfaction and rapid-eye movement sleep behaviour disorder) – Cognitive impairment	– SNc dopaminergic degeneration – Variable neuron loss in areas including locus coeruleus, dorsal motor nucleus of the vagus and the olfactory bulb – αSyn Lewy body and neurite pathology in neurons
Parkinson's disease with dementia (PDD) or dementia with Lewy bodies (DLB)	– Parkinsonism and dementia – In DLB versus PDD: fewer resting tremors, bilateral parkinsonism	Cholinergic/SNc dopaminergic degeneration – αSyn Lewy body and neurite pathology in neurons – Aβ amyloid plaques and tau tangles
Multiple system atrophy (MSA)	– Parkinsonism, cerebellar ataxia, autonomic failure – Nonmotor symptoms (sexual dysfunction, urinary incontinence and rapid-eye movement sleep)	– SNc/olivopontocerebellar degeneration – αSyn pathology in oligodendrocytes
Gaucher's disease	– Type I adult-onset (thrombocytopenia, anaemia, hepatosplenomegaly and bone pain) – Type II (infant)/III (juvenile) neuropathic form (seizures, cognitive impairment and oculomotor problems)	– αSyn Lewy body and neurite pathology in some patients
Additional lysosomal storage disorders	– Multisystem disorder	– αSyn Lewy body and neurite pathology in some patients
Neurodegeneration with brain iron accumulation	– Variable symptoms that may include dystonia, muscle rigidity, spasticity and ataxia	– Iron accumulation in the globus pallidus and SNc – αSyn Lewy body and neurite pathology in some patients
Alzheimer's disease (AD)	– Progressive memory loss – Cognitive impairment	– Cortical and CA1 hippocampal degeneration – Aβ amyloid plaques and tau tangles – αSyn Lewy body and neurite pathology in some patients

Table 1-2. Summary of diseases associated with αSyn toxicity. Adapted from Wong and Krainc 2017[179]

1.5.3.1 The neurotoxic species of αSyn

Recent studies have yielded insights into the molecular structure of the amyloid fibrils of αSyn[100,242,243] including a 3D structure at atomic detail of the full length protein showing a β-serpentine Greek key morphology[100]. Despite the realisation that oligomers may play a leading role, both generally for amyloidogenic proteins (as discussed in Sections 1.4.3 and 1.4.3.1) and for αSyn, and the fact that insoluble inclusions rarely correlate with disease progression, fibrils of αSyn have been shown to be toxic to cells in various studies[244-247]. In fact one study proposed that αSyn fibrils may be 1000-times more toxic than their precursors[161], another showed that injections of fibrils (rather than oligomers) into the SNc of rats induced the greatest amount of motor impairment, dopaminergic cell loss and synaptic impairment[244].

Conversely, there is a large body of evidence that points to the formation of αSyn oligomeric species as major toxic species in the aggregation pathway. A range of oligomers with a plethora of different morphologies has been identified, including spherical, elliptical, circular tubular and flat[104,128,141,248]. Similarly, oligomers of various secondary structures have been identified and seem to point to structural transitions[104] from α-helical to β-sheet, as the population shifts from early to late aggregation stages[249,250]. Much of the studies on αSyn oligomers has been focused on these species as toxic elements[251]. The fact that the PD familial mutations A30P and A53T, increase the rate of oligomerisation and not fibrillation suggests that oligomeric species may be the most toxic forms of αSyn[128]. Additionally, identification of other variants prone to oligomerisation and the demonstration of their increased toxicity relative to WT αSyn also supports this hypothesis[252-254].

It has to be noted, however, that physiological structured oligomers (or multimers) have been reported for αSyn *in vivo* and have been shown to be protective against toxicity[46,255-259]. The population of multimers has been shown to be modulated by the sequence/ presence of the N-terminal KTKEGV repeat motifs. PD familial mutations and synthetic mutations that occurred within the repeat motifs were found to decrease the multimeric conformation leading to increased toxicity[46,258]. These observations remain controversial due to the great body of preceding evidence showing αSyn as an intrinsically disordered monomer.

1.5.4 Effect of PD familial mutations

Currently, six familial mutations associated with early onset PD have been identified (A30P, E46K, H50Q, G51D, A53T, A53E)[181,222-227] (Figure 1-14). Extensive studies have focused on characterising the conformation and aggregation characteristics of these mutants. None of the

mutants promote a major structural change in the monomeric conformation of αSyn[260-262] with only A30P displaying reduced propensity to form α-helical structures[205].

The mutants are all localised in the N-terminal region of αSyn which is likely to affect the protein's membrane binding properties. Indeed, A53T and H50Q exhibit a higher propensity for membrane interaction[263], whereas A30P[264], G51D[265] and A53E[266] attenuate this propensity. As discussed in Section 1.5.2, transient, long range interactions play an important role in the conformation of αSyn and are partially driven by electrostatics, which are likely disrupted by these mutations affecting the stability of the protein in its native and misfolded states[216]. Aberrant intramolecular contacts may promote the formation of folded intermediates that would increase self-assembly propensity[267].

The aggregation propensity of αSyn is altered by PD familial mutations. E46K, H50Q and A53T increase aggregation propensity measured *in vitro*[260-262,268-271] whereas G51D and A53E[266,270,271] have the opposite effect. A30P is more prone to oligomerisation at the expense of fibril formation[260,268-270]. Familial mutations also display varied fibril morphologies including differences in diameter, periodicity and length[271,272]. However, despite intense study, the pathological role of these mutations remains unknown.

1.5.5 The synuclein homologues

αSyn is a member of a small family of highly homologous synuclein proteins Other members of this family include β- and γ-synuclein (βSyn, γSyn), proteins that are expressed predominantly in neurons, although their cellular location varies. βSyn, like αSyn, is expressed predominantly in neurons of the CNS at presynaptic termini, whereas γSyn is predominantly expressed in neuronal cells of the peripheral nervous system, abundant in spinal cord, sensory ganglia, and retina as well as in metastatic breast cancer and other cancer tissue[273]. The sequence similarity between the homologues and αSyn is 78% and 60% for βSyn and γSyn respectively (a sequence alignment of the synucleins is shown in Figure 1-14). β- and γSyn aggregate slower than αSyn, if at all[274,275], however there is evidence from cell-based fluorescence studies that all of the synucleins dimerise[276] and crosslinking studies have shown that all the synuclein homologues can form higher order physiological multimers[46]. Multiple studies have shown that βSyn, and in some cases γSyn, are protective against the aggregation of αSyn both *in vitro* and *in vivo*[275,277-280]. Intermolecular NMR PRE experiments have shown that βSyn interacts with αSyn directly forming a transient dimer with high specificity and low affinity[280]. The αSyn homodimer interactions were shown to be more heterologous and showed interactions between the N-terminal region (residues 36-44) and the C-terminal region (residues 124-140) in both a "head

to head" and "head to tail" conformations. In contrast, the heterodimeric αSyn-βSyn was observed to interact in a more specific manner via a "head to tail" interaction between a more extensive C-terminal region of βSyn (residues 105-134) and the N-terminal of αSyn. Homodimeric interactions of βSyn were not observed. Heterodimeric αSyn-βSyn were shown to interact in a higher affinity complex (K_D = 100 μM) that the αSyn-αSyn homodimer (K_D = 500 μM). This tighter binding is thus the bases of inhibition of αSyn aggregation by βSyn. The formation of stable mixed tetramers of αSyn with βSyn may act as regulators of αSyn *in vivo*[281], potentially acting as chaperones to prevent αSyn from misfolding and aggregation.

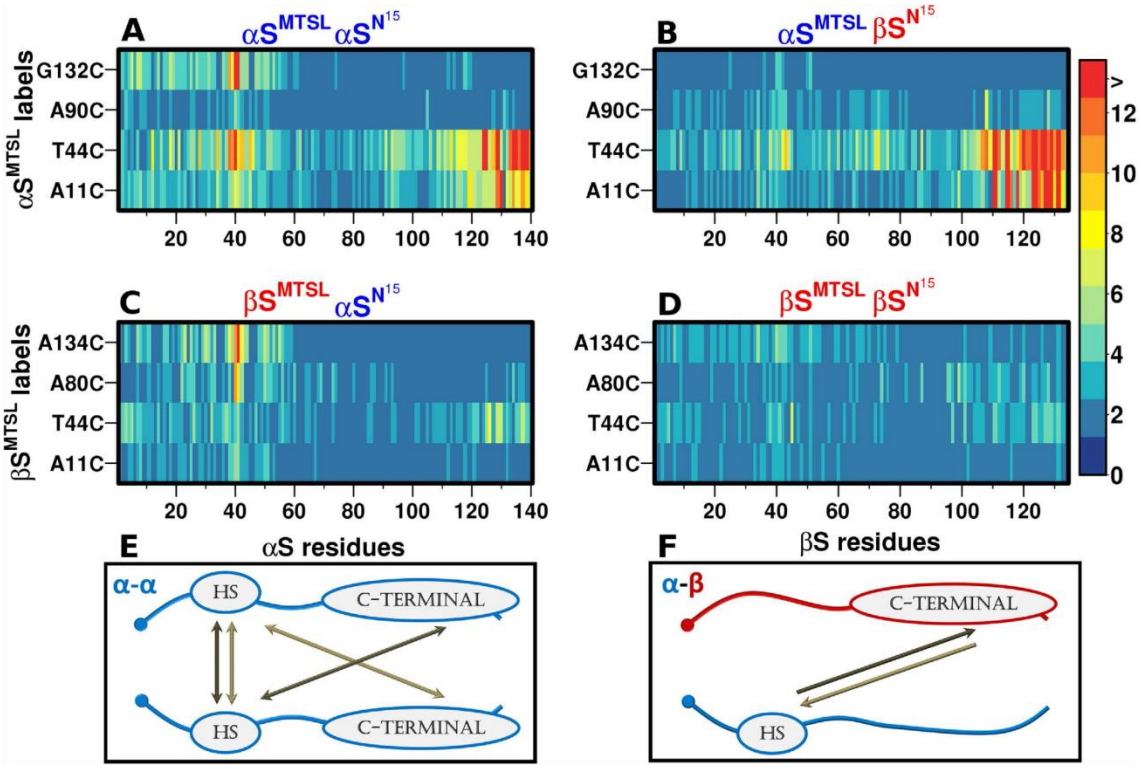


Figure 1-16. Intermolecular PRE NMR experiments of αSyn-αSyn homo-, and αSyn-βSyn heterodimeric interactions reveal distinctive profiles. A-D) Contact maps of PRE NMR data showing homo- and heterodimeric interactions of both αSyn and βSyn. Each strip represents the colour coded value of the residue-specific inter-chain PRE rate ($H_N\Gamma_2$) induced by the MTSL paramagnetic label on each of the proteins indicated. In all PRE experiments, both ^{15}N isotopically labelled synuclein (NMR "visible") and MTSL ^{14}N synuclein (NMR "blind") were present. This allows the observation intermolecular interactions as the MTSL labelled has an observable PRE effect on ^{15}N synuclein. A-D show four heat maps for all possible permutations of the spin label and NMR detectable chains of α S and β S: (A) ^{14}N-αSyn-MTSL/^{15}N-αSyn, (B) ^{14}N-αSyn-MTSL/^{15}N-βSyn, (C) ^{14}N-βSyn-MTSL/^{15}N-αSyn, (D) ^{14}N-βSyn-MTSL/^{15}N-βSyn. Each strip of the contact maps corresponds to a spin label in different positions throughout the ^{14}N labelled protein. The x-axis correspond to residue number of the ^{15}N protein. E and F) Schematic representations of the possible interactions of αSyn-αSyn (E) and αSyn-βSyn dimeric interactions. HS stands for "hot spot" and corresponds to residues 38-45. Taken from Janowska *et al* 2015[280].

Many aspects of αSyn intermolecular interactions remain to be determined and will be the focus of this thesis.

1.6 Force spectroscopy

One of the primary techniques utilised in this thesis is force spectroscopy, which has been used to interrogate protein-protein interactions on a single molecule scale. Force spectroscopy has been used to mechanically unfold proteins and to dissociate protein complexes[282-284]. The first mechanical unfolding experiments were performed over 20 years ago on the giant muscle protein titin using atomic force microscopy (AFM)[285]. Since then, our understanding of protein and protein complex mechanical stability has improved by using multiple complementary methods.

Three techniques that have been extensively used to study the effects of force on proteins are AFM, optical/magnetic tweezers and patch clamp, the former being extensively used in this thesis. The properties of these techniques are highlighted in Table 1-3.

	AFM	**Optical Tweezers**	**Patch Clamp**
Temporal resolution (ms)	1	0.1	0.001
Spatial resolution (nm)	0.1 (vertical)	0.1	50
Force Range (pN)	1-10,000	0.1-100	n/a
Applications	Unfolding/ dissociation	(Un)folding/ (un)binding	Single/ multiple ion channels
limitations	High level of noise	Narrow force range, photodamage and local heating from laser	Membrane proteins only

Table 1-3. Comparison of key parameters, features and limitations of AFM, optical tweezers and patch clamp techniques. Taken from Chen *et al* 2015[286].

1.6.1 Atomic force microscopy (single molecule force spectroscopy)

AFM-based single molecule force spectroscopy (SMFS) has been utilised extensively as a technique to measure the intramolecular unfolding forces of individual proteins[287], and the unbinding of non-covalent molecular interactions[282,288]. AFM is valuable in these scenarios due to its operational characteristics: high force sensitivity (theoretically 2-10 pN)[289], high dynamic range (0.001–5000 nN)[290] and high positional accuracy (0.01 nm), combined with the capability to operate in a wide range of buffers, including physiological conditions. The AFM has a unique

advantage of possessing high spatial resolution in order to detect molecular interactions at high lateral resolutions, in combination with possessing great force sensitivity and the wide dynamic range required for the mechanical study of proteins[282]. In this way, AFM can measure biologically relevant forces and distances. Recent studies however, have suggested that with technological advances, sub-pN forces with greater precision may be achievable in the near future[291,292].

A typical AFM instrument uses a cantilever which comprises a very sharp stylus on a flexible lever. Typically the cantilever is thin (10 μM) and composed of gold coated silicon nitride; it is used as a force-sensitive probe and is attached to an AFM optical head. A photo-diode laser is positioned onto the reflective cantilever which detects changes in the reflection signal. This is used as an accurate method of monitoring the vertical deflection of the lever, allowing the measurement of cantilever deflections in the pico-nanometer range (Figure 1-17 A). A known spring constant of the cantilever allows the force to be calculated from the deflection of the cantilever according to Hooke's law (Equation 2-3, Section 2.2.4.2). Spring constants of around 10-40 pN/nm are ideal in SMFS of protein molecules as the stiffness of the spring constant affects sensitivity (typically, cantilevers with spring constants of 30 pN/nm were used in experiments in this thesis). A typical experiment (Figure 1-17) consists of many approach retract cycles in which the AFM probe approaches and then retracts from a surface at a constant velocity. The AFM probe is positioned on the Z-axis by piezo expansion, controlling the approach retract cycles. The cantilever is pushed hard against the surface until a defined deflection threshold is reached. The cantilever is then withdrawn from the surface at a constant velocity which defines the amount of force loaded onto a molecule or complex.

In a typical SMFS experiment to analyse protein-protein or protein-ligand interactions, the proteins and/ or substrates are covalently immobilised to the AFM probe or surface by the use of polyethylene glycol (PEG) linkers derived with a functional group at one end in order to specifically immobilise the molecule of interest. Initial AFM protein recognition studies did not use linkers[289], but this simple approach failed when investigating antigen-epitope interactions[293] due to the lack of molecular mobility and to unspecific tip-probe adhesion forces which obscured the specific interactions of interest. The use of flexible linkers was developed to address some of these issues[294]. The use of PEG linkers is advantageous in SMFS experiments as it helps to reduce non-specific protein sticking, enables substrate flexibility and has a signature resisting force when stretched[295]. The SMFS procedure involves approaching a tip functionalised with proteins to a similarly functionalised surface so that intermolecular

complexes are formed (Figure 1-17 B and C). Upon retraction of the AFM probe if an interaction has formed, an entropic restoring force will be observed, increasing in a parabolic fashion until the interaction breaks and the force sharply falls to zero (Figure 1-17 B and C). The force required for dissociation, termed the rupture force (F_R), is the force at the apex of the parabolic curve (Figure 1-17 B iv and C iv).

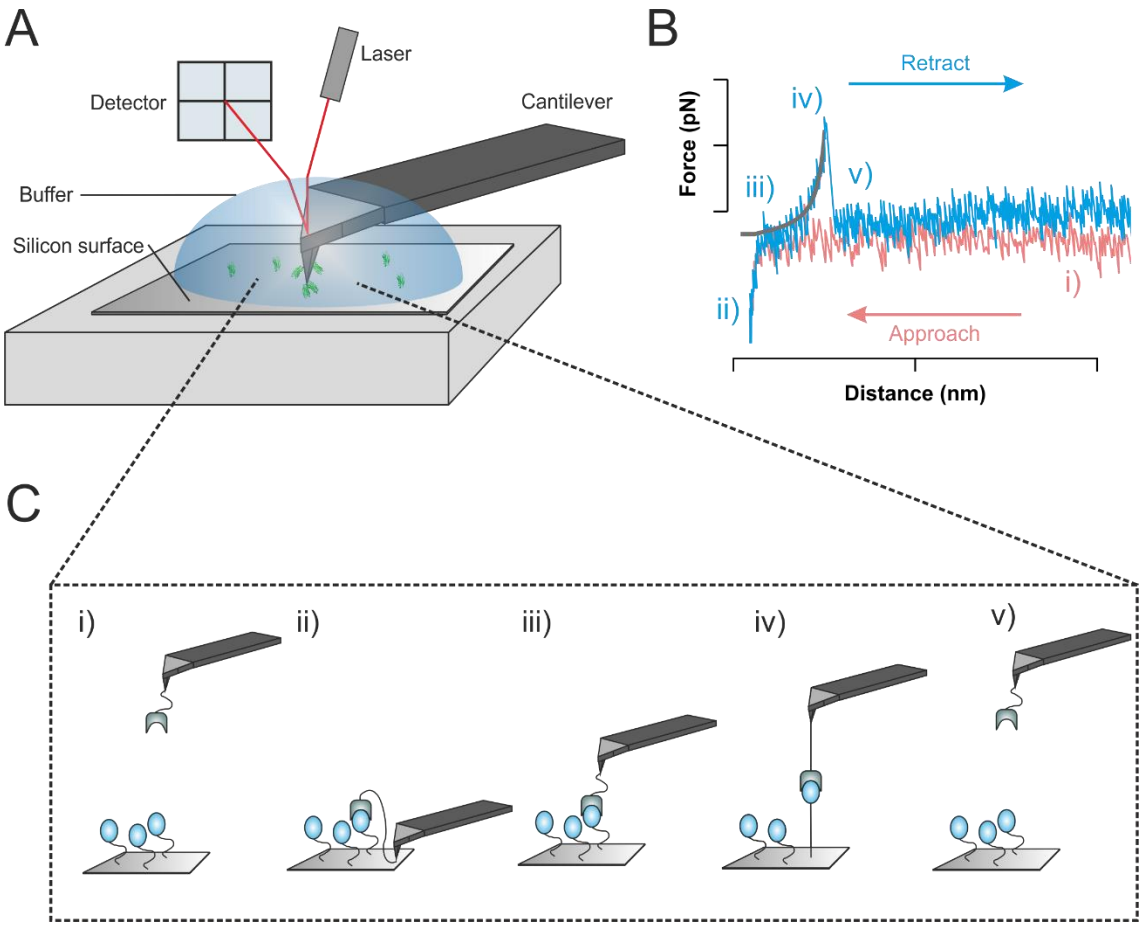


Figure 1-17. Schematic of a typical protein-ligand interaction in a SMFS experiment. (A) Typical set up of an AFM. A cantilever probe and surface are functionalised with proteins via flexible PEG linkers. A laser is positioned on the tip of the reflective probe which detects deformations caused by a resisting force. (B and C) Step by step outline of an approach-retract cycle in which a protein-ligand interaction is formed and then dissociated under force, plus a corresponding force-extension trace. i) The cantilever is initially situated away from the surface, ii) the piezo control approaches the cantilever to the surface until a defined deflection is reached and the cantilever is pressed into the surface, bringing interaction partners together, iii) an associated complex is stretched, iv) creating a force which bends the cantilever v) before the eventual dissociation of a complex and the return of the cantilever to its original position. This process causes a parabolic force increase in the force-extension profile (B) which can be fitted to a worm-like chain (WLC) model (grey line) in order to extract the parameters of rupture force (F_R) and contour length (L_C).

SMFS has been established as a powerful technique to study intermolecular interactions as it allows the estimation of unbinding forces (Table 1-4), dissociation rate constants, the nature of force resistant bonds and the length of the complex before dissociation. The first studies that demonstrated the capability of AFM to measure discrete and biologically-specific rupture forces of molecular complexes were carried out on avidin and biotin in 1994[282]. Since then an array of molecular complex interactions have been analysed, examples of which are highlighted in Table 1-4.

Molecular partners	**Pulling velocity (μms^{-1})**	**Unbinding force (pN)**	**Reference**
Avidin/biotin	5	173 ± 19	Lo *et al* 1999[296]
Strepavidin/biotin	5	326 ± 19	Lo *et al* 1999[296]
Human serum albumin (HSA)/anti-HSA	0.2	244 ± 22	Hinterdorfer *et al* 1996[294]
Actin/myosin	0.0334	14.8 ± 4 and 24.7 ± 1.4	Nakajima *et al* 1997[297]
Nitrilotriacetate (NAT)/histidine 6 (His6)	0.09–0.27	150–194	Kienberger *et al* 2000[298]
Intercellular adhesion molecule-1 (ICAM-1)/anti-ICAM-1 Ab	4.68	100 ± 50	Willemsen et al 1998[299]
Ganglioside G_{M1}/cholera toxin B-oligomer (ctB)	0.04–4	54 ± 46–62 ± 30	Cai and Yang 2003[300]

Table 1-4. Examples of protein ligand unbinding forces. Values taken from Lee et al 2007[282]

1.6.2 Worm-like chain model and contour length analysis

The force-extension data obtained from SMFS experiments in this thesis were analysed using a model for polymer elasticity termed the worm-like chain (WLC) model (Equation 1-1). The WLC describes a protein as a continuous flexible string and neglects any discrete structure in the polypeptide chain. This model of extension of a polypeptide chain (x), was empirically derived from the stretching of DNA[301,302], and predicts the entropic restoring force (F):

$$F(x) = \frac{k_B T}{p}\left(0.25\left(1 - \frac{x}{L_C}\right)\right)^2 - 0.25 + \frac{x}{L_C} \qquad 1\text{-}1$$

where L_C is contour length (the fully extended length of the protein or protein complex), T is temperature (in Kelvin), K_B is Boltzmann's constant and p is the persistence length (a measure of the flexibility of a polymer, the point at which a polymer ceases to be treated elastically and can thus be treated as a static chain. In these studies this is fixed as the length between two α-carbon atoms in a polypeptide chain (0.4nm)[303]).

The WLC model has been used extensively to delineate protein domain unfolding[285,304] and protein complex dissociation pathways[305,306], as the force extension profiles are well defined by the model. A valuable piece of information gained from the WLC model is the contour length (L_C). Firstly, the L_C value is important in identifying whether an interaction of two binding partners is specific, as the expected L_C can be estimated from structural information. Changes in L_C can report on structural changes in the protein/protein complex such as force-induced remodelling of protein domains[307] or report on different interactions between proteins. These scenarios are highlighted in Figure 1-18. Indeed both of these situations could exist concurrently.

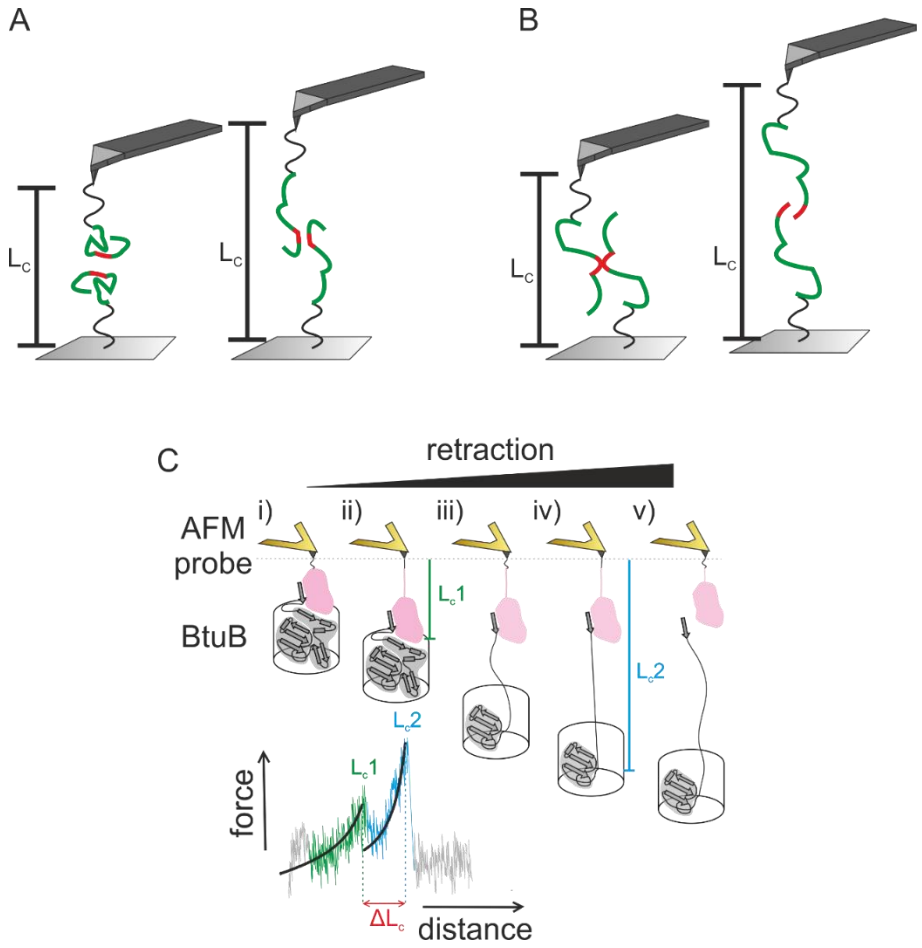


Figure 1-18. Properties of proteins which L_C may report on. (A) Different contour length values may result from a change in conformation of a protein with the same interaction region (red). In the example shown in (A), protein partners are either collapsed or expanded which results in different L_C values. (B) A different scenario in which the the protein partners interact via different interaction regions (red) which would result in different L_C values. (C) An experimental example taken from Hickman *et al* 2017[307] in which Gram negative TonB dependent transporters (BtuB in this example) are remodel under force as a method of gated transport in order to transport scarce nutrients across the outer membrane. In this example a double L_C profile was observed, the difference between these peaks (ΔL_C) corresponds to the amount of protein that unfolds under force.

From further calculations based on the WLC model (Equation 1-1), the instantaneous loading rate (the rate in which force is applied to a complex), which is strongly influenced by the L_C and not just the F_R, can be calculated. To do this, a differential WLC model (Equation 2-4) can be used to calculate the slope (WLC_{slope}) of the parabolic WLC fit which is the change in force as a function of distance (ie. the stiffness of the complex). When multiplied by the known retraction velocity, the loading rate (r_f) for a specific single event can be calculated. It is important to calculate the instantaneous loading rate as these events are kinetic processes and so are time dependent.

1.6.3 Dynamic force spectroscopy (DFS)

When force is applied to a protein or protein complex, the folded or bound states are destabilised. The rate at which force is loaded on a protein affects the apparent strength of a protein or protein-protein interaction since DFS is a kinetic measurement. In general, with higher velocities, more force is required to unfold/break a complex. This was demonstrated by Evans and Ritchie[308,309], based on Bell's model for off-rates[310], it was shown that externally applied force lowers the energy barrier between low and high energy states (Figure 1-19). A linear dependence is usually observed when the most probably rupture force (F_R) is plot as a function of ln(r_f) (Figure 1-19 and Equation 1-2 (Bell-Evens model)).

$$F_R = \left(\frac{k_B T}{x_u}\right) ln \left(\frac{r_f \, x_u}{k_{off}^{0F} k_B T}\right) \qquad 1\text{-}2$$

where k_B is Bolzmann's constant, T is temperature (in kelvin), r_f is the rate at which force is loaded onto the complex or loading rate, x_u is the distance along the reaction coordinate from the low energy state to the transition state and k_{off}^{0F} is the spontaneous unfolding or unbinding rate in the absence of force.

The effect of force on a protein or a protein complex can be depicted in a free energy landscape in which a low energy state representing the native or bound protein, is separated from a higher energy, unbound or unfolded state by an energy barrier which has to be overcome in order for a protein unfold or a complex to unbind, which becomes more likely under force. Force acts to tilt the energy landscape by a factor of Fx_u thereby reducing the size of the energy barrier to unfolding or unbinding. In this way, energy barriers that may have been hidden at zero force may become rate limiting at certain velocities, giving two linear correlations between (F*) and ln(r_f)[305] (see below).

Extending a protein or protein complex at various velocities allows the acquisition of a dynamic force spectrum to be acquired. This allows for the quantification of parameters that are used to describe the free-energy landscape such as x_u and the k_{off}^{0F} (Figure 1-19).

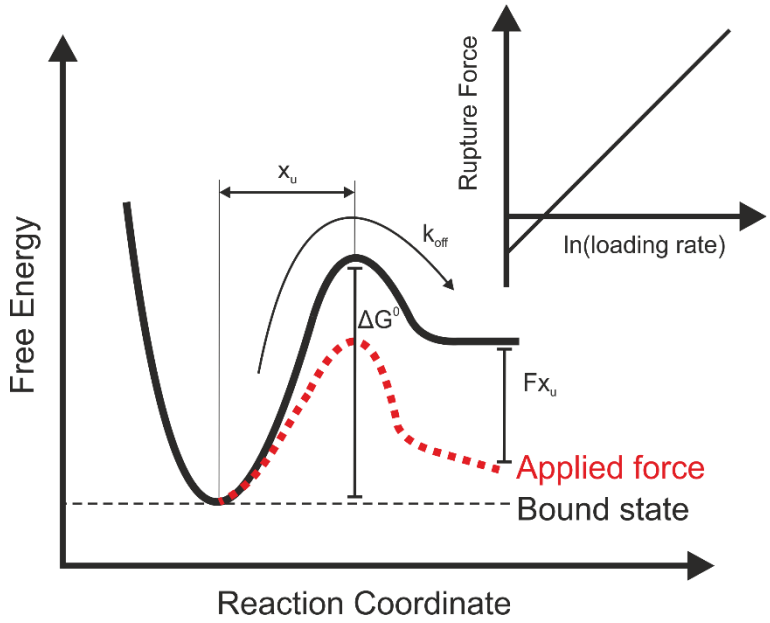


Figure 1-19. Free energy landscape and a dynamic force spectrum. A typical energy landscape of a two-state system that could be reporting on either protein unfolding or the dissociation of a protein complex (black trace). A single energy barrier with a value of ΔG separates a lower energy folded or bound state from a higher energy unfolded/ unbound state. The distance between the ground state and the energy barrier peak or transition state is x_u along the reaction coordinate. The energy barrier is spontaneously crossed at a specific rate (k_{off}). Application of an external force (red dashed trace) tilts the energy landscape a function of Fx_u, which lowers the energy barrier, making spontaneous unbinding or unfolding more likely. A typical dynamic force spectrum for a two-state system is shown inset. The spectrum is governed by a singular linear regime with a slope proportional to $1/x_u$. The y-intercept of the plot allows extrapolation of the k_{off} rate at zero force (k_{off}^{0F}).

1.6.4 Studies of aggregation with force spectroscopy

Force spectroscopy has been utilised previously in the study of aggregation-prone or amyloidogenic proteins[311-323]. Studying protein aggregation is a significant challenge, owing to the fact that the aggregating species are heterogeneous, only transiently and lowly populated, and the fact that aggregation-prone proteins progress on an exponential timescale towards higher order, end point species. Force spectroscopy offers an ideal technique to study aggregating systems as experiments can be carried out at low concentrations and single molecule events can be isolated without the contributions of heterogeneous, higher order species that often occur in ensemble methods. Both the intramolecular conformation and the

intermolecular interactions between aggregation-prone molecules have been studied using various force spectroscopy techniques, as discussed below.

1.6.4.1 Force studies on intramolecular interactions

An AFM SMFS study by Carrion-Vazquez and colleagues[311] used a system in which aggregation-prone proteins were engineered into a polyprotein construct in order to interrogate the intramolecular conformations of the amyloidogenic IDP molecules (Figure 1-20 A). This was achieved by engineering the protein of interest into a carrier protein that is part of the polyprotein relay in what was termed a "carrier-guest" technique. In this way, force events corresponding to the molecule of interest only occur after the carrier molecule has unfolded and so the noisy region of the force curve proximal to the surface is avoided. This study investigated the mechanostability of several amyloidogenic systems of poly-glutamine proteins with different poly-glutamine (poly-Q) expansions (Q_{19}, Q_{35} and Q_{62}) (Figure 1-20); $A\beta_{42}$ along with fibrillation incompetent (F19S/L34P[324]) and a familial AD mutants (E22G); Sup35 and αSyn along with familial PD mutants A53T and A30P. The study showed that mechanostable force events >20pN (at 400 nm/s) could be attributed to the aggregation-prone proteins as shown in the force-extension profiles in Figure 1-20 B and C, this was postulated to arise from the acquisition of β-structure which has previously been proposed to be concomitant with toxic gain of function[104]. The proportions of mechanostable force events also correlated well with aggregation-promoting mutant proteins of αSyn, Aβ42 and polyQ proteins, supporting this hypothesis. Moreover, a fibrillation incompetent Aβ42 (F19S/L34P[324]) and non-toxic short stretches of poly-glutamate (Q19) (Figure 1-20 C) showed an absence of mechanostable events in a similar manner to a non-amyloidogenic IDP control: VAMP2 (Figure 1-20 C).

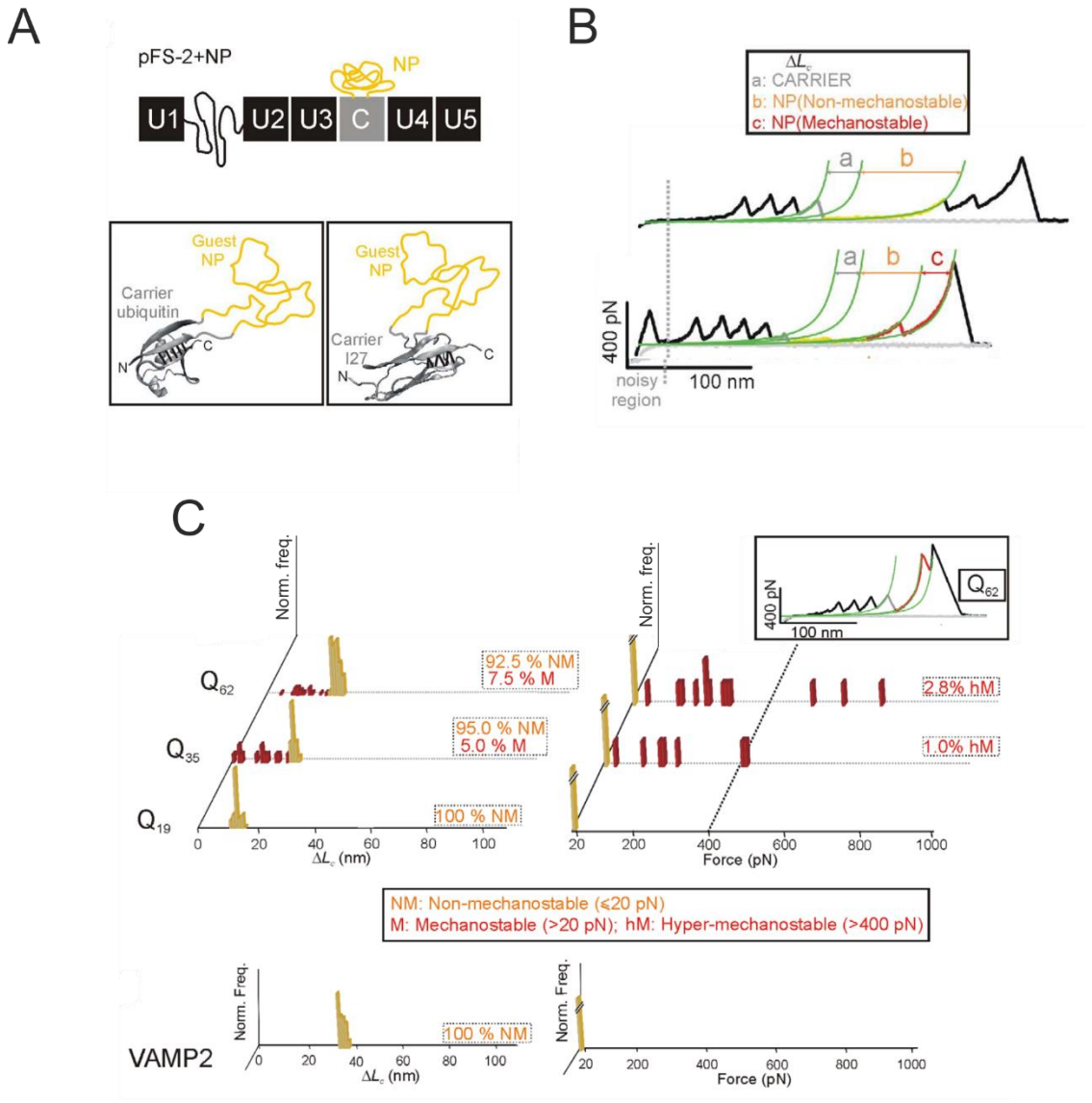


Figure 1-20. SMFS approach and experimental data from Hervas *et al* 2012[311] in which the conformational polymorphism of amyloidogenic IDPs were analysed in a single molecule approach. A) Top: schematic representation of the polyprotein system used in this study in which a guest NP (neurotoxic protein, in orange) is mechanically protected by a carrier protein (C in grey) in a polyprotein scaffold flanked by ubiquitin repeats (U in black). The carrier protein consists of either a monomeric ubiquitin or an I27 domain (A, bottom). B) Representative force-extension profiles from the system highlighted in A with Sup35 as the NP. Different conformations of the NP can be identified using this technique such as non-mechanostable events (b, orange) which occur after the unfolding of the carrier domain (a, grey) in which the ΔL_C corresponds to the full length of the NP without additional force peaks; and mechanostable events (c, red, bottom trace) in which an additional force peak can be attributed to the NP. C) Experimental data using the approach highlighted in A to interrogate polyQ tracts of various lengths. The non-disease, "sub-threshold" Q_{19} tract contains only non-mechanostable (M) events. The familial disease tracts (Q_{35} and Q_{62}) include mechanostable events which increase in proportion with increasing tract length. These mutants also contain hyper-mechanostable (hM) events of more than 400 pN which increase in proportion with increasing tract length. C) bottom histograms represent a non-aggregation control IDP, VAMP2, showing no M events. Adapted from Hervas *et al* 2012[311].

Similar studies by Samori and colleagues[312,313] reached similar conclusions. These studies also interrogated the intramolecular interactions of αSyn in a polyprotein context (Figure 1-21 A). Consisting of three tandem repeat domains of I27 flanking αSyn on either side. A conformational ensemble of αSyn was inferred from these studies. Random-coil conformational states were attributed to unfolding force events with no resolvable unfolding transition attributable to αSyn, while large force events were attributed to β-like conformations. These conformations with higher mechanostability increased in population when familial PD mutated αSyn variants were analysed (A30P, A53T and E46K) and also under aggregation-prone conditions such as lower pH and high ionic strength.

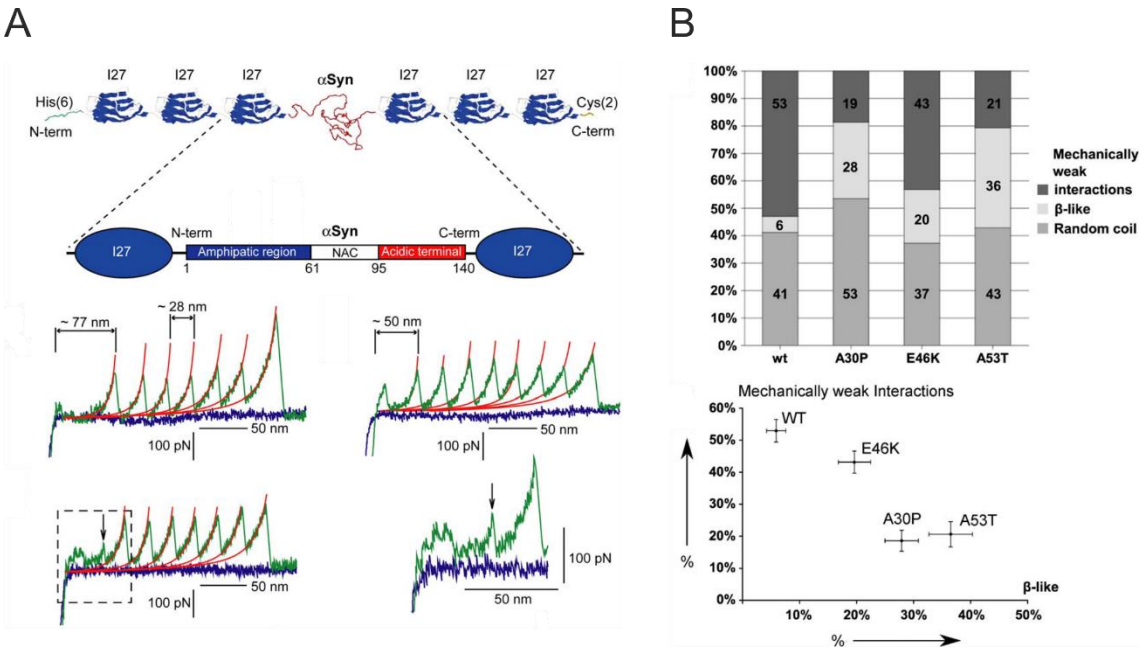


Figure 1-21. SMFS approach and experimental data from Sandal *et al* 2008[312] and Brucale *et al* 2009[313] in which the conformational polymorphism of αSyn and PD linked mutants (A30P, E46K, and A53T) were analysed in a SMFS approach. A) Schematic representation of the polyprotein construct used to interrogate the conformation of αSyn in these studies. αSyn is inserted into the polyprotein scaffold, flanked by three I27 domains at either side. Only force-extension data with at least 6 force events were binned for analysis in this study to allow for unequivocal stretching of the αSyn chain. Representative force-extension curves are shown below in A). Three types of events are identified in these studies, random coil interactions as shown in the top left force-extension trace characterised by a featureless region before the stretching of six tandem I27 domains. An example curve fitting a β-like signature is shown on the top right force-extension trace in which seven force events are observed, one of which corresponding to αSyn unfolding. Examples of mechanically weak conformations are shown in the bottom traces in which single or multiple small peaks (arrows) corresponding to the αSyn unfolding that precede the six saw-tooth-like force events from I27 domains. B) The relative proportions of these events of each of four αSyn variants (WT, A30P, E46K, A53T). PD familial mutants show a higher percentage of β-like events and a lower percentage of mechanically week events. Part A) adapted from Sandal *et al* 2008[312], part B) adapted from Brucale *et al* 2009[313].

Mechanostability of aggregation-prone molecules was also demonstrated in another force study[314], in which poly-Q tracts of different lengths were investigated. This study used a single molecule force clamp approach in order to pull a single chimeric polyprotein carrier with different length poly-Q tracts similar to the protocols implemented above, showing that poly-Q stretches formed collapsed, but a heterogeneous ensemble of mechanically stable structures.

All of these studies show evidence of highly mechanostable intramolecular interactions in these aggregation-prone IDPs. It was hypothesised that this highly mechanically stable ensemble of conformations, which would presumably be kinetically trapped, could play a part in mechanistically 'jamming' protein processing machinery *in vivo*, unbalancing cell proteostasis, leading to toxicity[311]. This was postulated because of the existence of AAA+ ATPases from degradative machines have been shown to unfold their substrate mechanically using relatively low forces[325,326]. The forces required to mechanically unfold some of the misfolded intramolecular interactions revealed by force studies would thus not be accessible by the cell processing machinery.

1.6.4.2 Force studies on intermolecular interactions

AFM-based SMFS studies have also been used to investigate the intermolecular interactions of aggregation-prone proteins on a single molecule scale similar to the approach used in this thesis. In this way, the initial interaction event in the aggregation cascade: that of dimerisation, can be interrogated. The Lyubchenko group is one of the few groups to have studied the interactions of amyloidogenic proteins in this way. Their force studies to date are summarised in Table 1-5.

Lybchencho and colleagues have studied various proteins using SMFS, with most of the work being carried out on Aβ and αSyn. The first study of the strength of inter-protein interactions was carried out by the group in 2005 and focused on αSyn, Aβ40 and lysozyme[315]. The proteins in this study were immobilised covalently to amino-functionalised mica and to similarly functionalised Si_3N_4 cantilever using glutaraldyhyde crosslinking. A problem with this method is that glutaraldehyde reactions take place at all amide moieties in addition to the N-terminus of the protein. This immobilisation method would be difficult to use to discern single molecule interactions due to the heterogeneity of the immobilisation.

The same group developed a different experimental method that addressed this problem, originally developed for Aβ40[321] and subsequently used for other proteins (outlined in Table 1-5). This method added an engineered cysteine residue at the N-terminus of Aβ40 allowing

the use of thiol-specific chemistry for protein immobilisation. The studies using this method to study αSyn used an engineered cysteine at the C-terminus as opposed to the N-terminus. The mica substrate and the cantilever tip were functionalised with maleimide-PEG-silantrane (MAS) and so were capable of reacting with the unique thiol group on the SH-terminated peptide.

Table 1-5. Studies looking at the intermolecular interactions of amyloidogenic proteins

Protein	Method of immobilisation	Conditions examined	Pulling rate	Source	Points of note
αSyn	Glutaraldehyde cross-linking	pH 9.8 - 1.0	Not stated	McAllister et al 2005[315]	First study into the strength of inter-protein interactions.
	C-terminal cysteine	pH 6.0 - 2.7	DFS pulling range: 100pN/s – 100,000pN/s	Yu et al 2008[316] and 2009[317]	Dimer lifetime in the range of seconds compared with ns range of monomer. Lifetimes increase with decreasing pH
	C-terminal cysteine	Zinc and aluminium cations, dopamine and ionic strength	Not stated	Yu et al 2011[318]	No effect of dopamine and ionic strength. Cations increase rupture events and decrease contour lengths- cause interactions closer to the C-terminus
αSyn and A30P	C-terminal cysteine	polyamide spermidine	600nm/s	Krasnoslobodtsev et al 2012[319]	Zero yield of rupture events in absence of spremidine at physiological pH. Increased to ~2% in presence of spermidine. A30P showed more interactions closer to the C-terminus than wild-type
αSyn and A30P, E46K, A53T	C-terminal cysteine	Famillial point mutation A30P, E46K and A53T	600nm/s	Krasnoslobodtsev et al 2013[320]	Mutations decreased L_C, increased interactions closer to the C-terminus, decreased interactions closer to the N-terminus compared with wild-type
Aβ40	Glutaraldehyde cross-linking	pH 9.8 - 1.0	Not stated	McAllister et al 2005[315]	First study into the strength of inter-protein interactions.
	C-terminal cysteine	pH 9.8- 5.0	Not stated	Krasnoslobodtsev et al 2005[321]	More reproducible and reliable immobilisation of protein technique produced
	C-terminal cysteine	pH 2.7 – 7.0	DFS extension rates: 1000pN/s-200,000pN/s	Kim et al 2011[322]	DFS analysis showed two discrete dissociation rates for force experiments below pH 7.0. Only one rate at pH 7.0. increased lifetimes of dimers at lower pH
Aβ40, Aβ42,	C-terminal	Mutations Aβ40 VPV*	5000- 7000pN/s	Lv et al 2013[327]	Aβ40 VPV Mutations decrease interactions closer to the

VPV mutant, *p*P mutant	cysteine	and Aβ40 *p*P**			N-terminal and increase interactions closer to the C-terminal. Aβ42 pP mutations do the opposite and revert the L_C profile back to an Aβ40 like profile.
Lysozyme	Glutaraldehyde cross-linking	pH 1.5 – 9.8	Not stated	McAllister et al 2005[315]	Force increases with lowering pH. Decreases dramatically below pH 4.0
Sup35 peptide	N-terminal cysteine	pH 2.0, 3.7, 5.6 in PepP***, PepQ****	102pN/s – 105,000pN/s	Portillo et al 2012 [323]	DFS analysis showed that PepQ formed low and high force interactions and ThT analysis showed that it formed fibrils. PepP didn't form fibrils, didn't have steep high force slope (inner barrier) suggesting the inner barrier characterises the formation of aggregates.

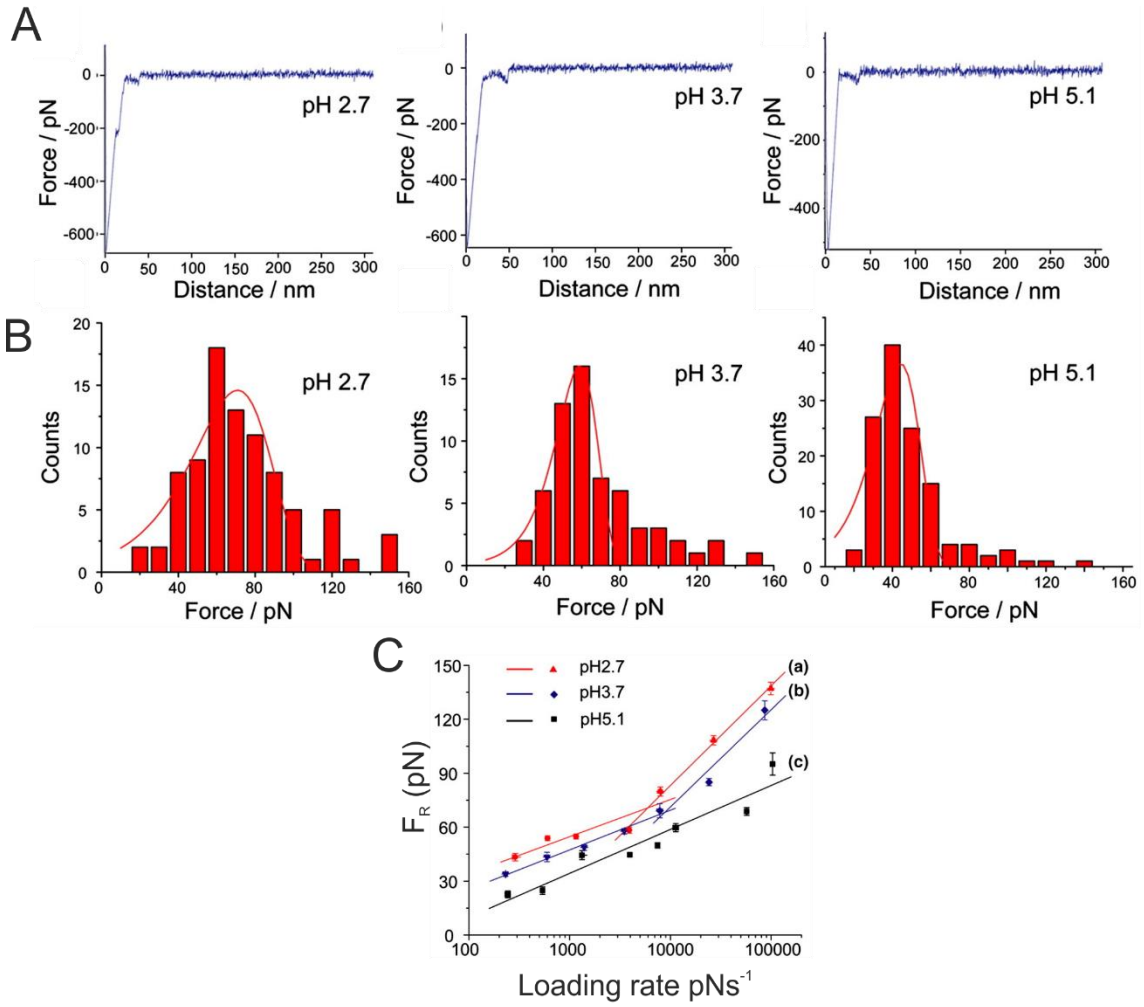
* VPV: Gly33Val-Val36Pro-Gly38 Val (mutations in both Aβ40 and Aβ42, reduces the flexibility of the C-terminus, stabilises proposed β-turn in C-terminus[328])

** *p*P: Val36/D-Pro and Gly37/L-Pro (mutations in both Aβ40 and Aβ42, destabilises proposed β-turn in C-terminus[328])

*** PepP: CGNNPQNY (N-terminal region of Sup35 with engineered Cysteine and a Q10P mutation)

**** PepQ: CGNNQQNY (N-terminal region of Sup35 with engineered Cysteine

Several studies to investigate single molecule dimerisation have been were carried out in this way in order to interrogate the initial intermolecular interactions of αSyn as highlighted in Table 1-5 and in Figure 1-22 andFigure 1-23. Various factors were investigated in the studies including pH (Figure 1-22), ionic strength, metal cations (Figure 1-23), the presence of dopamine, the presence of the polyamide spermidine (Figure 1-23), plus, the effects of pathological mutations (A30P, E46K and A53T) (Figure 1-23) on the self-association of α-Syn. It was shown that decreasing pH increases the rupture forces[315-317] from 43.1 pN at pH 5.1 to 71 pN at pH 2.7[316]. DFS analysis of αSyn allowed derivation of energy landscapes that showed dimer lifetimes in the range of seconds[316] (4 s for αSyn dimerisation at pH 2.7, Figure 1-22 C) compared with that of nanoseconds for monomer (predicted by molecular dynamics simulations[329]). This is an important finding as aggregation-prone dimeric species with longer lifetimes increase the probability of their further involvement in interactions that may be on pathway in the aggregation cascade. This analysis also revealed that reducing the pH increases the dimeric lifetime of the αSyn dimeric interaction from 0.27 s at pH 5.1 to 4 s at pH 2.7[316] (Figure 1-22). Moreover, DFS analysis revealed two linear regimes for αSyn in a Bell-Evens analyses[316] (Figure 1-22) at pH 2.7 and pH 3.7 but not at pH 5.1, thus suggesting that the dimerisation interaction of αSyn at lower pH values is governed by two barriers to dissociation whereas at higher values, there is only one barrier suggesting that the dimeric species can have different stabilities at different pH values. Given that pH has been showed to play an important role in the *in vitro* aggregation of αSyn, it may be logical to propose that the different properties of dimers at different pH values play an important role[330]. The finding that pH plays an important role in the affinity of the dimerisation interaction has implication for the disease process *in vivo* as αSyn comes into contact with various cellular compartments[331,332].



pH	X_{u1}(Å)	k_{off1} (s^{-1})	$\tau_0(1)$ (s)	X_{u2} (Å)	k_{off2} (s^{-1})	$\tau_0(2)$ (s)
2.7	4.59 ± 1.25	0.25 ± 0.38	4.00	1.71 ± 0.07	13.2 ± 1.8	0.08
3.7	4.29 ± 0.31	0.74 ± 0.24	1.35	1.75 ± 0.37	20.5 ± 13.6	0.05
5.1	3.88 ± 0.48	3.74 ± 1.99	0.27	–	–	–

Figure 1-22. DFS SMFS analysis of the αSyn dimerisation interaction as a function of pH. A) Representative force-extension plots of αSyn dimerisation at different pH values (2.7, 3.7 and 5.1 from left to right). B) Example force distributions at the same pH values as in A. C) DFS analysis of αSyn as a function of pH, (a) red triangles, pH 2.7; (b) blue diamonds, pH 3.7; and (c) black squares, pH 5.1. Error bars show the standard error. Bell-Evens parameters are denoted in the table at the bottom of the figure. Dimerisation interactions at both pH 2.7 and pH 3.7 show two linear regimes indicating an outer and an inner barrier to dissociation is present in the energy landscape of dissociation. X_{u1} denotes the outer dissociation barrier (lower gradient linear regime), K_{off1} indicates the off rate at zero force of the same linear regime and $\tau_0(1)$ indicates the lifetime of this interaction. Similarly, X_{u2}, K_{off2} and $\tau_0(2)$ correspond to the inner dissociation barrier. Adapted from Yu *et al* 2008[316].

Familial PD mutants (A30P, E46K and A53T) of αSyn were investigated by this group and showed differences in the dimerisation interactions relative to WT at pH 3.7 (Figure 1-23 A).

The observed L_C values decreased for PD mutants which indicated that there was an increase in interactions closer to the C-terminus in PD mutant dimers compared with WT[319,320].

Most of the SMFS studies carried out on αSyn by the Lyubchenko group were conducted under acidic pH as events at neutral pH were not observed. The presence of both metal cations (Al^{3+} and Zn^{2+}) and the polyamide spermidine, however, induce the formation of dimers at neutral pH (Figure 1-23 B and C). Cations increased the yield of dimerisation interactions form 0.8 % at at neutral pH to 3.9 and 7.0 % for Zn^{2+} and Al^{3+} respectively[318]. The polyamide spermidine at physiological pH increases the frequency of interactions from 0 to ~2%[319]. This data shows that external factors can promote the proposed aggregation-prone dimerisation interaction of αSyn and be important in promoting conformations that may promote higher order aggregate formation and be involved in the implication of disease.

As mentioned above, most of the dimerisation events analysed for αSyn were analysed at acidic pH. However, studying these interactions at neutral pH may be more applicable in regards to the relationship between dimerisation formation and disease. The hit rate for αSyn dimerisation at pH 7.0 was reported at 0.8 % by this group[318] and not further analysed. The immobilisation concentration is relatively low (19 nM) in these studies. It is likely, therefore, that interactions of αSyn are missed and not analysed in these studies at neutral pH. Furthermore, there are other weaknesses with the approach used in these studies. Firstly the force-extension traces show a large amount of noise in the regions proximal to the surface (Figure 1-22 A and Figure 1-23 B and C). It is likely therefore that interactions that may occur at lower L_C values are masked by noise in these experiments and therefore not able to be analysed. Secondly, L_C distributions may have been over-analysed in these studies (Figure 1-23 A and B). Several distributions were consistently fit to L_C data[319,320] (Figure 1-23 A and B) and the different distributions were interpreted as different dimeric conformations.

A final caveat with the data obtained in these studies involves the interpretation of the L_C values. The L_C data in these studies were taken as the length of linkers and protein extended in an interaction. However, this is only the case in an idealistic scenario in which a protein is immobilised onto the cantilever tip apex and the protein partner is situated directly below it on the substrate. In reality, this scenario is rare and would be the maximum L_C value observed. The observed modal L_C is therefore lower than the true length of the protein and linkers. This is due to the important effects of linker length and AFM tip geometry have on experimental L_C values. This issue has been discussed by this group previously[306].

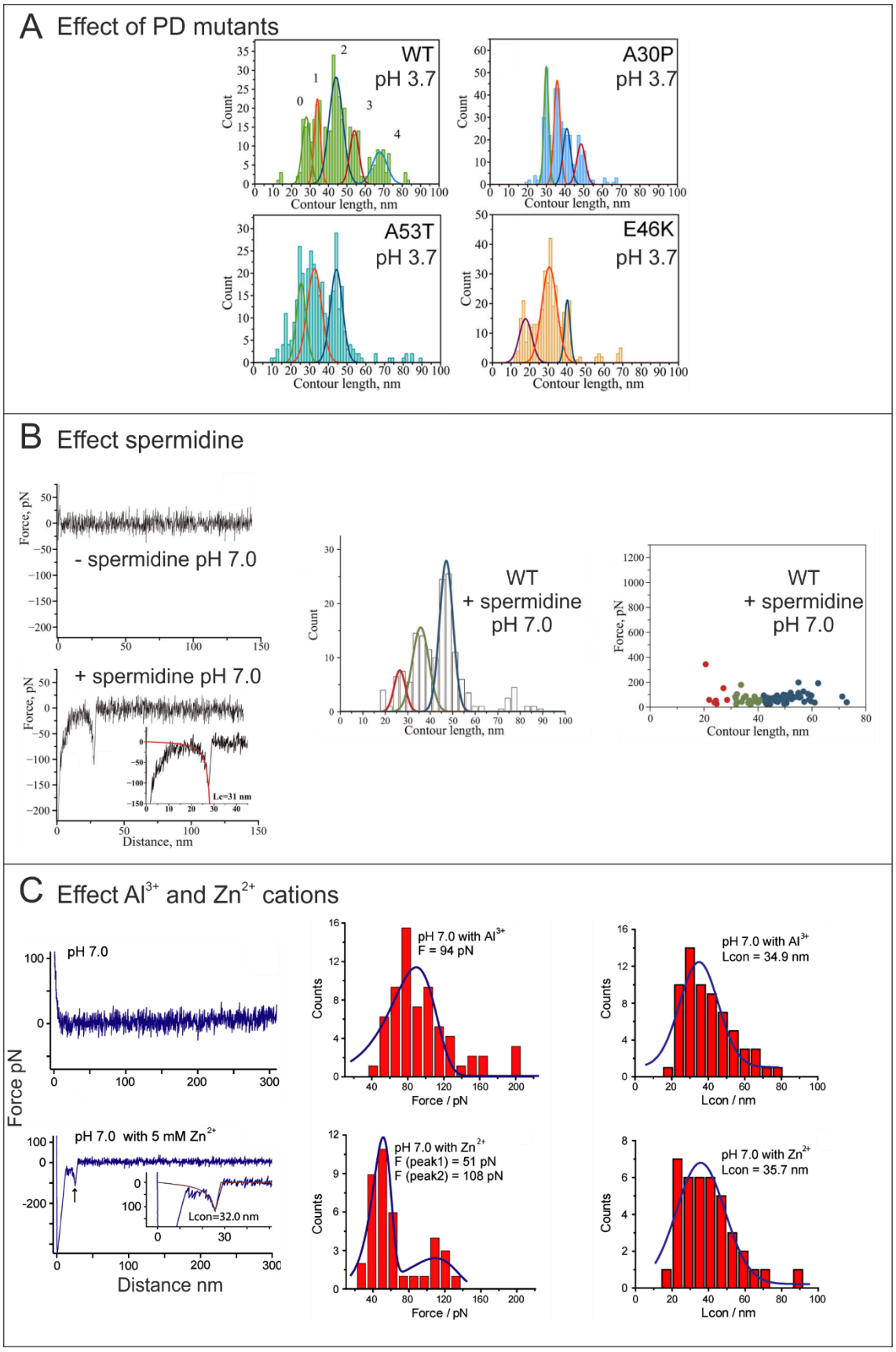

A Effect of PD mutants
WT
pH 3.7
A30P
pH 3.7
A53T
pH 3.7
E46K
pH 3.7
Count
Contour length, nm
B Effect spermidine
- spermidine pH 7.0
+ spermidine pH 7.0
Force, pN
Distance, nm
Lc=31 nm
WT
+ spermidine
pH 7.0
C Effect Al3+ and Zn2+ cations
pH 7.0
pH 7.0 with Al3+
F = 94 pN
pH 7.0 with Al3+
Lcon = 34.9 nm
pH 7.0 with 5 mM Zn2+
Lcon=32.0 nm
pH 7.0 with Zn2+
F (peak1) = 51 pN
F (peak2) = 108 pN
pH 7.0 with Zn2+
Lcon = 35.7 nm
Force pN
Distance nm
Counts
Force / pN
Lcon / nm

Figure 1-23. SMFS data of αSyn dimerisation and the effects of PD mutants (A), the presence of spermidine (B) and the presence of Al^{3+} and Zn^{2+} cations (C). A) L_C distributions of WT αSyn and PD variants (A30P, A53T and E46K) at pH 3.7[320]. The different variants give different L_C distributions relative to WT. Each distribution is fit to several Gaussians which was interpreted as different conformations in the dimeric species of αSyn and the PD mutants. B) SMFS data showing the effect of spermidine on the dimerisation interaction of αSyn[319]. Dimerisation events were absent at pH 7.0, however, in the presence of spermidine under these conditions, force events were able to be observed. L_C was again fit with several Gaussian distributions, taken to represent different dimeric conformations. A L_C vs F_R scatter plot is shown on the right. C) The effect of Zn^{2+} and Al^{3+} on the dimeric interactions of αSyn. As in B, dimerisation events were absent at pH 7.0, but were found to be present with the addition of cations[318].

1.7 Thesis aims

The principal aim of the work carried out in this thesis is to provide molecular insights into the initial interactions of the Parkinson's disease linked protein αSyn. SMFS is used primarily in order to interrogate dimerisation interactions on a single molecule scale.

As discussed in this chapter (Section 1.6.4), SMFS has been used previously to interrogate the intermolecular interactions of αSyn at a single molecule level[315-320]. However there are weaknesses in these studies as discussed above (Section 1.6.4.2). Similar methods with more robust experimental and analytical protocols, were utilised in order to gain further insights into the nature of the dimeric interactions of αSyn, an important event at the start of the aggregation cascade. More knowledge is needed about the properties that drive this interaction, identifying interaction regions and interrogating likely conformations that may occur in the dimeric species, as various SMFS studies have linked highly force resistant conformations to increased aggregation propensity[311-314]. Relating how the driving forces for dimerisation compare in aggregation-prone PD mutants and non-aggregation-prone homologues is important for understanding the role of this interaction in health and disease. The conformation of αSyn as an IDP is dependent on environmental conditions as discussed in this chapter (Section 1.5.2.1), it is therefore important to ascertain how the dimeric interaction is effected by these factors.

The work in this thesis aims to:

1. Gain further knowledge about the conformations observed in the dimeric species of αSyn and understand their context dependence.
2. Identify interaction regions driving dimeric interfaces.
3. Further understand how intrinsic (synuclein variants) and extrinsic (environmental conditions) factors effect αSyn dimerisation, and compare these to the rates of αSyn aggregation in bulk solution.

4. Postulate mechanisms as to how the dimerisation interaction may be related to the aggregation process of αSyn.
5. Develop a SMFS display system that allows robust measurement of intermolecular interactions of highly aggregation-prone-proteins.

Chapter 3 outlines the design and implementation of a protein displaying scaffold which can be utilised in SMFS studies of intermolecular interactions of aggregation-prone proteins. This technique allows the presentation of small, highly aggregation-prone peptides in a flexible loop of a mechanically strong protein. In this way, dimeric interactions in an aggregation-prone system can be analysed without the contributions of noisy region of force-distance data, proximal to the surface.

Chapters 4-6 use SMFS and other biophysical techniques to interrogate the nature of dimeric interactions of full length synuclein variants. Chapter 4 presents SMFS of full length αSyn dimerisation at a single molecule level were validated and shown that robustness of the experimental system.

Results in Chapter 5 indicate force-resistant structure forms in the dimeric species, and that this conformation is dependent on pH. Whereas previous SMFS of αSyn dimerisation used an experimental technique in which αSyn was immobilised at the C-terminus[315-320], several SMFS experiments were carried out here in which αSyn monomers were immobilised at different positions. Carrying out L_C simulations in parallel[306] allowed the localisation of areas of structure and also interaction regions in the dimer.

The experimental section of this thesis concludes with Chapter 6 in which environmental and intrinsic protein properties driving dimerisation are investigated. The data in this chapter allowed the postulation that the dimeric interactions observed in SMFS events may be protective to aggregation and, therefore, have significant implications into the physiological forms of αSyn. These findings may be important to consider when postulating disease modifying agents targeting αSyn.

The thesis ends with a conclusions chapter which draws findings together, correlating these findings with evidence from existing literature and points to future exciting questions.

2 Materials and Methods

2.1 Materials

2.1.1 Centrifuges

Avanti J-26 XP Centrifuge (Backman Coulter, Brea, CA, USA)

Bench top centrifuge: GenFuge 24D (Progen Scientific, London, UK)

Eppendorf 5810R Centrifuge (Fisher Scientific, Loughborough, UK Beckman Coulter)

Contifuge Stratos Continuous-Flow Rotor (Heraeus, Hanau, Germany)

2.1.2 Incubators

Innova 43 Shaker Incubator (New Brunswick Scientific, USA)

Innova 44 Shaker Incubator (New Brunswick Scientific, USA)

SI600 orbital incubator (Stuart, Staffordshire, UK)

SI500 orbital incubator (Stuart, Staffordshire, UK)

2.1.3 Protein purification equipment

ÄKTAprime plus (GE healthcare, Buckinghamshire, UK)

HisTrap FF 5 mL Ni Sepharose (GE healthcare, Little Chalfont, UK)

Superdex™ 75 Hiload 26/60 gel filtration column (GE healthcare, Buckinghamshire, UK)

Superdex™ 75 HR 10/30 gel filtration column (GE healthcare, Buckinghamshire, UK)

2.1.4 Spectrophotometers

NanoDrop 2000 UV-Vis Spectrophotometer (Thermo Scientific, MA, USA)

Ultrospec 2100 pro UV/Visible spectrophotometer (GE Healthcare, Buckinghamshire, UK)

2.1.5 PCR thermocycler

T100 thermal cycler (BioRad, CA, USA)

2.1.6 AFM

MFP-3D™ Stand Alone AFM (Asylum Research, Buckinghamshire, UK)

2.1.7 AFM probes

MLCT silicon nitride with reflective gold AFM probe (Bruker, CA, USA)

2.1.8 Electron microscope

JEOL JEM-1400 transmission electron microscope (JEOL USA Inc., Peabody, USA) with Gatan US1000XP 2k x 2k CCD camera (Gatan Inc., Pleasanton, USA)

2.1.9 Circular Dichroism

ChirascanTM plus CD Spectrometer (Applied Photophysics, U.K.)

2.1.10 Fluorometer

Photon Technology International fluorometer (Ford, West Sussex, UK)

2.1.11 NMR instrument

600 MHz NMR magnet (Oxford Instruments, Abingdon, UK)

QCI-P-cryoprobe and an Avance III HD console (Bruker Corpn., Coventry, UK)

2.1.12 Microplate readers and plates

FLUOstar Omega (BMG Labtech, Ortenburg, Germany)

FLUOstar Optima (BMG Labtech, Ortenburg, Germany)

Corning® 96 Well Half Area Black with Clear Flat Bottom Polystyrene NBS™ Microplate (Corning GmbH, Wiesbaden, Germany)

2.1.13 MS instrument

Synapt HDMS quadrupole-time-of-flight mass spectrometer (Waters Corpn., Manchester, UK), equipped with a Triversa NanoMate (Advion Biosciences, Ithaca, NY, USA) automated nano-ESI interface. The instrument has a travelling-wave IMS device situated between the quadrupole and the time-of-flight analysers.

2.1.14 Gel electrophoresis

Slab Gel Electrophoresis Chamber AE-6200 (ATTO, Tokyo, Japan)

Powerpac 3000 (Bio-Rad Lab., Hercules, CA, USA)

Powerpac Basic (Bio-Rad Lab., Hercules, CA, USA)

2.1.15 Gel ladders and dyes

Precision plus protein dual colour standards protein ladder (BioRad, CA, USA)

Instant Blue Stain (Expedeon Protein Solutions, UK)

2.1.16 Kits

Wizard® Plus SV Minipreps DNA Purification System (Promega Hampshire, UK)

Q5® Site-Directed Mutagenesis Kit (NEB, Hertfordshire, UK)

2.1.17 Buffers

Buffers used in experiments are described in the experimental section, more general buffers used in this thesis are described below.

2.1.17.1 pL purification buffers

Lysis buffer - 20 mM Tris-HCl pH 8.0, 300 mM NaCl, 20 mM imidazole, 2 mM DTT, 0.025 % (w/v) sodium azide, 1 mM PMSF, 2 mM benzamidine, 0.15% (v/v) Triton X100, 20 µg/ml DNase (Sigma Life Sciences, MO, USA), 100 µg/ml lysozyme (Sigma Life Sciences, MO, USA).

Wash buffer - 20 mM Tris-HCl pH 8.0, 300 mM NaCl, 20 mM imidazole, 2 mM DTT, 0.025 % (w/v) sodium azide, 1 mM PMSF, 2 mM benzamidine.

Elution buffer - 20 mM Tris-HCl pH 8.0, 300 mM NaCl, 250 mM imidazole, 2 mM DTT, 0.025 % (w/v) sodium azide, 1 mM PMSF, 2 mM benzamidine.

SEC buffer – 25 mM Tris-HCl, 300 mM NaCl, 2 mM DTT, pH 8.

2.1.17.2 Full leghth synuclein purification buffers

Lysis buffer – 20 mM Tris-HCl, 1 mM EDTA, 5 mM DTT, 1 mM PMSF, 2 mM benzamidine, 100 µg/ml lysozyme, 20 µg/ml DNase, pH 8.0

Wash buffer – 20 mM Tris-HCl, 5 mM DTT, pH 8.0

Elution buffer - 20 mM Tris-HCl, 5 mM DTT, 1 M NaCl, pH 8.0

SEC buffer – 20 mM 20 mM sodium phosphate, pH 7.5

2.1.17.3 SDS-PAGE buffers

SDS-PAGE resolving gel buffer - 1 M Tris-HCl pH 8.45, 0.1 % (w/v) SDS, 15 % (v/v) acrylamide, 13 % (v/v) glycerol, 0.7 % (w/v) APS, 0.07 % (v/v) TEMED.

SDS-PAGE stacking gel buffer - 750 mM Tris-HCl pH 8.45, 0.07 % (w/v) SDS, 4 % (v/v) acrylamide, 0.32 % (w/v) APS, 0.16 % (v/v) TEMED.

SDS-PAGE loading buffer (x2 concentrated stock) - 2 % (w/v) SDS, 10 % (v/v) glycerol, 0.1 % bromophenol blue, 100 mM DTT.

SDS-PAGE cathode buffer (x10 concentrated stock) - 1 M Tris, 1 M Tricine, 1 % (w/v) SDS.

SDS-PAGE anode buffer (x10 concentrated stock)2 M Tris-HCl pH 8.9.

2.1.17.4 TAE buffer (x25 concentrated stock)

1 M Tris, 25 mM EDTA, 2.8 % (v/v) glacial acetic acid.

2.1.18 Peptides

Peptides were purchased from Genscript, NJ, USA at > 99% purity. All peptides were N-terminally acetylated and C-terminally amidated.

2.1.19 DNA primers

All DNA primers were purchased from Eurofins MWG Operon, Ebersberg, Germany.

2.1.20 Plasmids

The pET23a plasmid (expression under control of T7 promotor) encoding pL I60F with an existing protein sequence present (TolBox, TolB binding epitope) in an extended loop region between β-sheets 3 and 4 was kindly provided by Dr David Brockwell (Astbury Centre for Structural Molecular Biology, University of Leeds, UK).

pET23a-pL αSyn_{71-82}: (central NAC region of αSyn, residues 71-82 engineered within an extended loop region between β-sheets 3 and 4, in an expression vector under the control of a T7 promotor) .

pET23a-pL γSyn_{71-82}: (central NAC region of γSyn, residues 71-82 engineered within an extended loop region between β-sheets 3 and 4, in an expression vector under the control of a T7 promotor) .

pET23a-GS: (A non-aggregation control construct in which repeating units of Gly-Ser residues of the same size as the peptide inserts, 12 residues/ 6 repeating units, within an extended loop region between β-sheets 3 and 4, in an expression vector under the control of a T7 promotor).

The pET23a plasmid encoding αSyn (expression under control of T7 promotor) was provided by Prof Jean Baum (Department of Chemistry and Chemical Biology, Rutgers University, NJ, USA).

pET23a-αSyn A140C: (human αSyn with a C-terminal Cys residue for SMFS studies, in an expression vector under the control of a T7 promotor).

pET23a-αSyn A90C: (human αSyn with Cys residue in the NAC region for SMFS studies, in an expression vector under the control of a T7 promotor) kindly provided by Dr Matthew Jackson (Astbury Centre for Structural Molecular Biology, University of Leeds, UK).

pET23a-αSyn A18C: (human αSyn with Cys residue in the N-terminal region for SMFS studies, in an expression vector under the control of a T7 promotor) kindly provided by Dr Matthew Jackson (Astbury Centre for Structural Molecular Biology, University of Leeds, UK).

pET23a-αSyn A140C E46K: (human αSyn familial PD mutant E46K with a C-terminal Cys residue for SMFS studies, in an expression vector under the control of a T7 promotor).

pET23a-βSyn A134C: (human βSyn with a C-terminal Cys residue for SMFS studies, in an expression vector under the control of a T7 promotor).

pET23a-γSyn D127C: (human γSyn with a C-terminal Cys residue for SMFS studies, in an expression vector under the control of a T7 promotor).

2.1.21 Bacterial strains

E. coli XL1-Blue competent cells (efficiency: >1 x 108 cfu/µg) (Agilent technologies, Berkshire, UK).

*rec*A1 *end*A1 *gyr*A96 *thi*-1 *hsd*R17 *sup*E44 *rel*A1 *lac* [F´ *pro*AB *lacI*qZΔM15 *Tn*10 (Tetr)].

E. coli DH5-α competent cells (NEB, Hertfordshire, UK).

F– Φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17 (rK–, mK+) *pho*A *sup*E44 λ– *thi*-1 *gyr*A96 *rel*A1.

E. coli BL21(DE3) competent cells (efficiency: >1 x 106 cfu/µg) (Agilent technologies, Berkshire, UK).

F- dcm ompT hsdS(r_B– m_B–) gal λ(DE3)

2.1.22 Chemicals

A	**Supplier**
Acetic acid, glacial	Fisher Scientific, Loughborough, UK
Acrylamide 30 % (v/v)	Severn Biotech, Kidderminster, UK
Agar	Fisher Scientific, Loughborough, UK
(3-Aminopropyl)triethoxysilane (APTES)	Sigma Life Sciences, MO, USA
Ammonium persulphate (APS)	Sigma Life Sciences, MO, USA
Ammonium Bicarbonate	Sigma Life Sciences, MO, USA
Ammonium sulfate	Acros Organics, Geel, Belgium
B	
Benzamidine dihydrocholride	Sigma Life Sciences, MO, USA
C	
Carbenicillin	Formedium, Norfolk, UK
Chloroform	Fisher Scientific, Loughborough, UK
Calcium chloride ($CaCl_2$)	Melford Laboratories, Suffolk, UK
D	
Dithiothreitol (DTT)	Formedium, Norfolk, UK
Deuterium oxide (D_2O)	Fluorochem, UK
E	
Ethanol	Sigma Life Sciences, MO, USA
Ethylenediaminetetraacetic acid (EDTA)	Fisher Scientific, Loughborough, UK
G	
Glycerol	Fisher Scientific, Loughborough, UK
H	
Hexoflouroisopropanol (HFIP)	Sigma Life Sciences, MO, USA
Hydrogen peroxide (H_2O_2)	Sigma Life Sciences, MO, USA
Hydrochloric acid (HCl)	Fisher Scientific, Loughborough, UK
I	
Imidizole	Sigma Life Sciences, MO, USA
Isopropanol	Fisher Scientific, Loughborough, UK
Isopropyl β-D-1-thiogalactopyranoside (IPTG)	Melford Laboratories, Suffolk, UK
L	
LB broth, granulated	Melford Laboratories, Suffolk, UK
M	

Monosodium phosphate	Fisher Scientific, Loughborough, UK
N	
N-Hydroxysuccinimide-PEG24-maleimide (SM PEG)	Fisher Scientific, Loughborough, UK
N,N-diisopropylethylamine (DIPEA)	Fisher Scientific, Loughborough, UK
P	
Phenylmethanesufonyl fluoride (PMSF)	Sigma Life Sciences, MO, USA
S	
Sodium azide (NaN_3)	Sigma Life Sciences, MO, USA
Sodium chloride (NaCl)	Fisher Scientific, Loughborough, UK
Sulphuric acid	Fisher Scientific, Loughborough, UK
Sodium hydroxide (NaOH)	Fisher Scientific, Loughborough, UK
Sodium dodecyl sulphate (SDS)	Severn Biotech, Kidderminster, UK
Sodium phosphate dibasic, Na_2HPO_4	Thermo Scientific, Surrey, UK
Sodium phosphate monobasic, NaH_2PO_4	Sigma Life Sciences, St. Louis, USA
T	
Triton X-100	Sigma Life Sciences, MO, USA
Tris	Fisher Scientific, Loughborough, UK
Tetramethylethylenediamine (TEMED)	Sigma Life Sciences, MO, USA
Tris(2-carboxyethyl)phosphine hydrochloride (TCEP)	Sigma Life Sciences, MO, USA
U	
Uranyl acetate	Sigma Life Sciences, MO, USA
Urea	MP biomedicals, Loughborough , UK

2.2 Methods

2.2.1 Molecular biology

2.2.1.1 Preparation of competent *E. coli* for transformation

10 mL of LB medium was inoculated with a desired strain of *E. coli*; this was grown at 37 ˚C, 200 rpm for 16 hours. 5 mL of the culture was used to inoculate 100 mL of LB medium and was grown to an optical density at 600 nm (OD_{600}) of 0.45 before the cells were harvested by centrifugation at 4,000 rpm for 10 minutes at 4 ˚C (Eppendorf 5804R Refrigerated Benchtop

Centrifuge). The supernatant was discarded and the pellet gently resuspended in 10 mL of sterile pre-chilled (4 ˚C) 100 mM $CaCl_2$. The resuspension was incubated on ice for 10 minutes before centrifugation to re-pellet the cells. The supernatant was discarded and the pellet was gently resuspended in 2 mL of pre-chilled (4 ˚C) 100 mM $CaCl_2$, 30 % (w/v) glycerol. 100 µl aliquots were pipetted into sterile Eppendorf tubes sat on dry ice. Once frozen, the competent cells were stored at -80 ˚C.

2.2.1.2 Q5 NEB site directed mutagenesis

Site directed mutagenesis was carried out using the NEB Q5 site-directed mutagenesis kit (Section 2.1.16). Primers were designed using the NEB online tool (http://nebasechanger.neb.com/) and purchased from Eurofins MWG Operon. Both small changes such as when substituting an amino acid for a Cys residue into full length αSyn, as well as larger insertions such as that required to insert the 12 residue αSyn central NAC region into pL were carried out using this method.

The following reagents were assembled in a 0.2 ml PCR tubes:

Reagent	**Volume (µl)**
Q5 Hot start high-fidelity master mix (2X)	12.5
10 µM Forward primer	1.25
10 µM Reverse primer	1.25
Template DNA (1-25 ng/µl)	1
Nuclease free water	9

This was the subjected to the following PCR thermal cycle:

Step	**Temperature (ºC)**	**Duration (s)**	
Initial denaturation	98	30	
PCR	98	10	
	50-72*	30	X 35
	72	30 /kb	
Final extension	72	120	
Hold	4	∞	

*Annealing temperature depends on mutagenic primers and were taken from the online tool NEBaseChanger (http://nebasechanger.neb.com/).

The PCR reaction mixture from the previous step was the subjected to a kinase, ligase and DpnI (KLD) enzyme treatment. The following reagent were set up in a new 0.2 ml PCR tube:

Reagent	Volume (µl)
PCR product	1
KLD reaction buffer (2x)	5
KLD enzyme mix (10x)	1
Nuclease free water	3

5 µL of the reaction product was transformed into DH5α cells (Section 2.1.21) by heat shock (42 °C for 30 seconds). This was then plated onto antibiotic selection agar plates (100 µg/ml carbenicillin) after an hour incubation at 37 °C, 200 rpm in SOC medium provided with the kit.

2.2.1.3 Preparation of plasmid DNA for sequencing

Single colonies were picked from antibiotic selection agar plates and used to inoculate 10 mL of LB medium with the same selection antibiotic as the agar plate (100 µg/mL carbenicillin). The inoculated medium was incubated for 16 hours at 37 °C, 200 rpm and the cells were subsequently harvested by centrifugation at 4,000 rpm for 10 minutes, 4 ˚C (Eppendorf 5804R Refrigerated Benchtop Centrifuge). Using a Wizard® Plus SV Minipreps DNA Purification System kit (Section 2.1.16), the plasmids were purified from the bacteria and the concentration in water was calculated using a nanodrop 1000 spectrophotometer using the optical density at 260 nm (A_{260}) (concentration (µg/mL) = 50 µg/mL x A_{260}). The plasmid DNA was diluted to 100 ng/µl and sent for sequencing (Beckman Coulter Genomics).

2.2.1.4 Protein expression and purification

2.2.1.5 Starter culture

200 mL of sterile LB medium containing 100 µg/mL carbenicillin was inoculated with a single colony from an ampicillin selection plate of freshly transformed expression cells, BL21 (DE3), with the required plasmid constructs (Section 2.1.20). The inoculated medium was incubated at 37 °C, 200 rpm for 16 hours.

2.2.1.6 Expression and purification of pL constructs

10 X 1L cultures were inoculated with 10 mL starter culture (2.2.1.5). The cells were grown to OD_{600} = 0.6 before protein expression was induced with 1 mM final concentration IPTG. The cultures were allowed to grow for a further 4 hours before harvesting. Cell pellets were resuspended in lysis buffer (Section 2.1.17), homogenised and further lysed by cell disruption at 30K PSI. The cell lysis solution was then centrifuged at 12000 rpm for 30 minutes. The cleared lysate was decanted and syringe-filtered through a 0.22 µm membrane. The proteins were subjected to Ni-NTA purification on a 5mL His-Trap FF column (GE Healthcare). The column was washed with pL wash buffer (Section 2.1.17), before the protein was eluted in high imidazole containing pL elution buffer (Section 2.1.17). The eluted proteins were then concentrated using Vivaspin centrifugal columns (20 ml columns 5000 kDa MWCO) to less than ~3 mL and applied onto a HiLoadTM 26/60 Superdex 75 prep grade SEC column (GE Healthcare). The SEC column was equilibrated and run with pL SEC buffer (Section 2.1.17) at 3 mL/ min. Fractions were analysed by SDS-PAGE. The main peak corresponding to chimeric pL variants was pooled. The protein was then dialysed into 20 mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) pH 7.5 (for pL SMFS experiments) before being flash frozen with liquid N_2. The presence, purity and the correct mass of proteins were confirmed by MS.

2.2.1.7 Expression and purification of full length αSyn constructs

All full length synuclein constructs (including Cys mutants and homologues) were expressed in BL21 (DE3) *E. coli* cells from a pET23a vector (Section 2.1.20). Cells were grown in LB medium at 37 °C. Expression was induced with 1 mM IPTG at OD_{600} = 0.6. The protein was expressed for 4 h post-induction, before harvesting by centrifugation (6000 x g, 30 min, 4 °C).

Pellets were resuspended in lysis buffer (25 mM Tris-HCl, pH 8.0, 100 µg/mL lysozyme, 50 µg/mL PMSF and 20 µg/mL DNase). The pellet was homogenised before heating to 80 °C for 10 min. The homogenate was then centrifuged (35,000 x g, 4 °C, 30 min) and the protein, isolated in the soluble fraction, was precipitated with 50 % (w/v) ammonium sulphate, incubated at 4 °C, 30 min. The suspension was centrifuged at 35,000 x g and the pellet resuspended and precipitated again in 50% (w/v) ammonium sulphate, 4 °C, 30 min. After a further centrifugation (35,000 x g, 4 °C, 30 min) the pellet was resuspended in 20 mM Tris-HCl, pH 8.0 prior to anion exchange.

The partially purified α-syn was loaded onto a Q-Sepharose anion exchange column with a 20 mM Tris- HCl, pH 8.0 mobile phase. The protein was eluted with a linear gradient of 0-500 mM

NaCl and monitored by absorbance at 280 nm. Fractions containing α-syn were analysed by SDS-PAGE, dialysed against 50 mM ammonium bicarbonate and lyophilised.

The protein was the resuspended in 20 mM sodium phosphate, pH 7.5, and loaded onto a HiLoadTM 26/60 Superdex 75 prep grade gel filtration column. The protein was eluted from the column with 20 mM sodium phosphate, pH 7.5 at a flow rate of 3 mL/min. The protein was dialysed against 50 mM ammonium bicarbonate and lyophilised. Purified protein was stored at – 20 °C. Examples of chromatographic traces and SDS-PAGE gels of the synuclein purification are shown in Chapter 3.

2.2.2 Biochemistry techniques

2.2.2.1 Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

SDS resolving and stacking gel solutions (Section 2.1.17.3) were freshly prepared and APS and TEMED were added directly before pouring the gel mixture into a sealed clean casting chamber. Once the resolving gel had set, the stacking gel was added and a comb inserted into the top of the casting chamber. Excess stacking gel was removed using a pipette tip. The rubber seals were removed once the gel had set and the casting chamber containing the gel was inserted into the cathode chamber of an electrophoresis cell, which was then filled with 1x cathode buffer (Section 2.1.17.3). The anode chamber was then filled with 1x anode buffer (Section 2.1.17.3). 10 µL of the protein sample was mixed 1:1 with 2x SDS loading buffer (Section 2.1.17.3) and boiled for 10 minutes. 15 µL of this solution was loaded into the wells of the stacking gel. 5 µL of Precision plus protein dual colour standards protein ladder (BioRad, CA, USA) was also loaded into a single well for molecular weight determination. The electrodes were then connected to a power supply and 35 mA per gel was applied until the stained sample rested at the surface of the resolving gel. The voltage was then increased to 65 mA per gel and was electrophoresed until the stain left the bottom of the resolving gel. The gel was removed from the casting chamber and submerged into Instantblue coomassie stain (Expedeon) and incubated on a rocking table for ~2 hours before images were recorded.

2.2.2.2 Thioflavin T (ThT) aggregation assays

2.2.2.2.1 Synuclein peptide ThT

α- and γSyn 71-82 peptides were dissolved in 100 % (v/v) HFIP (hexoflouroisopropanol) at 450 µM and dispensed into Corning 96-well flat bottom assay plates. 50 µl was dispensed into wells, and the HFIP was left to evaporate. The dry peptide in the well was dissolved into 100 µl of 20 mM HEPES, 20 µM ThT, pH 7.5 to give a concentration of 225 µM peptide. Incubations were carried out at 37 ^{o}C shaking at 600 rpm. The samples were excited at 444 nm and the fluorescence emission was monitored at 480 nm on a BMG Labtech FLUOstar optima plate reader with a gain set at 1450.

2.2.2.2.2 pL construct ThT

Using the same plate reader conditions as above (Section 2.2.2.2.1), 100 µM pL constructs were incubated at 37^{o}C shaking at 600 rpm in 20 mM HEPES, 20 µM ThT, pH 7.5.

2.2.2.2.3 Full length synuclein ThT

Using the same plate reader conditions as above (Section 2.2.2.2.1) 100 µM synuclein constructs were incubated at 37^{o}C shaking at 600 rpm in various conditions as discussed in the experimental section. The gain was typically set at 350.

2.2.2.2.4 Fibril yield analysis

Fibril yields were assessed by densitometry analysis using bands from SDS-PAGE (Section 2.2.2.1) gels using the GeneTools programme (Syngene). Samples of full length synuclein from the end point of ThT experiments were spun down in a bench top GenFuge 24D centrifuge (Progen Scientific, London, UK) at 16,000 x g for 30 minutes. Supernatant samples were taken and the pelleted samples were resuspended in buffer using the same volume as that of the supernatant for accurate comparisons. Two dilutions of each sample were made and all samples were prepared and run on SDS-PAGE gels (Section 2.2.2.1). Fibril yields were calculated from the percentage of pelleted material related to the whole material in pelleted and supernatant samples.

2.2.2.2.5 Data analysis

Where normalised data is presented, it has been processed on the plate reader software and normalised (after buffer subtraction) between 0 and 100. Lag time analysis was carried out by manually fitting a linear regression of the steepest exponential region of the ThT curves on

Origin Pro 9.1 software. The fit was extrapolated to calculate the x-intercept (quoted lag times).

2.2.3 Biophysical techniques

2.2.3.1 Circular dichroism (CD)

To gain secondary structural information for the purified proteins, Far UV (190-260 nm) circular dichroism (CD) spectroscopy was performed in a 1 mm path length curvette (Hellma) using a ChirascanTM plus CD Spectrometer (Applied Photophysics, U.K.).

200 µL of 50 µM protein solution was placed into the cuvette before a far-UV CD spectrum was acquired using a 1 nm bandwidth at room temperature, 1 s time step. An average of 3 scans (190-260 nm) were taken per sample.

CD experiments on pL constructs were carried out in 25 mM sodium phosphate buffer, 2 mM DTT, pH 8.0. Experiments on full length αSyn were carried out in either 20 mM Tris, 200 mM NaCl, 2 mM DTT pH 7.5 or 20 mM acetate, 200 mM NaCl, 2 mM DTT, pH 4.5.

2.2.3.1.1 CD temperature ramp

The thermal denaturation of pL variants was measured by setting up a temperature gradient from 20 to 90 ^{o}C in 1 ^{o}C steps. Protein samples were incubated for 180 s at each temperature before CD spectra were taken as above. After the thermal melt was carried out, a spectra at 20 ^{o}C was recorded to ensure that folding was reversible. The thermal melt data were analysed on Photophysics Global3 software.

2.2.3.2 Fluorescence spectroscopy

Intrinsic tryptophan emission spectra of pL constructs were recorded on a Photon Technology International fluorometer (Ford, West Sussex, UK). Excitation and emission slit widths were set to 1 and 2 nm, respectively. Proteins were excited at 280 nm and emission spectra were recorded at 290-400 nm.

Spectra were recorder for 200 µl of 50 µM pL constructs in 25 mM sodium phosphate buffer, 2mM DTT, pH 8.0. All data were normalised to the unfolded state. The unfolded state was recorded in the same conditions as above in the presence of 8M urea.

2.2.3.3 Size exclusion chromatography - small angle X-ray scattering (SEC-SAXS)

All SEC-SAXS experiments were performed in partnership with, and analysed by, Dr Samuel Lenton (University of York).

Size exclusion chromatography in-line with small-angle X-ray scattering (SEC-SAXS) experiments were performed at Diamond light source beamline B21. A 2.4 mL Superdex™ 200 increase 3.2/300 column connected to an Agilent 1200 HPLC system was equilibrated with either 20 mM Tris, 200 mM NaCl, pH 7.5 or 20 mM acetate, 200 mM NaCl, pH 4.5. 55 µl of αSyn construct (10 mg/mL) was loaded into a well of a 96 well plate; the amount loaded onto the column was 45 µL. A fixed wavelength of 1.0 Å (12.4 keV) was used with the X-ray detector (PILATUS 2M) placed 4 meters from the sample (suitable for particles with an Rg < 200 Å), the flow rate through the detector was 0.075 mL/min.

Buffer contributions were subtracted using the program SCATTER (version 3.0a) and a Guinier approximation (Equation 2.9) of the scattering vector ((ln[I(q)] vs q^2) where I(q) is the background corrected intensity and q is the momentum transfer ($q = 4\pi \sin(\theta)/ \lambda$, where 2θ is the angle between the incident and scattered beam and λ is the beam wavelength)) was used to estimate the radius of gyration (Rg) by limiting q x Rg < 1.3, typical for an IDP.

$$\ln I(q) = \ln I(0) - \frac{{R_g}^2}{3}.q^2 \qquad 2\text{-}1$$

2.2.3.4 Ensemble optimisation method (EOM)

To gain structural information of αSyn in different conditions from the SAXS experimental data, the software EOM (version 2.0)[333] was used. EOM fits the averaged theoretical scattering intensity from an ensemble of conformations to experimental SAXS data. A pool of 10,000 independent models based upon the sequence of αSyn was initially generated. A genetic algorithm compares the averaged theoretical scattering intensity from the independent ensembles of conformations against the experimental SAXS data. The conformations that best describe the experimental data are selected.

2.2.3.5 Heteronuclear single quantum coherence (HSQC) nuclear magnetic resonance (NMR) spectroscopy

NMR experiments were performed in partnership with, and initial analysis carried out by Dr Theo Karamanos (University of Leeds). Final processing of the data was carried out by Hugh Smith (University of Leeds).

HSQC spectra of pL αSyn_{71-82} and pL γSyn_{71-82} (both 400 µM) were recorded on an AVANCE III Bruker spectrometer (600 MHz) equipped with a cryogenic probe in the presence of 20 mM HEPES buffer pH 7.5. Spectra were processed in NMRPipe and analysed in CCPN analysis. Experiments were carried out in 20 mM HEPES buffer, pH 7.5 with 15 ul D_2O and 0.02% (w/v) sodium azide.

2.2.3.6 Negative stain transmission electron microscopy (TEM)

The formation and morphology of fibrils was examined by negative stain TEM. 3 µl samples were pipetted onto the surface of carbon coated copper grids provided by the University of Leeds. The sample was left for 30 seconds and then removed by blotting with filter paper. The grid was washed by the addition of 3 µl water which was then removed as described previously. The grid was washed with water twice more. Two x 3 µl of 1 % (w/v) uranyl acetate was added and blotted as before. The last stain with uranyl acetate was left for 15 seconds before blotting. Images were taken on an FEI T12 electron microscope.

2.2.4 Atomic force microscopy (AFM) based single molecule force spectroscopy (SMFS)

2.2.4.1 Surface functionalisation (for SMFS)

2.2.4.1.1 Oxidisation of silicon nitride AFM probe and surface

Silicon nitride surfaces were cut into 1c m^2 pieces from larger disks (Rockwood electronic material. In order to clean and oxidise both the silicon nitride surfaces and AFM cantilever probes, incubations with piranha solution (3:1 ratio of 0.5 M (>95 %) H_2SO_4 to 30 % (v/v) H_2O_2) was performed. Surfaces were incubated for 30 minutes and then washed with water before being dried in N_2. Fresh AFM probes were submerged in piranha for 30 seconds and washed in the same way as the silicon nitride surfaces. The surfaces and AFM probes were then placed on a microscope slide inside a petri dish with a hole in the lid and placed under a UV lamp (UVIlite, UVItec) set to 254 nm for 30 minutes (ozone cleaning).

2.2.4.1.2 Aminosilination of silicon nitride

Oxidised AFM probes and surfaces were placed into a desiccator along with 80 µL of (3-aminopropyl)triethoxysilane (APTES) and 20 µL of N,-N-diisopropylethylamine (DIPEA) held in separate 1.5 mL Eppendorf tube lids. The desiccator was evacuated using a vacuum pump for 1 minute and left to incubate at room temperature for 2 hours. After the incubation, the APTES and DIPEA solutions were removed and the desiccator was flooded with N_2 and left to cure for 48 hours.

2.2.4.1.3 Functionalisation with NHS-PEG_{24}-maleimide linkers

Amino-silanated AFM probes and surfaces were immersed in ~1 mL chloroform containing 20 µL of 250 mM SM PEG_{24} (succinimidyl-[(N-maleimidopropionamido)-diethyleneglycol] ester linkers in DMSO) and left to incubate at room temperature for 1 hour. Both AFM probes and surfaces were then washed with chloroform and dried with N_2.

2.2.4.1.4 Protein immobilisation

Full length synuclein variants (5 µM) or pL constructs (50 µM) both containing engineered cysteine residues were deposited over the functionalised surfaces and AFM probe and left to incubate in a covered container for 30 minutes at room temperature. All proteins analysed by SMFS were incubated onto functionalised probes and surfaces in the presence of 1 mM TCEP (tris(2-carboxyethyl)phosphine) in order to limit disulfide linkage of proteins, thereby increasing the free cysteines available for immobilisation. Unreacted protein was then washed from the surface and AFM probe with the reaction buffer as quoted in the experimental section.

2.2.4.2 Cantilever calibration

The AFM probe (MLCT with reflective gold, Bruker) functionalised with the analyte protein was inserted into a cantilever holder and secured. The silicon surfaces with functionalised protein were attached to a microscope slide with Loctite superglue and secured to the XY scanner with magnetic bars.

A droplet of reaction buffer was applied to the silicon surface and was held by surface tension. The AFM probe in the holder was mounted to the MFP-3D head (Asylum) and approached towards the surface until the probe was fully submerged in the buffer droplet. Using the inbuilt optics, the laser was positioned to the tip of cantilever D (manufacturers spring constant: 30 pN/nm) and the deflection was set to zero. The cantilever was engaged using the Asylum Research software (MFP version 11), engagement causes a Z-piezo voltage maximum (+150),

an indication of full Z-piezo extension and zero surface contact. Using the thumb wheel on the MFP-3D head (Asylum), the cantilever was approached to the surface. The z-voltage was adjusted to 70, assuring that the piezo was in the middle of its z range (7.5 µm). Spring constant calibration was carried out as a two-step procedure: determining the slope of contact from a force curve to find the sensitivity of the cantilever (nm/V) and then performing the non-destructive thermal tune to determine resonant frequency of the cantilever[334]. A single force-extension plot was recorded with the trigger (amount of deflection the cantilever undergoes before retraction) set to 20-40 nm, which gives a quantifiable deflection slope upon hard contact of the cantilever and the surface. The slope of the contact region (inverse optical lever sensitivity (InvOLS)) was measured by a linear fit. The cantilever was withdrawn from the surface and the deflection was set to 0. A thermal tune was carried out to detect the natural thermal fluctuation of the cantilever by performing ~100 frequency sweeps (0-1 MHz). The drive frequency (first major resonance peak) was selected and a Lorentzian function was fit. The area of the thermal fluctuations (P) is used to find the spring constant (k) using Equation 2-2:

$$k = \frac{k_B T}{P} \qquad 2\text{-}2$$

where k_B is the Boltzmann constant, T is temperature in Kelvin.

The spring constant was always within error of the manufacturer's guidelines before data collection. Hooke's Law (Equation 2-3) allows the calculation of the force applied (*F*) to the cantilever with a known spring constant by the deflection (or extension (*x*)).

$$F = kx \qquad 2\text{-}3$$

2.2.4.3 Data collection in SMFS

The feedback loop trigger point was set to 10 nm, the retraction distance was typically set to 1000 nm and approach velocity was set to 2 µm/s during data collection and kept constant whilst using various retraction velocities and sample rates (10 kHz per µm/s velocity). The buffer on the sample surface was replenished frequently to prevent evaporation. Force maps of 20 μm^2 with 500 approach-retract cycles were taken to maximise surface coverage, and between force maps the AFM probe was repositioned manually using the XY scanner. The retraction velocities used in the DFS study on pL variants were 200, 500, 1000, 3000 and 5000

nm/s. Typically all other SMFS experiments were carried out at 1000 nm/s including experiments on full length synuclein variants.

2.2.4.4 Contour length (L_C) simulations

The L_C simulation method carried out in this thesis are described in Farrance *et al,* 2015[306] and further developed by Dr Yun Chen (Universisty of Leeds, unpublished work). The model predicts L_C measurements from SMFS experiments that consider attachment position on the AFM tip, geometric effects, and polymer dynamics of the linkers[306]. In the ideal situation, one immobilisation point for protein would be located on the apex of the cantilever tip whereas the second immobilisation point would be directly below on the substrate. In this scenario, the observed L_C would simply be a sum of the length of the components of the tethered complex. In reality, this situation occurs very rarely, it is possible that binding may occur between partners whose linkers are immobilized at other locations. The model therefore uses an iterative Monte Carlo to randomly pick immobilisation points on both the AFM tip (with a modelled geometry of a square-based pyramid) and the substrate surface. The separation of the two proteins is then used to calculate the binding probability as described in Farrance *et al,* 2015[306].

Additional parameters were added to the published model, such as a Monte Carlo procedure which accounted for the stochastic rupture of protein complexes which can occur anywhere along the theoretical WLC parabolic curve. This property has an effect on the L_C values obtained.

The protein sequences were split into 10 residue segments, each of the 10 residues were modelled as interacting with the corresponding region of the second immobilised protein monomer. The linker length was set to 10 nm as in SMFS experiments (SM-PEG_{24}). The length of αSyn was also set as 140 residues with the carbon α-carbon α position taken as 0.4 nm (total length of αSyn monomer = 56 nm). Different immobilisation points were taken into account in the fitting. Each L_C simulation was carried out for a total of 50 million iterations.

2.2.4.5 Data processing

All force spectroscopy data were analysed using IGOR pro 6.32A with an Asylum Research extension (MFP3DXop v30).

The hard contact (0 nm) and baseline (0 pN) of all the force-extension retraction traces were manually set, by taking a section of the hard contact or the retraction with no events and

setting the zero to the centre of the data (force baseline has a thermal noise of ± 10 pN). The WLC model (with a fixed persistence length of 0.4 nm, Equation 1-1) was fit manually to all unbinding events that are parabolic by inserting locks at the apex and the base of the curve and recording the contour length (L_C) and extension (x).

Force-distance plots were binned for analysis when single characteristic, parabolic WLC events were observed. Single Gaussian distributions were fit to unbinding force and contour length histograms in order to determine the most probable force and contour length at rupture for each retraction velocity investigated. For each pulling velocity used, data were collected in triplicate (using a freshly prepared cantilever for each repeat).

For DFS analysis, loading rates were calculated by fitting a WLC model to the rising edge of each unbinding profile when plotted as force versus tip-sample separation. The instantaneous gradient of this fit at rupture (WLC_{slope}) was calculated by inserting the derived contour length and extension at rupture into a differentiated form of the same equation (Equation 2-4). The loading rate at rupture was then obtained by multiplying this value by the retraction velocity.

$$WLC_{slope} = \frac{k_B T}{p}\left(\frac{1}{2L_c\left(1-\frac{x}{L_c}\right)^3}\right)+\frac{1}{L_c} \qquad 2\text{-}4$$

where; p is the persistence length, L_C is the contour length, x is the extension, k_B is the Boltzmann constant and T is the temperature.

The natural logarithm of the mean loading rate (N/s) at each velocity was plot against the mean rupture force (N) which gives a linear relationship (dynamic force spectrum see Section 1.6.3). The Bell-Evans model[309] (Equation 1-2) was rearranged to use the gradient of the linear fit to calculate the distance from the transition state (x_u) (Equation 2-5) and the y-intercept for the off rate at zero force (k_{off}^{0F}) (Equation 2-6).

$$x_u = \frac{k_B T}{m} \qquad 2\text{-}5$$

Where m is the gradient parameter in the linear fit $y = mx + c$

$$\frac{1}{k_{off}^{0F}} = \frac{k_B T}{x_u} \exp\left[\frac{c\, x_u}{k_B T}\right] \qquad 2\text{-}6$$

Where C is the y-intercept parameter in the linear fit $y = mx + c$

2.2.4.6 Data processing utilising an automated process

An automated process for analysing force-distance data was developed by Dr Yun Chen (University of Leeds) using python scripting software. The force and separation channels from force-distance data were exported from IGOR pro 6.32A software as .txt files. The horizontal baseline was fit to the first 100 nm of force-extension plots, avoiding later points distal to the surface that can be subject to drift. The vertical baseline was fit to 500 points before the start of the horizontal section (pre-surface engagement). WLC models were fit to data that deviated from the horizontal baseline. Each force distance plot was subjected to 400 fitting attempts. L_C thresholds were set at > 10nm and < 150 nm, this eliminates events that are close to the surface and have a high probability of noise in the data and also events that are more than the theoretical maximum L_C. Any peaks that fall outside these thresholds are discounted. The lower F_R threshold was set to 20 pN. Lower forces than this are outside the force detection limit. A complete peak was described as one in which the entropic restoration part of the peak returns to within 25 pN of the horizontal baseline over 5 nm.

The script cycled through each force-distance plot in an experiment and using the parameters above, WLC events were fit. A quality control parameter was included in the fitting as a reduced chi-square value. This was set so that any fit above 10 was discounted (this value is quite high so that filtering is conservative and real hits are not missed). After events had been selected via the automated script, a manual screen was then carried out in order to filter any false-positive data. The majority of the SMFS on the full synuclein variants were analysed utilising this method of data processing.

2.2.4.7 Data analysis

Positive hits were fit with a worm-like chain model and the L_C and F_R distributions were plotted in histograms with 5 nm and 7 pN bin sizes respectively. Both of these data sets were combined into 2D scatter or contour plots. The ability of this method to observe bone fide protein dimerisation interactions was validated by the presence of obvious peaks in L_C and F_R distributions as well as correlations of these in 2D scatterplots, these aspects of the data would

not be present if non-specific interactions were being analysed. Correlation of L_C and F_R in 2D scatter or contour plots were assessed in a qualitative manner.

Hit rate quoted throughout the thesis are the percentage of positive hits out of the total number of approach retract cycles. The errors on hit rate are SD between different experiments.

2.2.5 Mass spectrometry (MS)

All mass spectrometry experiments in this thesis were carried out by Dr Lydia Young.

2.2.5.1 Native-MS

All native MS analyses of αSyn fragments, full-length synuclein variants and pL constructs were performed using a Synapt HDMS quadrupole-time-of-flight mass spectrometer (Waters Corpn., Manchester, UK), equipped with a Triversa NanoMate (Advion Biosciences, Ithaca, NY, USA) automated nano-ESI interface. The instrument has three travelling wave ion guides (TWIG) situated between the quadrupole and the time-of-flight analysers (Figure 2-1). The first TWIG is known as the trap region, the second as the IMS cell and the third as the transfer region. The use of hybrid instruments, such as the Synapt HDMS, allows additional characterisation of protein or peptide samples using tandem mass spectrometry (MS/MS), most commonly via a collisional induced dissociation (CID) mechanism[335].

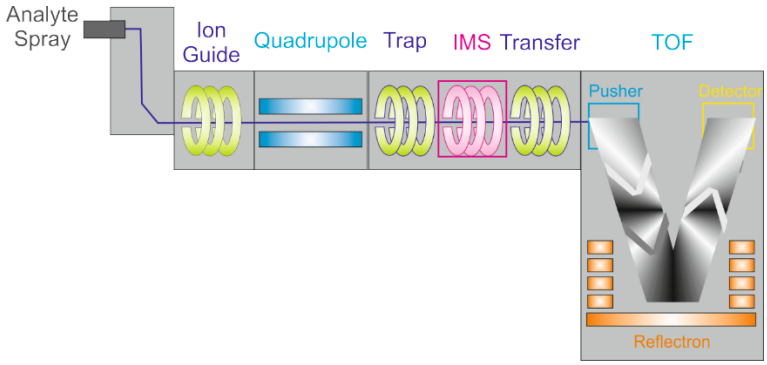


Figure 2-1. Schematic showing the layout of the Synapt G1 HDMS used for native MS analysis. Figure adapted from Pringle *et al*. 2007[336].

2.2.5.2 Peptide, full length synuclein variants and pL construct analysis

Fragment peptides (αSyn and γSyn 71-82), full-length synuclein variant proteins (αSyn, βSyn, γSyn and αSyn E46K) and pL construct (pL GS, pL αSyn_{71-82}, pL γSyn_{71-82}) samples were all analysed at a 100 µM final concentration in 100 mM ammonium acetate buffer, pH 6.8. Samples for MS were analysed using positive mode nanoESI with a capillary voltage of 1.7 kV

and a nitrogen nebulising gas pressure of 0.8 psi. The following instrumental parameters were used: cone voltage 30 V; source temperature 60 °C; backing pressure 3.2 mBar; ramped travelling wave height 7–20 V; travelling wave speed 300 m/s; IMS cell pressure 0.55 mBar. Data were acquired over the range m/z 500–6,000. Data were processed by use of MassLynx v4.1 and Driftscope v2.4 software supplied with the mass spectrometer. Mass calibration was achieved using caesium iodide solution, prepared by dissolving the compound in 50 % (*v/v*) water/ isopropanol to a concentration of 2 mg/ml.

2.2.5.3 Collision induced dissociation (CID)-MS/MS of synuclein variant dimers

CID-MS/MS experiments were performed using the quadrupole analyser to select synuclein variant dimer ions. Increasing collision energy was applied to the trap collision cell in 10 V increments from 0-80 V, until the dimer ions were completely dissociated into monomer ions.

2.2.6 Zyggregator bioinformatics analysis

The bioinformatics tool Zyggregator was utilised to study the aggregation/amyloid propensity of full length synuclein sequences. Developed by M. Vendruscolo and colleagues[337-339], the software takes into account various intrinsic physio-chemical properties of the protein sequence and their contribution to aggregation propensity (as discussed further in Chapter 6).

3 Development and implementation of a display system for aggregation-prone peptides in SMFS

3.1 Abstract

Protein aggregation is linked with the onset of various neurodegenerative disorders, including PD, which is associated with the aggregation of αSyn as discussed in Chapter 1. The mechanistic details of protein aggregation remain elusive[1,51,56], and methods to interrogate the species formed during aggregation, especially in the initial stages involving the first dimer formation, are distinctly lacking. In this chapter, we have attempted to probe the first intermolecular contacts that initiate aggregation at the single molecule level, utilising single molecule force spectroscopy (SMFS) to study the fragment αSyn_{71-82}. This region has been shown to be necessary for the aggregation of full length αSyn[204], and is capable of forming amyloid fibrils in isolation[204]. We demonstrate that the interaction of αSyn_{71-82} monomers can be studied using SMFS. We have achieved this by designing and engineering residues 71-82, a highly aggregation-prone peptide of αSyn[204] into protein L, a mechanically strong soluble scaffold protein[340-342] in order to act as a display system for SMFS studies. The corresponding fragment of the homologous protein γ-synuclein (γSyn), which has a lower aggregation propensity[275], has also been studied here. The results from SMFS, in combination with native ESI-mass spectrometry analyses and aggregation assays, demonstrate that the dimerisation propensity of γSyn_{71-82} is lower than that of αSyn_{71-82}, but that mixed αSyn_{71-82}: γSyn_{71-82} dimer forms with a similar propensity to the αSyn_{71-82} homodimer. This study illustrates the utility of a novel display protocol for the study of small aggregation peptides using SMFS which would otherwise be difficult to study. We have also shown that this display SMFS technique has yielded new mechanistic insights into the specific system studied.

3.2 Introduction

Studying protein aggregation is a significant challenge, owing to the fact that the aggregating species are heterogeneous, only transiently and lowly populated, and the fact that aggregation- prone proteins progress on an exponential timescale towards higher order, end point species as discussed in reviews and Chapter 1[1,51,56]. Early events in the aggregation cascade, therefore, are extremely difficult to study. However, these events are of critical importance in relation to protein aggregation. Gaining a greater mechanistic or structural understanding of these early events would yield an attractive target for preventative and curative treatments for these diseases.

Ensemble methods have traditionally been used to investigate protein aggregation and despite their limitations have revealed valuable information about the conformation of aggregation-prone protein monomers and their self-interactions. Small angle X-ray scattering (SAXS) experiments have been used to gain insights about the conformation of aggregation-prone proteins in an ensemble. Uversky and co-workers have utilised SAXS and a computational method of ensemble optimisation modelling to show that αSyn and its aggregation-prone mutants have two major conformer groups representing the beginning of oligomer and fibril formation pathways[343]. Paramagnetic relaxation enhancement (PRE) NMR experiments have been used by our group to investigate the interchain interactions of a folded aggregation-prone protein β_2M and have demonstrated that weak interactions ($K_D \approx 100$ μM) play an important role in the promotion and inhibition of protein aggregation[344]. Similar intermolecular PRE experiments have been carried out on αSyn and its homologue βSyn in order to investigate the inhibitory effect of βSyn on the aggregation of αSyn[280]. This study showed that a heterologous interaction is of five-fold higher affinity than the homologous αSyn interaction, giving evidence for a kinetic trap mechanism for inhibition[280].

Dye binding assays have been used extensively in studies of aggregation-prone proteins, the most prevalent of which is Thioflavin T (ThT)[345]. These experiments are valuable in obtaining information about an aggregating system. They report on the presence of cross-β amyloid structures[345] but can give false positives or false negatives as the dye is not perfectly specific for amyloid[345]. They also lacks resolution regarding the early lower oligomeric species in an aggregation pathway as they may not bind these species.

Fluorescence experiments at the single molecule level have been developed to address some of the issues with traditional ensemble methods of interrogating an aggregating system as they can be performed at low concentrations. Single-molecule fluorescence resonance energy transfer (FRET) and fluorescence correlation spectroscopy (FCS) methods have been utilised to elucidate the distribution of oligomeric species[104,313]. The Lyubchenko group have used a total internal reflection fluorescence microscopy (TIRFM) in which monomeric αSyn was immobilised to a surface and then incubated with fluorescently labelled αSyn[346]. These experiments were able to show lifetimes of dimeric αSyn and how this varied for aggregation-prone αSyn variants[346].

SMFS studies have also been utilised to address some of the issues associated with studying aggregation-prone proteins using ensemble techniques (as discussed in Section 1.6.4). These studies are able to probe the dimerisation of aggregation-prone molecules at a single molecule

level, as the experiments can be carried out at low protein concentrations, limiting the amount of concentration-dependent aggregation. By studying the interactions of these proteins in a single molecule setting, the contributions from heterogeneous higher order species do not complicate data analysis. These studies are however complicated by the fact that non-specific interactions inherent to the experimental set up and/or interactions of higher order events may occur, which may mask monomer-monomer interactions.

In this chapter, we have circumvented some of the problems encountered in aggregation studies and in existing SMFS experiments of aggregation-prone proteins by developing a method in which a highly aggregation-prone region of a protein of interest is engineered into a loop of a mechanically stable monomeric carrier protein. The central NAC region (71-82) of αSyn was chosen as a test system for this display system.

The NAC region of αSyn contains a central, hydrophobic 12 residue long sequence at residues 71-82 (VTGVTAVAVAQK), known to be both sufficient and necessary for aggregation[204], for example, deletion of this sequence from full length αSyn ameliorating fibril formation[204]. Additionally, a recent ssNMR structure of an αSyn fibril showed that the NAC sequence forms the fibril core[100], thus playing a key role in αSyn aggregation. This region of αSyn has been used in this study to demonstrate how our developed method can be utilised to study the self-associations of even small stretches of aggregation-prone proteins at a single molecule scale.

αSyn_{71-82} and the same region from the homologue γSyn_{71-82} were engineered into a mechanically strong scaffold protein: protein L (pL)[340-342], and the chimeric protein was used as a display system for both αSyn_{71-82} and γSyn_{71-82} thereby solubilising highly hydrophobic peptides (schematic of experimental set up depicted in Figure 3-1, the peptide insertion sequences are shown in Figure 3-2). pL is a small 62 residue β-grasp protein that comprises four β-strands and one α-helix in the structure ββαββ. Previous work by this group has described the mechanical properties of pL in detail [340-342]. These studies demonstrated that pL is mechanically strong when pulled apart via the C- and N-termini due to the sheer force required to rupture the hydrogen-bonded parallel terminal β strands[340,342]. The mechanical strength of pL mutants were analysed[341] with the I60F mutation resulting in a 2-fold increase in mechanical strength with respect to WT pL (increased from 108 to 216 pN at $447 nms^{-1}$)[341]. This mechanically stabilised I60F variant was chosen for use in this study (termed pL throughout). The application of force is applied in the same geometry as the terminal β-strands which has been shown to yield greater mechanical resistance. In addition, this allows some control over the direction of force applied to the αSyn_{71-82} dimer complex. The insertion site is in an

extended loop region between β-sheets 3 and 4 (Figure 3-1). In this way, we are able to use pL as a display system in SMFS experiments in a similar manner in which small adheron proteins present peptide loops for binding to target molecules[347,348].

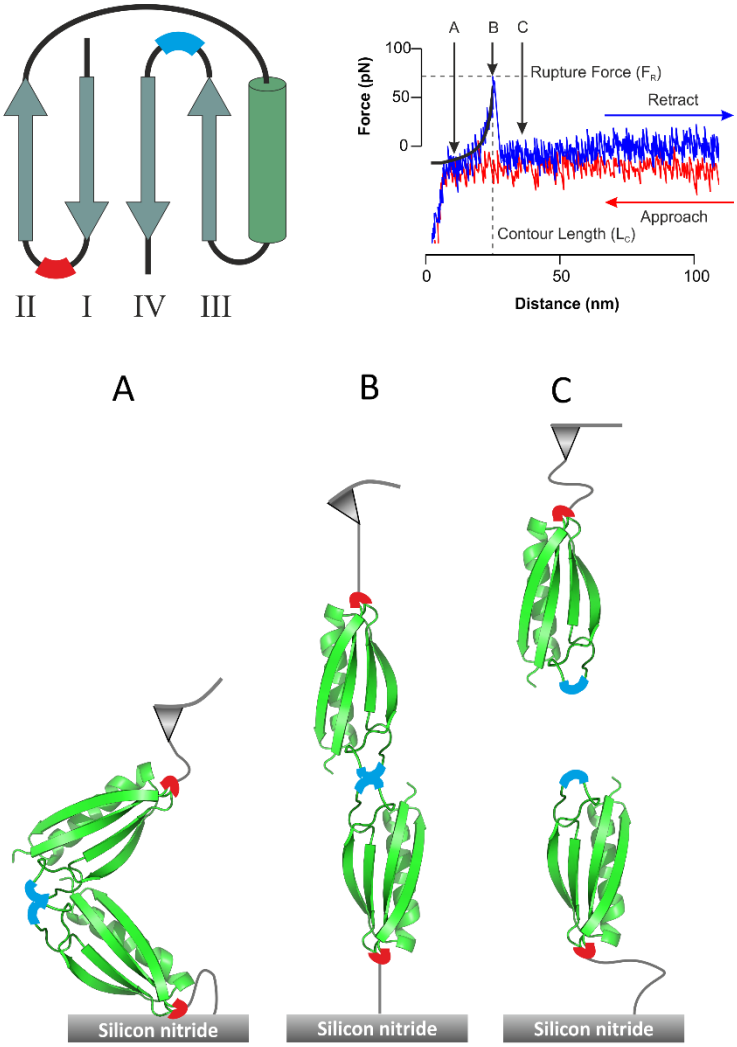


Figure 3-1. Schematic representation of the approach used in this study. Top left: secondary structure topology of the chimeric pL constructs used. The position of the engineered Cys residue and synuclein insert are shown in red and blue respectively. Bottom: Different peptide sequences, shown in blue were inserted into an unstructured loop of pL (loop has been extended using SWISS-MODEL based on PDB file 1HZ6[349]). The proteins were attached to a cantilever tip and a silicon nitride surface derivitised with hetero-bifunctional maleimide/ succinimide polyethylene glycol linkers via unique free cysteine residues present in the proteins. (A) The AFM tip derivitised with the chimeric pL constructs is brought into contact with a similarly derivitised surface and synuclein inserts form an interaction. The complex is then extended by retracting the cantilever until a force is reached which bends the cantilever (B) until a maximum force is reached (rupture force (F_R)) which dissociates the protein complex (C) at rupture force (F_R). By applying the Worm Like Chain (WLC) model shown by the black curve in the force-distance plot (top right), the length of the fully stretched complex before dissociation, or contour length (L_C) and the F_R can be calculated.

pL was suitable for use as a scaffold for the aggregation-prone peptide inserts given that the interactions between the aggregation-prone peptides would be expected to have a lower dissociation force than the unfolding forces required for pL[319,320,341]. In this experimental design, residues 71-82 of αSyn were engineered into a flexible loop extension in pL between β-strands III and IV. A cysteine residue was engineered into the loop region between β-strands I and II (Figure 3-1 and Figure 3-3). The cysteine was used to immobilise pL constructs to both Si_3N_4 cantilever tips and similarly functionalised surfaces. This method involves both surfaces and cantilevers being subjected to aminosilanisation, treated with heterofunctional PEG linker (MAL-PEG-NHS, MW 3400 Da), before covalently immobilising the chimeric pL molecules to the cantilever tips and surfaces. Single intermolecular interactions between chimeric pL monomers (which equates to the first step in the amyloid pathway) can be detected when the cantilever tip is brought into contact with, and retracted from a similarly derivitised surface (Figure 3-1).

In addition, we have verified our observations with an orthogonal technique: native electrospray ionisation-ion mobility spectrometry-mass spectrometry (ESI-IMS-MS) to study the assembly of the peptides, both within the pL scaffold and in isolation. This technique has been successfully applied to the study of multiple amyloid systems[350-354] and enables the detailed interrogation of every species within a heterogeneous, assembling mixture.

3.3 Engineering chimeric display proteins

3.3.1 Molecular biology

Molecular biology techniques were used to engineer the αSyn_{71-82} peptide and its homologue γSyn_{71-82} into pL. The residue sequences for these regions are shown in Figure 3-2. An existing protein construct already available in this lab was used as a starting point for molecular biology protocols. The starting construct was pL with a C-terminal hexa-His tag, a Cys residue in the flexible loop extension in pL between β-strands I and II a different peptide sequence in an extended loop region between β-strands III and IV. This peptide would be substituted with αSyn_{71-82} and γSyn_{71-82}. The starting construct contained an extended flexible loop region before and after the inserted peptide sequence. All constructs were engineered in a pET-23a plasmid.

New England Biolabs Q5© mutagenesis was performed in which the desired peptide inserts were substituted into the extended loop region between β-sheets 3 and 4 (Figure 3-3). Two extra flexible linker residues were added at either side of the peptide inserts using the same

mutagenesis protocol in order to effectively display the peptides in a flexible loop and to not cause steric constraints of this region.

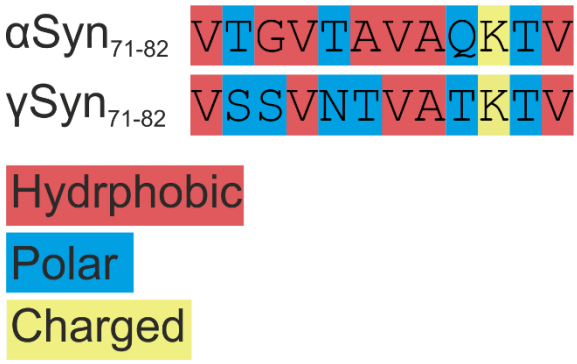


Figure 3-2. Sequence alignment of the central NAC regions of α and γ synucleins: αSyn_{71-82} and γSyn_{71-82}. Residue properties are colour coded: red, blue and yellow show hydrophobic, polar and charged residues respectively. The sequences are highly homologous, 7/12 residues are identical in the same positions between α and γ, 10/12 residues have the same properties in the same positions between α and γ.

The final protein sequence and structure is presented in Figure 3-3 as the pL αSyn_{71-82} construct that will be utilised in SMFS as a display protein. A non-aggregation control construct has was designed and engineered (termed pL GS) in which repeating units of Gly-Ser residues of the same length as the peptide inserts (12 residues/ 6 repeating units) and inserted into the same region of the pL display protein as αSyn_{71-82} and γSyn_{71-82}.

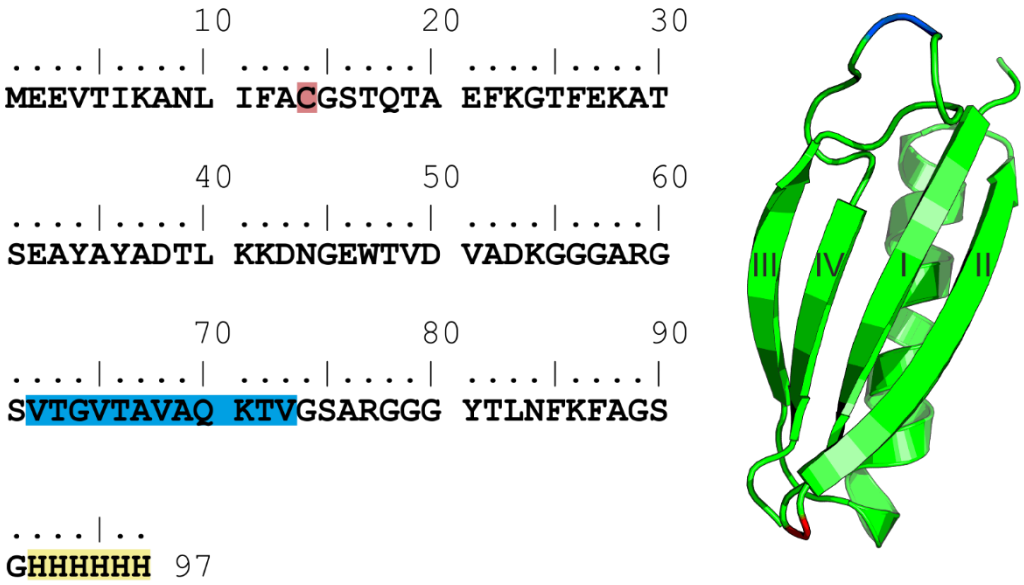


Figure 3-3. Sequence and modelled structure schematic of pL αSyn_{71-82}. The red highlighted text indicates the position of the engineered cysteine residue used for immobilisation in SMFS experiments. Blue highlighted text indicates the αSyn_{71-82} sequence. This is the same position also used for γSyn_{71-82} insertion and the non-aggregation control GS insertion. Yellow highlighted text indicates the presence of a hexa-His tag to enable Ni-affinity purification. The structural schematic on the right shows the engineered cysteine residue in the loop between β strands I and II in red; and the position of the peptide inserts is shown in blue in the extended loop region between β strands III and IV. The loop has been extended using SWISS-MODEL based on PDB file 1HZ6[349].

3.3.2 Expression and purification

pL constructs were transformed into *E. coli* BL21 (DE3) cells and protein expression was carried out as described in Section 2.2.1.6. The purification involved a two-step chromatographic process (Figure 3-4) in which the soluble fraction of lysed cells was taken forward for initial affinity capture chromatography step (Ni-affinity chromatography) before a second chromatography size exclusion chromatography (SEC) step was carried out. At each step in the purification process, the presence of ~10 kDa pL constructs were confirmed by SDS-PAGE analysis (Figure 3-4 A). The purity of the proteins were confirmed by SDS-PAGE and ESI-MS analysis. Mass spectra of the pL constructs showed masses 133 Da lower than that expected as the N-terminal Met residue is excised, due to the activity of the *E. coli* enzyme methonyl-aminopeptidase which has increased activity with decreasing residue size in the penultimate position (alanine in this case)[355].

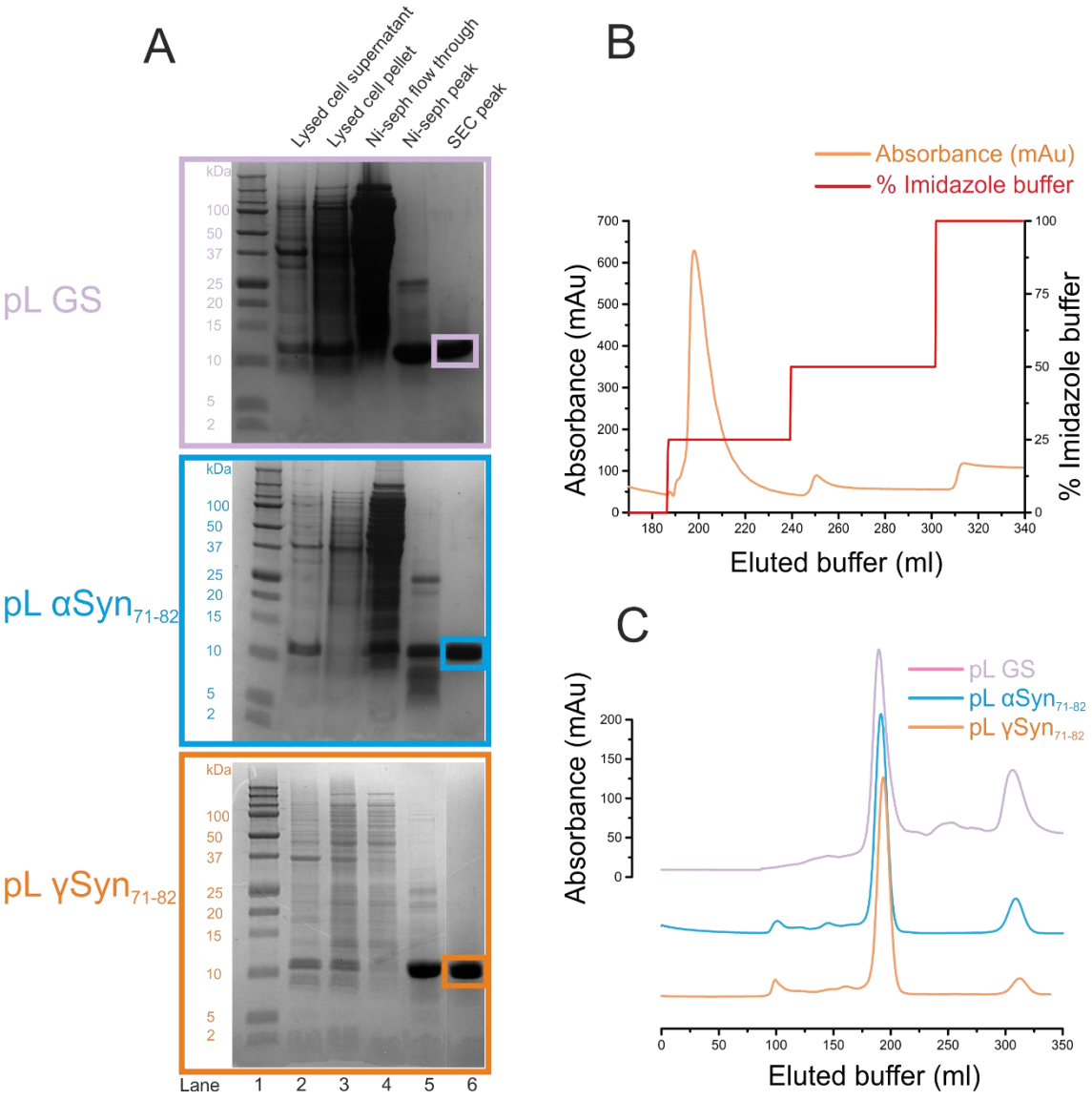


Figure 3-4. Overview of the purification process of pL variants used in this chapter. A) SDS-PAGE gels of the purification process of pL GS, pL αSyn_{71-82} and pL γSyn_{71-82} highlighted in pink, blue and orange respectively. The final purified protein after SEC is shown as an intense band without obvious impurities in the right hand lane (lane 6) of the gels highlighted with a coloured box (colour coded as above). B) Example of the initial chromatography step of affinity capture via Ni-sepharose (for pL γSyn71-82). The protein is eluted by a step-wise gradient of increasing imidazole concentration (Section 2.2.1.6). The main protein peak and the peak that was taken forward corresponded to 25% imidazole containing buffer (100%: 250mM imidazole). This protein fraction correspond to the fifth lane in A). C) The second purification step of size exclusion chromatography (SEC). The main peak fractions were taken, these correspond to the final lane in A).

3.3.3 Spectroscopic analysis of the protein fold

In order to ensure that pL constructs were natively, folded various spectroscopic analyses were carried out. The correct fold on the pL constructs was important as the mechanical strength of pL is dependent on the topology of the correctly folded protein.

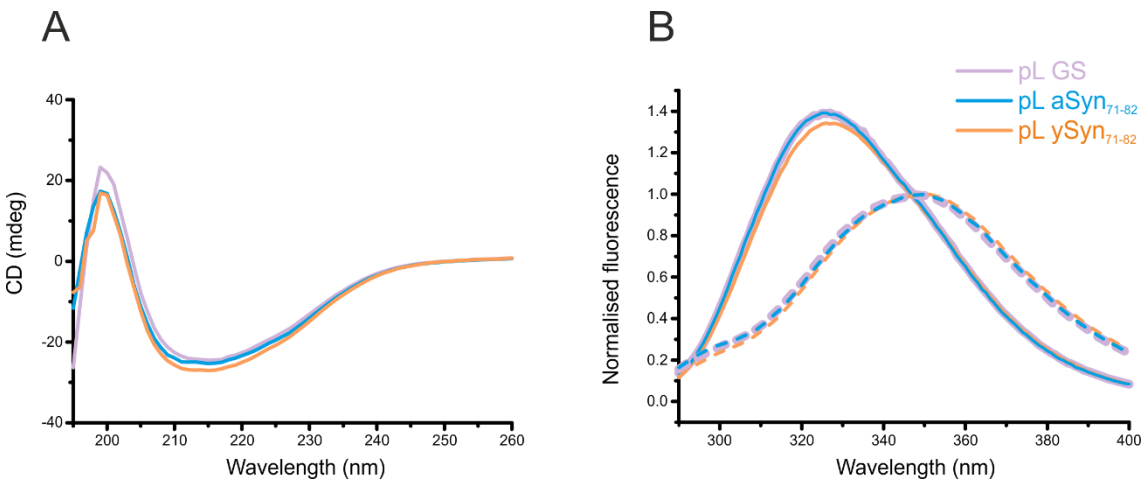


Figure 3-5. Far-UV CD spectra and intrinsic tryptophan emission spectra of pL variants. A) Far-UV CD spectra of pL variants used in this project analysing the secondary structural content of these constructs. The spectra of all 3 proteins are very similar shapes and intensities indicating similar secondary structural content. The spectra show characteristics of a folded protein with a broad minima from ~210-220 nm, most likely occurring from characteristic β-sheet contributions at ~215nm and α-helical contributions at ~209 and 222 nm consistent with a mixed α/β topology. B) Intrinsic tryptophan fluorescence of pL constructs under folded conditions (solid lines) and under unfolded conditions in the presence of 8M of the chaotropic agent urea (dashed lines). Pink, blue and orange spectra show pL GS, pL αSyn_{71-82} and pL γSyn_{71-82} respectively in both A and B.

In order to analyse whether pL variants maintain a secondary structure after purification, far-UV CD spectroscopy was carried out (Figure 3-5). The spectra of all 3 proteins are very similar shapes and intensities indicating similar secondary structural content. The spectra shows characteristics of a folded protein with a broad minima from ~210-220 nm most likely occurring from characteristic β-sheet contributions at ~215nm and α-helical contributions at ~209 and 222 nm consistent with a mixed α/β topology[356].

Tryptophan fluorescence emission spectra were also acquired to analyse whether engineered pL variants adopted a correctly folded tertiary structure. All pL constructs contain a single tryptophan (W47) that should be within the hydrophobic core of the protein when a native tertiary fold is adopted. Fluorescence emission spectra were recorded (290-400 nm) after excitation of tryptophan residues at 280 nm both in the presence or absence of 8 M urea (Figure 3-5). In the absence of urea, all pL constructs possess spectra typical of a folded

globular protein with a λ_{max} of 326 nm for pL αSyn_{71-82} and pL γSyn_{71-82} and 327 nm for pL GS, all of which possessed similar fluorescence intensities indicating a similar tertiary fold. The reduction in fluorescence intensity and a characteristic red shift (326-349 nm) of the λ_{max} in the presence of 8 M urea was indicative that the tryptophan was no longer packed into the hydrophobic core of pL constructs showing protein unfolding in these conditions. These spectrophotometric analyses of the proteins suggested that all pL variants are folded globular proteins with the expected secondary and tertiary structure from crystal structures of pL, and were similar to one another. Importantly, both pL αSyn_{71-82} and pL γSyn_{71-82} far-UV CD and intrinsic tryptophan fluorescence emission spectra are very similar to that of the non-aggregation control pL GS indicating that the aggregation-prone inserts are not perturbing the pL scaffold.

In order to determine whether the insertion of an amyloidogenic peptide into the pL scaffold protein perturbs the stability of the protein, far-UV CD was carried out as a function of temperature from 20-90 ^{o}C (Figure 3-6). Far-UV CD temperature gradient data was fitted to a two-state transition model in Global3 software (Applied Photophysics). The proteins possess typical spectra of a mixed α/β topology (Figure 3-5 A) at lower temperatures before unfolding and exhibiting spectra characteristic of a random coil at higher temperatures after a transition with a T_m of 45.5 and 46.6 ^{o}C for pL GS and pL αSyn_{71-82} respectively. Importantly, the stability of the protein variants are similar and the data shows that the insertion of an amyloidogenic peptide sequence does not significantly perturb the structure of the protein scaffold relative to a non-aggregation control (GS).

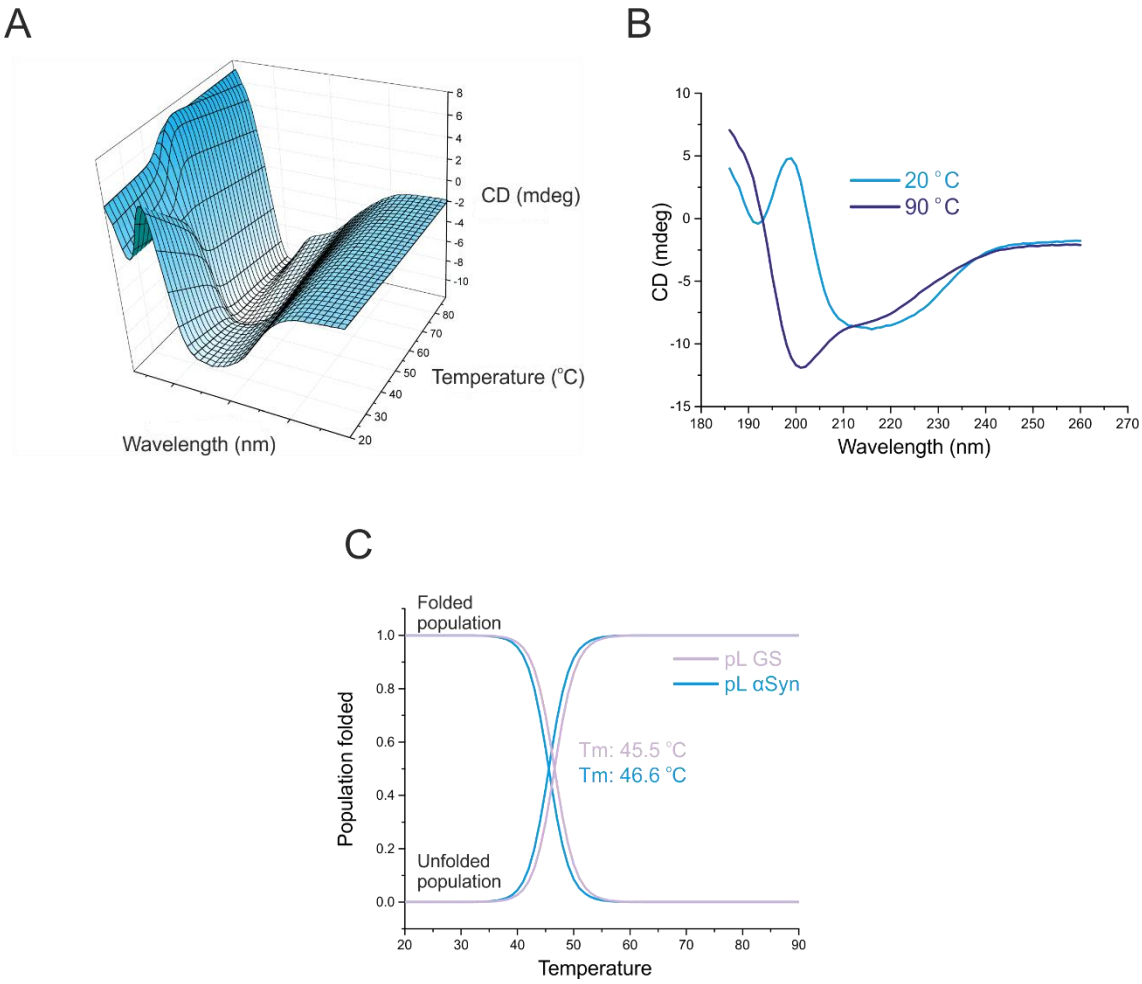


Figure 3-6. Far-UV CD temperature ramp of pL constructs. A) 3D representation of CD spectra as a function of temperature for pL αSyn_{71-82}. The spectra changes from a folded mixed α/β topology to that of a random coil at higher temperatures. B) Spectra of pL αSyn_{71-82} at 20 and 90 °C (light and dark blue, respectively) C) The relative populations of folded and unfolded conformations of both pL αSyn_{71-82} and pL GS (blue and pink respectively) calculated from the global analysis of CD thermal melt data (Global3 software from Applied Photophysics) the T_m values are shown inset.

3.3.4 NMR analysis of the chimeric protein display system

^{1}H-^{15}N HSQC spectra were taken of pL αSyn_{71-82} and pL GS constructs (Figure 3-7) in order to validate that the constructs were correctly folded and that inserting an amyloidogenic sequence doesn't perturb the structure. Both spectra show disperse, well defined peaks of similar intensity, characteristic of a folded protein. The majority of the peaks are identical with few chemical shift perturbations between the two constructs. CCPN software was used to automatically count peaks in the spectra, 83 main chain peaks were counted for pL αSyn_{71-82} and 78 for pL GS. There are likely to be less peaks visible for the pL GS variant because of overlap of the resonances arising from the GS linker region. Extra peaks in pL αSyn are likely attributable to the 71-82 αSyn insert region (VTGVTAVAQKTV). Moreover, most of the additional peaks present in pL αSyn_{71-82} occupy the unfolded region of the spectrum at ~8.0

ppm in the ^{1}H dimension suggesting that these peaks indeed arise from the αSyn_{71-82} insert and that they are present in an unfolded conformation within the loop region. The few chemical shift perturbations that are observed between the two constructs are also present in this region, suggesting that these differences in the two proteins arise from the residues that are close to, or are part of, the flexible loop region.

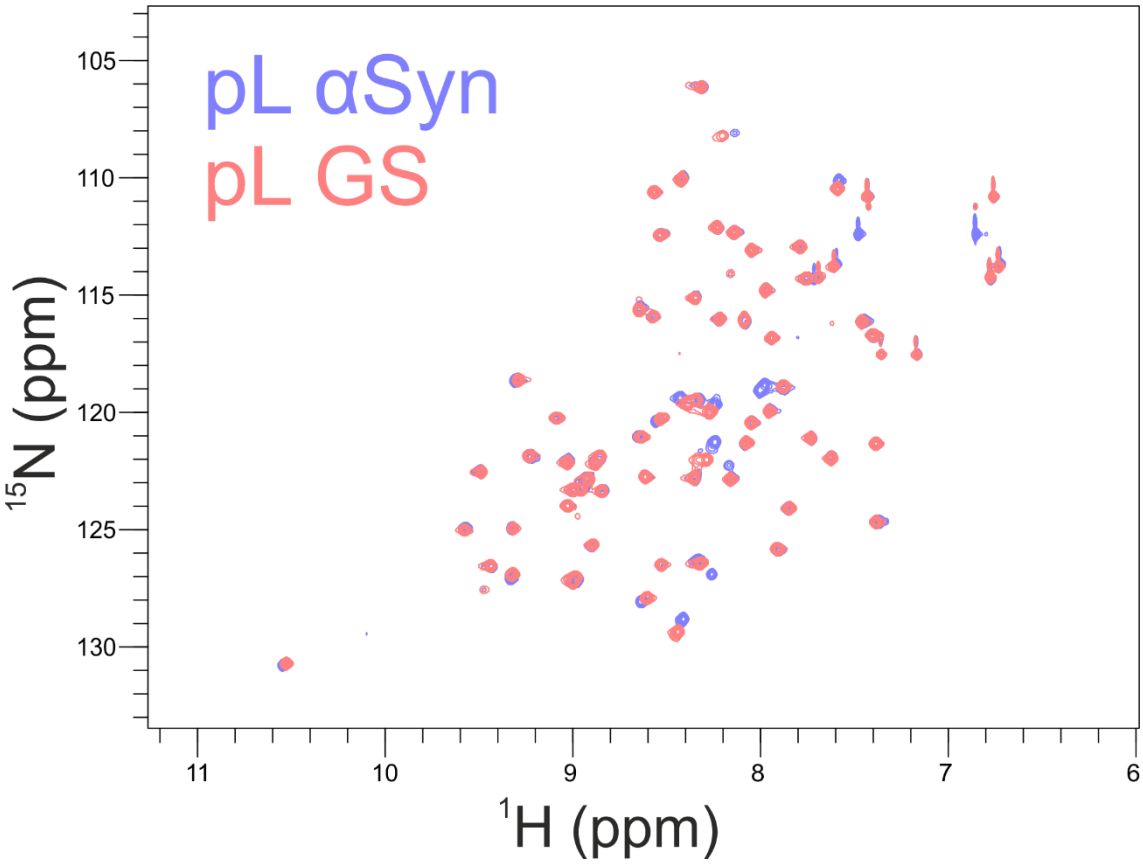


Figure 3-7. 600 MHz ^{1}H-^{15}N HSQC-NMR spectra of ^{15}N labelled pL αSyn_{71-82} and pL GS (labelled blue and pink respectively). Both spectra have dispersed and well defined peaks characteristic of folded proteins. Both constructs also have very similar spectra with few shifted peaks, indicating similar overall structure of the proteins.

Additional side chain peaks are present in the spectrum of pL αSyn_{71-82}, most likely due to the presence of amide side chains in the αSyn_{71-82} insert region from Q or K residues (Figure 3-8). The single tryptophan (W47) is visible at ≈ 10.5 ppm ^{1}H frequency (Figure 3-7) with identical chemical shift, suggesting that the hydrophobic core surrounding W47 is identical in both pL αSyn_{71-82} and pL GS. This also validates the intrinsic fluorescence data (Figure 3-5) indicating a similar chemical environment of the tryptophan in the core of the protein.

Overall, the spectroscopic and NMR data suggest the pL variants are folded and that inserting an amyloidogenic region does not perturb the structure of pL, which is important for utilising these display proteins in SMFS experiments.

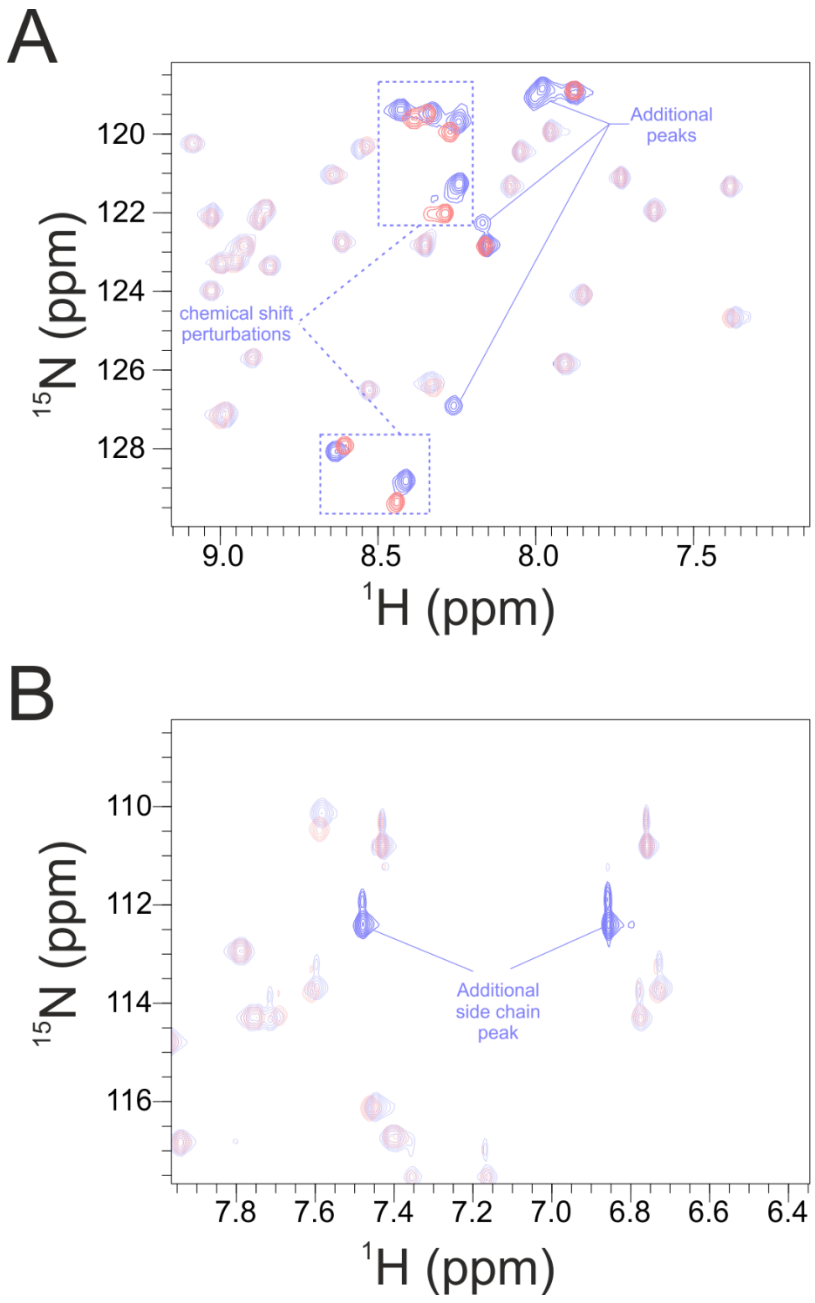


Figure 3-8. Zoomed region of ^{1}H-^{15}N HSQC spectra of pL αSyn$_{71-82}$ and pL GS (labelled blue and pink respectively). A) Expanded central region of the spectrum, extra peaks for pL αSyn$_{71-82}$ are present in the unfolded region of the spectum. A small number of residues also show chemical shift perturbations (blue dashed boxes) possible due to these residues being part of or close to the insertion loop. B) Additional side chains present in pL αSyn$_{71-82}$ most likely arising from amide containing side chains in the αSyn 71-82 region.

3.4 Dimerisation of the αSyn central NAC region detected by SMFS

Dimerisation interactions of αSyn_{71-82} were probed using AFM-based SMFS methods. The pulling velocity was kept constant at 1000 nms^{-1}. All SMFS experiments on pL constructs were carried out in 20 mM HEPES, pH 7.5. Thousands of force distance curves were analysed, and those which showed a characteristic single molecule "saw tooth" plot were fitted with the WLC model (Equation 1-1). From this model, L_C data were calculated. These data, along with F_R were plotted as both scatter plots and the corresponding contour plots (Figure 3-9). Correlated values of L_C and F_R visualised as "hotspots" in the contour plot (Figure 3-9 A) represents a distinct interaction in an inherently noisy experimental approach. The histograms in the third panel of Figure 3-9 indicate the distributions of L_C and F_R. The histograms in Figure 3-9 A show normal distributions, typical of a stochastic interaction, with Gaussian peak maxima at 22 ± 1 nm and 57 ± 1 pN for L_C and F_R, respectively. The estimated length of the pL molecule between Cys immobilisation point and the insert is 2.5 nm and the PEG linkers used are approximately 20 -25 nm each in length, giving a total contour length of 45 - 55nm. The experimental value differs from this true L_C value due to the geometry of the cantilever. The true L_C value should fall on the far right-hand side of the histogram of experimental L_C values, as previously discussed by this group[306]. Indeed, this is the case for the L_C of αSyn_{71-82} dimerisation (22.0 nm, Figure 3-9), confirming that a specific single molecule interaction at the expected L_C is being observed. A control experiment was carried out with pL GS. In the control variant, no specific dimerisation interaction is observed in the SMFS experiments (Figure 3-9 B), indicating that in this setup the interactions observed result from the presence of the αSyn_{71-82} insert, and not from the pL scaffold.

These results demonstrate dimerisation interactions of amyloidogenic regions can be measured at the single molecule scale using SMFS. Such interactions of small amyloidogenic peptides which would be extremely difficult to achieve via other methods.

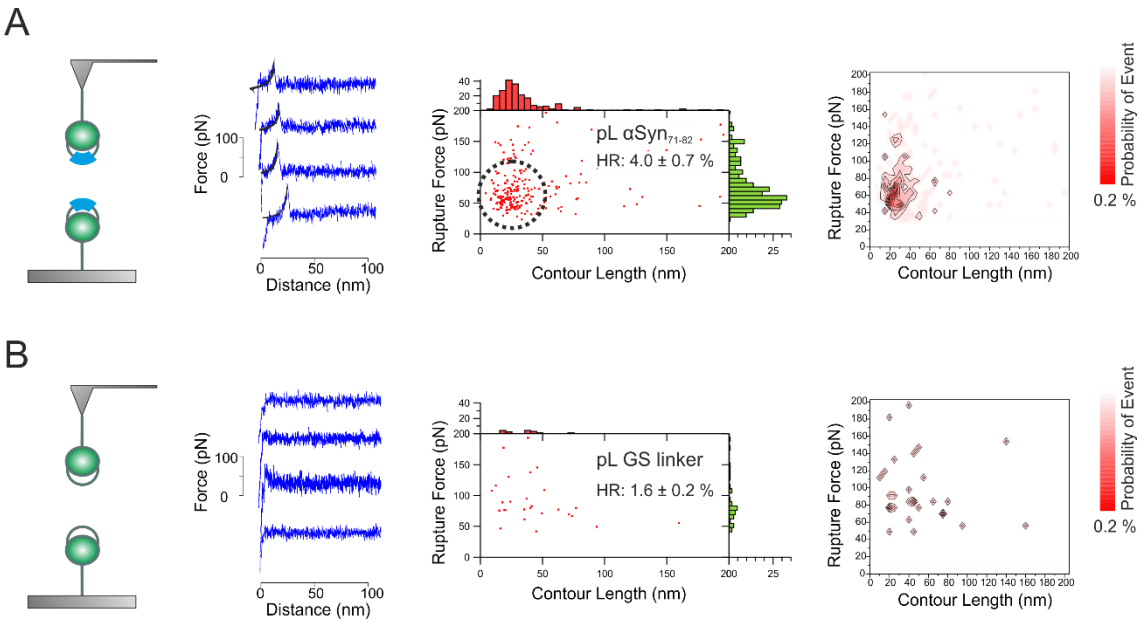


Figure 3-9. SMFS data for interactions between chimeric pL monomers. Raw data are shown alongside schematics on the left. WLC fits (black line) are shown on characteristic single molecule interactions. (A) The dimerisation interactions of αSyn_{71-82} shown as a scatter plot (middle) of individual data points with marginal histograms for L_C and F_R. The modal L_C and F_R were calculated to be 22 ± 1 nm and 57 ± 1 pN respectively. A "hotspot" of data is highlighted by a broken circle and shown in the corresponding contour map (far right). (B) Force events between the non-aggregation-prone pL GS shown as a scatter plot of individual data points with histograms. This data lacks the "hotspot" that is clearly present in A. The hit rate (HR) is denoted with SD between experiments. The total number of force-retract cycles are 8000 and 2500 for pL αSyn_{71-82} and pL GS dimerisation interactions, respectively.

3.5 DFS of the αSyn central NAC region

Despite the small size of the peptide fragment αSyn 71-82 under study here, the data demonstrate that this interaction is sufficiently strong to be detected by SMFS. In order to characterise this dimer interaction further, DFS was next carried out (Figure 3-10). These analyses allow parameters of the unbinding energy landscape to be calculated (as discussed in Section 1.6.3). The SMFS experiments were carried out at different pulling velocities ranging from 200 – 5000 nms^{-1}. Using the Bell-Evans model[308,309] (Equation 1-2) the dissociation rate constant in the absence of force (k_{off}^{0F}) and the distance along the free energy landscape of bound to transition state (x_u) can be calculated. The k_{off}^{0F} rate for the dimerisation interaction of αSyn_{71-82} using DFS was calculated to be 0.18 s^{-1}, which equates to a dimer lifetime in the range of seconds (5.79 s). Comparable k_{off}^{0F} rates, also on a seconds time scale, have been reported for the dimerisation of full length αSyn (4.00 s at pH 2.7 and 1.35 s at pH 3.7)[316]. The finding that the lifetimes of full length αSyn and αSyn_{71-82} dimers are of comparable magnitudes suggests that residues 71-82 play a key role in the stability of dimeric species

formed from the intact protein. Moreover, molecular dynamic studies on the dynamics of full length αSyn monomer reveal dramatic conformational changes on a nanosecond timescale[329] suggesting that the dimeric species of αSyn and $αSyn_{71-82}$ are relatively stable.

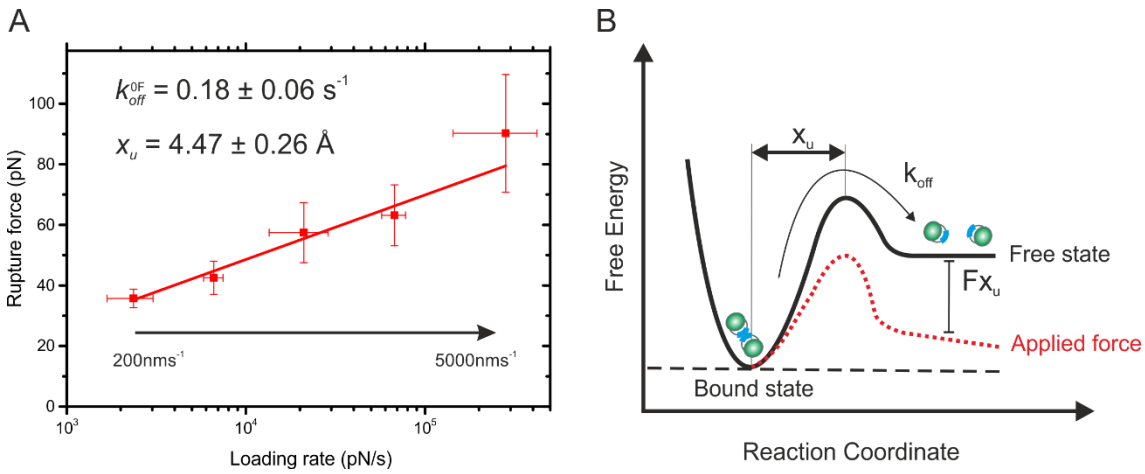


Figure 3-10. DFS speed dependence of $aSyn_{71-82}$ dimerisation. (A) Schematic representation of an unbinding energy landscape. The dotted line depicts the tilted energy landscape under force. x_U is depicted as the distance from the native energy well to the transition state. k_{off}^{0F} is shown as the stochastic process of crossing the transition energy barrier. The error bars shown are the standard deviation between the mean of the triplicate data. The errors of the k_{off}^{0F} and x_U were calculated by manual bootstrapping. (B) Dynamic force spectrum from the dimerisation interaction of $αSyn_{71-82}$ fitted with a linear fit, used to extrapolate the values k_{off}^{0F} and x_U using the Bell-Evans model[308,309].

3.6 Novel interaction between $αSyn_{71-82}$ and its homologue $γSyn_{71-82}$

αSyn is one of a small family of highly homologous synuclein proteins (a sequence alignment is shown in Figure 1-14). Other members of this family include βSyn and γSyn, (αSyn homologues are discussed in Section 1.5.5). βSyn lacks the central 71-82 sequence of αSyn essential for aggregation[204] and shown to form the core of fibrillary structures[100]. βSyn protein is not found in pathogenic LBs *in vivo*[273], nor does it form fibrils in vitro[274,275], providing further evidence that this sequence is key to the aggregation process of αSyn. Unlike βSyn, the 127-amino acid γSyn protein, contains a central NAC domain (residues 71-82). The central NAC region differs in sequence between αSyn ($_{71}$VTGVTAVAQKTV$_{82}$) and γSyn ($_{71}$VSSVNTVATKTV$_{82}$) (Figure 3-2). A chimeric pL displaying $γSyn_{71-82}$, therefore, was engineered for SMFS studies as discussed above in order to investigate relative strength of the dimerisation interaction and how this might relate to aggregation. The scatter and contour plots in Figure 3-11 show heterodimerisation of $γSyn_{71-82}$ and $αSyn_{71-82}$, as well as the homodimerisation of $γSyn_{71-82}$.

These data reveal that a heterodimerisation event between γSyn$_{71-82}$ and αSyn$_{71-82}$ detected using SMFS is similar to that of the αSyn$_{71-82}$ homodimer. The homodimerisation interaction of γSyn$_{71-82}$, however, has a significantly lower hit rate, with the data lacking a 'hot spot' indicating that under this experimental set up, the affinity of γSyn$_{71-82}$ to homodimerise is lower than that of αSyn$_{71-82}$:αSyn$_{71-82}$ or indeed αSyn$_{71-82}$:γSyn$_{71-82}$. This is the first observation of a heterodimeric interaction in an aggregating system by SMFS and also the first observation that the central NAC region of αSyn, important for the aggregation of the full length protein, interacts with the corresponding sequence of a homologue. In fact, the αSyn$_{71-82}$:γSyn$_{71-82}$ heterodimer has comparable L_C and F_R values (19 ± 1 nm and 53 ± 1 pN, respectively) to that of the αSyn$_{71-82}$ homodimeric interaction (22 ± 1 nm and 57 ± 1 pN, respectively). Uversky and colleagues have previously shown that full length αSyn fibrillation is inhibited by the presence of γSyn[275]. One possible interpretation is that the inhibitory effect reported is mediated by the interaction between the central NAC regions of these proteins.

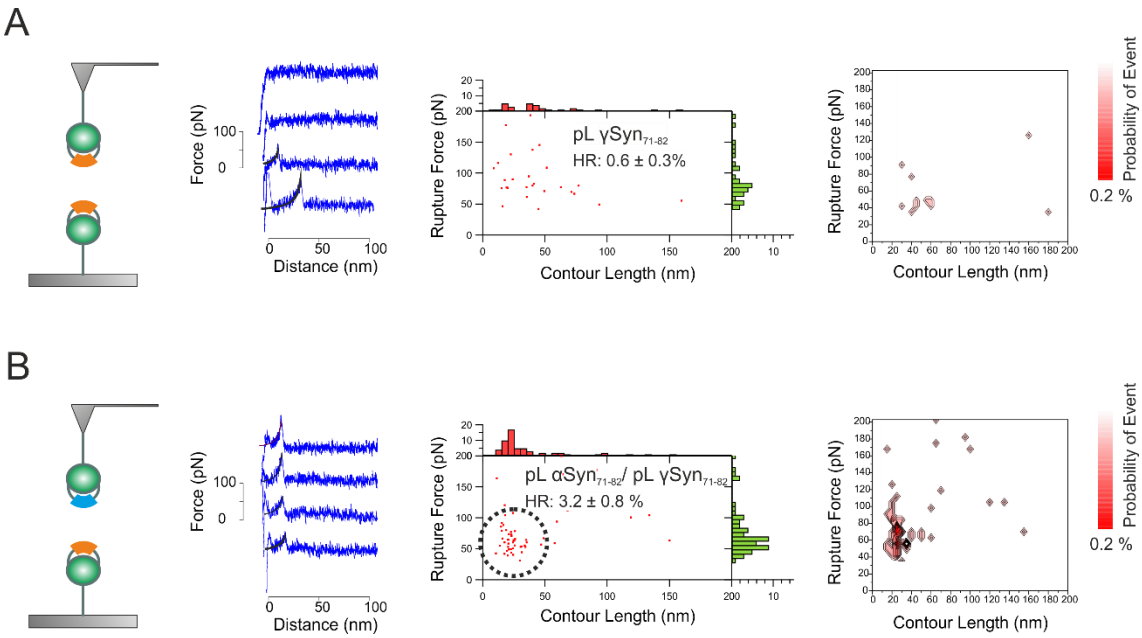


Figure 3-11. SMFS data for the dimerisation interactions of the γSyn$_{71-82}$ homodimer (A) and the γSyn$_{71-82}$/αSyn$_{71-82}$ heterodimer (B). Raw data are shown alongside schematics on the left. WLC fits (black line) are shown on characteristic single molecule interactions. The interactions between αSyn$_{71-82}$ and γSyn$_{71-82}$ are shown as a scatter plot of individual data points with marginal histograms for L_C and F_R. The modal L_C and F_R were calculated to be 19 ± 1 nm and 53 ± 1 pN respectively. A "hotspot" of data is highlighted by a broken circle and shown in the corresponding contour map. (B) The dimerisation interactions of γSyn$_{71-82}$ shown as a scatter plot of individual data points with histograms for L_C and F_R. These data lack the "hotspot" that is clearly present in A. The hit rate (HR) is denoted with SD between experiments. The total number of approach retract cycles for pL γSyn$_{71-82}$ homodimerisation and αSyn$_{71-82}$/ γSyn$_{71-82}$ heterodimerisation interactions are 2000 and 2000 respectively.

A SMFS experiment was carried out as highlighted in Figure 3-12 in order to help validate the observation of heterodimerisation events and that the apparent absence of a hotspot for γSyn$_{71-82}$ homodimeric interaction was not simply due to a problem with the experimental set up such as insufficient functionalisation of cantilever tips and surfaces. In the experimental setup, a cantilever tip and a silicon surface were functionalised with pL γSyn$_{71-82}$. SMFS experiments were then carried out carried. A hotspot of data was not observed in this experiment (Figure 3-12, as above). A new silicon surface was then functionalised with pL αSyn$_{71-82}$ and SMFS experiments were carried out. The data revealed that the αSyn$_{71-82}$:γSyn$_{71-82}$ heterodimer formed observable interactions. Changing back to the original pL γSyn$_{71-82}$ functionalised the surface revealed that a hot spot of events once again disappears.

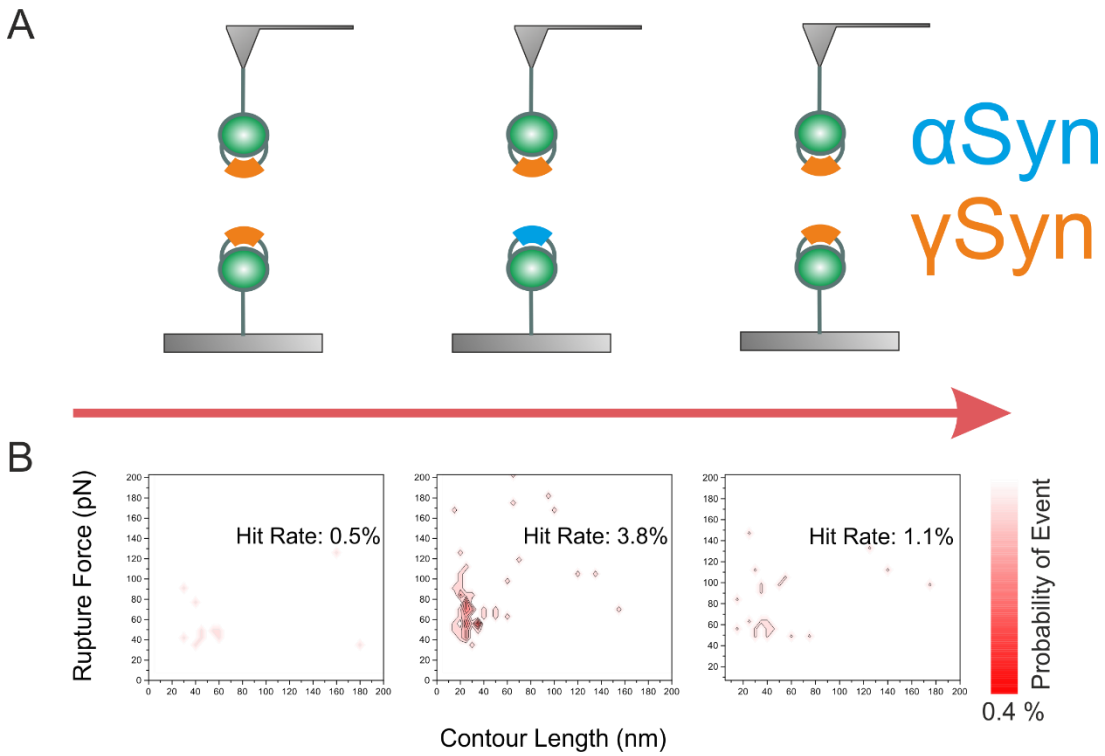


Figure 3-12. SMFS experiment showing appearance and subsequent disappearance of heterogeneous dimer with pL chimeric constructs immobilised on cantilever tips and surfaces. αSyn$_{71-82}$ and γSyn$_{71-82}$ inserts are shown in blue and orange, respectively. A) Schematic of experimental setup γSyn$_{71-82}$ homodimerisation was probed, followed by changing the functionalised surface to probe αSyn$_{71-82}$: γSyn$_{71-82}$ heterogeneous dimerisation; and subsequently the surface was then changed back to the original pL γSyn$_{71-82}$ functionalised surface in order to probe γSyn$_{71-82}$ homodimer interactions. B) Contour plots from the experimental set up described in (A) in which few interactions are observed for γSyn$_{71-82}$ homodimerisation, followed by an appearance of a αSyn$_{71-82}$: γSyn$_{71-82}$ heterogeneous dimerisation interaction when the surface is swapped, before a disappearance of the interaction when the surface is swapped back. These experiments were carried out in the same experimental set up in order to validate that the absence of a γSyn$_{71-82}$ homodimerisation interaction is real and not simply due to insufficient surface functionalisation.

3.7 Native Mass Spectrometry of pL constructs

Having observed the dimerisation interactions of the αSyn_{71-82} and γSyn_{71-82} sequences using SMFS, an orthogonal technique, ESI-IMS-MS, was next applied to the samples to validate that the dimerisation events observed using SMFS were a result of a specific interaction between the peptide inserts, and could not the pL scaffold. This soft ionisation method enables preservation of non-covalent interactions, and protein quaternary structure, in the gas phase[357]. ESI-IMS mass spectra were taken at time point t= 0 hours, and again after 4 hours (Figure 3-13). At t= 0 hours, all samples, including each pL peptide variant in isolation and a 1:1 mixture of pL αSyn_{71-82} and pL γSyn_{71-82} , appeared to be purely monomeric. After 4 hours, all samples, with the exception of the non-aggregation pL GS linker construct, contained oligomeric species (Figure 3-13 B). Both pL αSyn_{71-82} alone and a 1:1 mixture of pL αSyn_{71-82} and pL γSyn_{71-82} show monomer through to trimer, whereas the pL γSyn_{71-82} alone contains only monomer and dimer. This indicates, as in the SMFS, a lower self-association propensity for the γSyn_{71-82} relative to that of αSyn_{71-82} or the co-assembly of the two homologs, αSyn_{71-82} and γSyn_{71-82}.

Closer analysis of the dimeric species observed in the ESI-mass spectra after 4 hours (Figure 3-14) also confirmed that in the mixed samples, a heterodimer is observed alongside homodimers of pL αSyn_{71-82} and pL γSyn_{71-82}, corroborating the observation of a heterodimer in SMFS experiments. This is evidenced in the bottom panel of Figure 3-14 A, with a difference in mass between the pL αSyn_{71-82} and pL γSyn_{71-82} constructs (10.391 and 10.423 kDa, respectively), distinct peaks for both pL αSyn_{71-82} and pL γSyn_{71-82} homodimers, plus heterodimers can be identified in the same spectra. Moreover, these peaks show a higher intensity for the pL αSyn_{71-82} homodimer, followed by a slightly lower intensity for the pL αSyn_{71-82}: pL γSyn_{71-82} heterodimer, with a significantly reduced, lower intensity for the pL γSyn_{71-82} homodimer. This order of apparent affinity is the same as in the SMFS further corroborating the relative order of affinities (from highest to lowest apparent dimer affinity: $\alpha\alpha > \alpha\gamma > \gamma\gamma$).

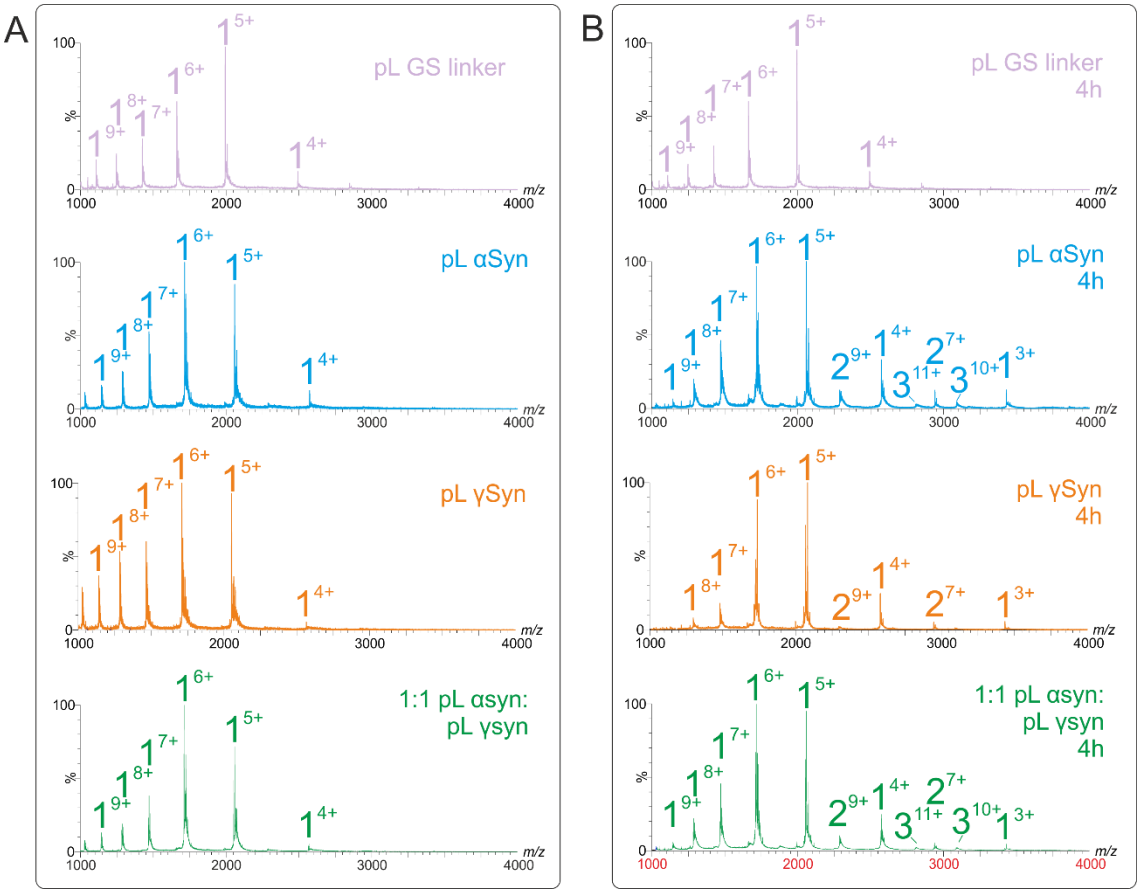


Figure 3-13. Native ESI-mass spectra of pL constructs immediately after dilution (t = 0) and after 4 hours. A) and B) Spectra at time point 0 and after 4 hours, respectively. Pink, blue, orange and green show pL GS linker, pL $\alpha Syn_{71\text{-}82}$, pL $\gamma Syn_{71\text{-}82}$ and 1:1 mix of pL $\alpha Syn_{71\text{-}82}$ and pL $\gamma Syn_{71\text{-}82}$ constructs, respectively. The numbers above the peaks denote the oligomer order, with the positive-charge state of ions in superscript. All variants in isolation and a 1:1 heterogeneous mixture of α and γ were present as monomer (left side) at t = 0. After 4 hours (right side), all variants except the non-aggregation pL GS linker construct showed self-aggregation. All pL constructs were diluted to a final concentration of 100 µM in 100 mM ammonium acetate buffer, pH 6.8. MS analysis carried out by Dr Lydia Young (University of Leeds).

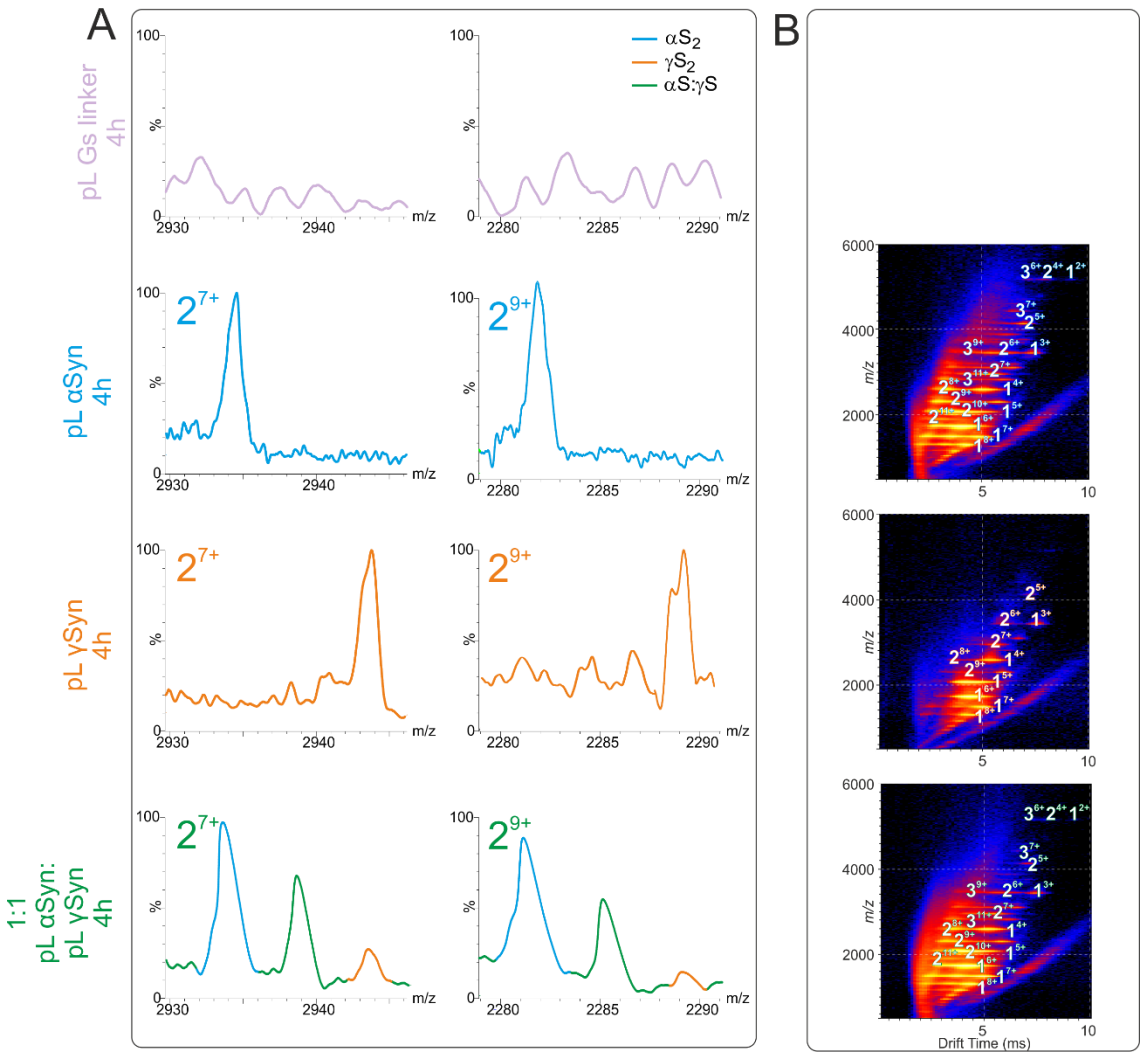


Figure 3-14. Zoomed ID ESI-native mass spectra of dimeric pL constructs and 3D ESI-IMS-MS of pL constructs after 4h incubation. A) Zoomed ESI-mass spectra of two different charge states (7+ and 9+) of dimeric pL. Pink, blue, orange and green indicate pL GS, pL αSyn_{71-82}, pL γSyn_{71-82} and mixed 1:1 pL αSyn_{71-82} and pL γSyn_{71-82} respectively. The numbers above the peaks denote the oligomer order, with the positive-charge state of ions in superscript. The spectra of pL GS shows the absence of dimer after 4h. The difference in mass between pL αSyn_{71-82} and pL γSyn_{71-82} allows for pL αSyn_{71-82} and pL γSyn_{71-82} homodimers, plus heterodimeric species to be discerned from one another. In order of relative intensities there is the most pL αSyn_{71-82} homodimer, followed by αSyn_{71-82}: pL γSyn_{71-82} heterodimer and then least amount of pL γSyn_{71-82} homodimer in the mixed α/γ spectra. B) ESI–IMS–MS Driftscope plots of the pL variants showing IMS drift time versus mass/charge (m/z) versus intensity (z, square-root scale). Both pL αSyn_{71-82} and αSyn_{71-82}: pL γSyn_{71-82} mix show monomer (1) up to trimer (3), γ only forms monomer and dimer (2) species. MS analysis carried out by Dr Lydia Young (University of Leeds).

3.8 Characterising the aggregation of pL constructs

Whether the chimeric constructs were still able to aggregate when aggregation-prone peptides were engineered into the pL scaffold protein was next investigated using ThT fluorescence. ThT fibril growth experiments (Figure 3-15) confirmed that both pL αSyn_{71-81} and pL γSyn_{71-82} show

an increase in ThT fluorescence. However, the mixed sample appeared unable to aggregate as studied by this assay. The reduced aggregation observed in the pL αSyn_{71-81}: pL γSyn_{71-82} mixed incubation is dependent of either pL αSyn_{71-81} or pL γSyn_{71-82} aggregation in isolation. This suggests that the homodimeric interaction previously identified by SMFS and MS, inhibits further aggregation. This supports the finding that the inhibitory effect reported via full length γSyn on the aggregation of αSyn is mediated by the interaction between the central NAC regions of both these proteins. It is important to note that the observed pL αSyn_{71-81} homodimeric interaction and the pL αSyn_{71-81}: pL γSyn_{71-82} heterodimeric interaction by SMFS appear identical.

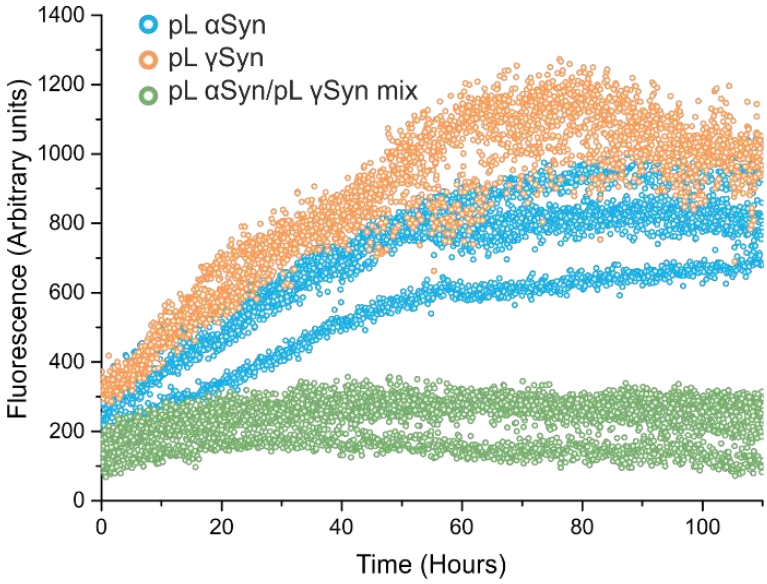


Figure 3-15. ThT fluorescence assay of pL variants. Incubations were carried out at 37°C, 600 rpm. Blue, orange and green data points show the normalised fluorescence over time of pL αSyn_{71-82}, pL γSyn_{71-82} and the 1:1 mix of pL αSyn_{71-82} and pL γSyn_{71-82}. Proteins are at a final concentration of 100 µM (α/γ mix at 50µM of each protein). The mixed α/γ incubation shows reduced aggregation that is independent of either α or γ in isolation.

3.9 Native Mass Spectrometry of synuclein peptides used in force studies

ESI-IMS-MS was also used to study the self-assembly properties of synthetic peptides equivalent to the central NAC (residues 71-82) in isolation i.e. not in the context of the pL scaffold to further validate that the interactions observed in SMFS. This analysis also allowed determination of the role of the pL scaffold in perturbing or encouraging oligomer formation of the peptide inserts. ESI-mass spectra of all the samples (αSyn_{71-82} and γSyn_{71-82} and a 1:1 mixture of the two) confirm self-association of the peptides, with oligomers forming up to pentamer, with the 1:1 heterogenic mix showing a random distribution between number of α

and γ monomers in the oligomers (Figure 3-16). ESI-IMS-MS Driftscope plots of the peptides alone, and the 1:1 mixture show monomer (1) through to the tetradecamer (14) present in all of the samples. These results confirm that the peptides alone without pL as a display scaffold are able to interact and indeed assemble into higher order oligomers. The fact that oligomers up to, and including, tetradecamer are observed in the absence of the scaffold, compared with a maximum order of trimers when engineered in chimeric pL (Figure 3-13), suggests that the pL scaffold disfavours larger oligomer formation of the peptides, likely a result of steric hindrance. The data again also confirm that αSyn_{71-82} and γSyn_{71-82} sequences directly interact forming hetero-oligomeric constructs (Figure 3-16 bottom panel).

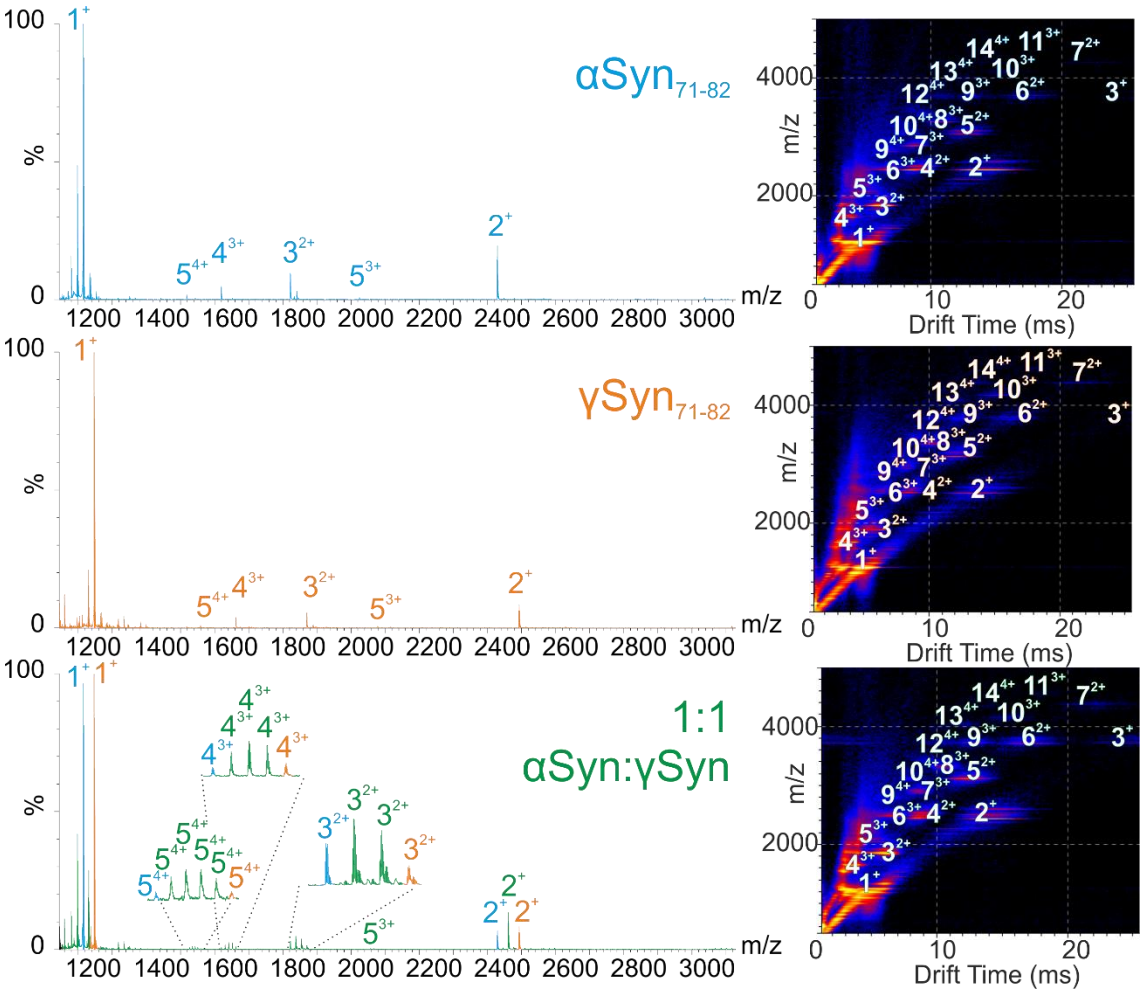


Figure 3-16. ESI-mass spectra and 3D ESI-IMS-MS of αSyn_{71-82} and γSyn_{71-82} peptides. Top, middle and bottom spectra and corresponding IMS-MS show spectra of αSyn_{71-82}, γSyn_{71-82} and a 1:1 mix of αSyn_{71-82} to γSyn_{71-82} respectively. The numbers above the peaks denote the oligomer order, with the positive-charge state of ions in superscript. All mass spectra (left hand side) confirm self-association of the peptides up to pentamer, with the 1:1 mix of each peptide showing a random distribution between number of αSyn_{71-82} and γSyn_{71-82} monomers in the oligomers. ESI–IMS–MS Driftscope plots of the peptides alone, and the 1:1 mixture (right hand side) show monomer (1) through to the tetradecamer (14) present in all of the samples, two minutes after diluting the monomer to a final peptide concentration of 100 µM in 100mM acetate buffer pH 6.8. ESI–IMS–MS Driftscope plots show the IMS drift time versus mass/charge (m/z) versus intensity (z, square-root scale). MS analysis carried out by Dr Lydia Young (University of Leeds).

3.10 Characterising the aggregation of synuclein peptides used in force studies

The aggregation of the peptides αSyn_{71-82} and γSyn_{71-82} were also monitored in fibril growth assays. Both αSyn_{71-82} and γSyn_{71-82} peptides form ThT-positive amyloid fibrils with a comparable lag-time (Figure 3-17). However, a 1:1 mixture of αSyn_{71-82} and γSyn_{71-82} resulted in a significant increase in lag-time, compared to the peptides in isolation. Transmission electron

microscopy (TEM) images show that fibrillar species were present in all samples (αSyn_{71-82} and γSyn_{71-82} and a 1:1 mixture of the two) after a 90 h incubation. Indeed, the mixture assembles at a different rate and results in a distinct fibril morphology with respect to each of the peptides alone. This aggregation behaviour of the peptides in isolation mirrors that of the chimeric display proteins shown in Section 3.8, both these analyses show that the heterodimeric interactions retard the rate of aggregation.

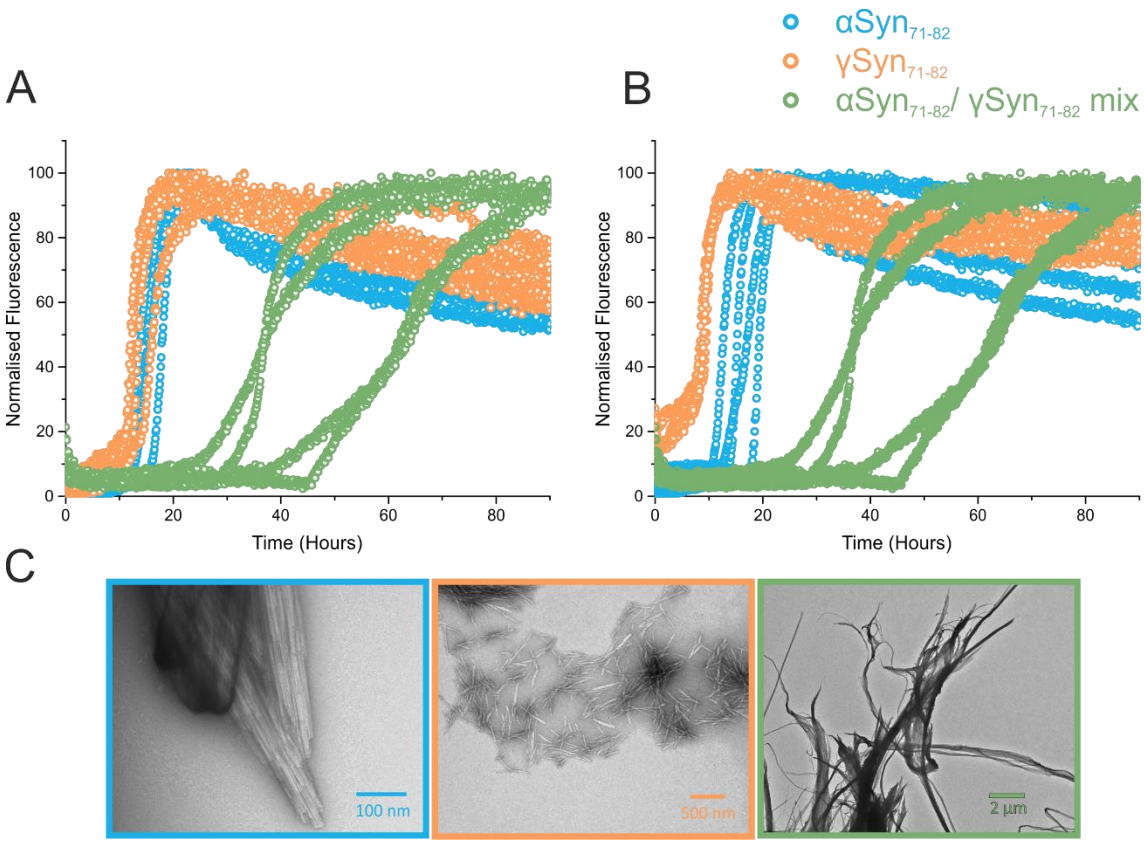


Figure 3-17. ThT fluorescence assay of αSyn_{71-82} and γSyn_{71-82} peptides. A) and B) Blue, orange and green data points show the normalised fluorescence signal over time of αSyn_{71-82}, γSyn_{71-82} and the 1:1 mix of αSyn_{71-82} and γSyn_{71-82} peptides, respectively. A) Both αSyn_{71-82} and γSyn_{71-82} are at 225 µM final concentration, the αSyn_{71-82} + γSyn_{71-82} mix is at 225µM of both peptides: 450 µM final concentration. B) Both αSyn_{71-82} and γSyn_{71-82} are at 450 µM final concentration, the αSyn_{71-82} + γSyn_{71-82} mix is the same data as presented in A) for comparison at 450 µM final concentration. The αSyn_{71-82}: γSyn_{71-82} mix showed retarded rates of aggregation with an increased lag time in respect to αSyn_{71-82} and γSyn_{71-82}. The data shows that this cannot simply be attributed to either α or γSyn_{71-82} in isolation, or indeed an artefact of concentration. C) The corresponding TEM images were taken at the end points of the incubations and are colour coded in the same manner as above. All three incubations form fibrillary structures, the morphology of the mixed incubation is independent of that of either αSyn_{71-82} or γSyn_{71-82}.

3.11 Discussion

The results in this chapter have demonstrated and validated a novel technique for use in SMFS experiments investigating the dimerisation of amyloidogenic proteins at the single molecule scale. Using a chimeric protein display system for the analysis of these interactions, we were able to interrogate specific interactions and minimise the contributions of non-specific events that often occur in the surface-proximal regions of force-distance curves. This technique also restricts the application of force to a defined geometry, thus giving more confidence that specific dimer interactions are being interrogated in a consistent manner. Importantly, we have validated that the structure and stability of chimeric display proteins is not perturbed by the insertion of amyloidogenic sequences. This is significant as the structure of the chimeric pL constructs is integral to the ability to display protein sequences.

By implementing this technique, we have been able to reveal novel results regarding the interactions of αSyn and its homologue γSyn. We probed specific single molecule dimerisation events between the central NAC regions (residues 71-82), integral to the aggregation of αSyn[100,204]. Understanding dimerisation events is crucial in our quest to shed light upon the aggregation process as a whole, given that dimer formation is the first stage in the aggregation cascade, the toxic species of which remains elusive. DFS of αSyn_{71-82} dimerisation revealed novel insights into the nature of the αSyn_{71-82} dimerisation event including its relatively long lifetime (in the range of seconds) consistent with previous force data on full length αSyn[316]. It is important to note, however, that previous SMFS experiments on full length αSyn were conducted at acidic pH where the full length protein is closer to its pI and therefore more prone to aggregate[315-320]. The experiments carried out here were conducted at a more physiological pH (pH 7.5) and are therefore more applicable to the conditions in which the protein self-associates *in vivo*. The fact that the dimerisation lifetime of αSyn_{71-82} in this study was higher than that determined for the full length protein previously[316] suggests that this highly hydrophobic region of αSyn dimerises with higher affinity than that of full length protein. These data provide further support for the central role of this central NAC sequence in the aggregation of the full length protein. These data may suggest, therefore, a regulatory role for sequences outside of NAC in full length αSyn.

A novel interaction between αSyn_{71-82} and its homologue sequence, γSyn_{71-82}, was also identified using SMFS and validated using native mass spectroscopy techniques, revealing similar L_C and F_R values to the αSyn_{71-82} homodimerisation interaction and similar oligomers detected by ESI-MS. We have shown that this interaction inhibits aggregation using both

chimeric pL constructs and peptides equivalent to the central NAC region in isolation, which might suggest that the mechanism by which full length γSyn inhibits the aggregation of αSyn[275] may be mediated by the central NAC region of the proteins. Recently, Baum and colleagues have proposed a mechanism in which βSyn inhibits the aggregation of αSyn[280] that is dependent on the higher affinity interaction of the heterodimeric interactions relative to the homodimeric interactions. Thus βSyn inhibits the self-association of αSyn via kinetic competition. Here, SMFS shows that the affinity of the αSyn_{71-82} homodimer and the αSyn_{71-82}: γSyn_{71-82} heterodimer are very similar, at least revealed via SMFS and MS methods, and much more avid than the γSyn_{71-82} homodimer. This points to a similar mechanism of kinetic competition as the bases of aggregation inhibition. It can be postulated that heterogeneous interactions between α- and γSyn occur as normal physiological complexes, and that these interactions may function in chaperoning αSyn away from aggregation under normal cellular conditions. Indeed, both α- and γSyn are highly expressed in many of the same cells of the brain[273].

4 Interrogation of the dimerisation events of α-Synuclein

4.1 Abstract

In this chapter, the utility of SMFS has been demonstrated in the single molecule study of αSyn dimerisation. It has been shown that these experiments can identify single molecule dimerisation interactions and that these data can be used to probe the conformation and/or interaction regions of αSyn dimers. Biophysical analysis using CD and SEC-SAXS showed that although αSyn and E46K αSyn PD early onset PD familial mutant showed similar structural and conformational characteristics, there are subtle differences which may play a key role in the relative aggregation rates and their role in disease. Indeed, it was shown that the charge substitution E46K, decreases the lag time for fibril formation compared to WT αSyn. This variant also displays different L_C values in SMFS experiments, suggesting differences in the dimerisation interaction between the variants. The interaction measured using SMFS may therefore be a key step in the aggregation pathway that imparts differential aggregation propensity. SMFS studies indicated that the interaction of β- and γSyn which are non-aggregation-prone, non-PD causing, human homologues of αSyn, form similar dimerisation interactions. It was therefore hypothesised that the interactions observed for αSyn in these conditions are not linked to the aggregation pathway.

4.2 Introduction

As discussed previously in this thesis, studying protein aggregation is a significant challenge, especially the important early events in the aggregation pathway. We have therefore utilised SMFS in this chapter in order to interrogate the homodimeric interactions of full length αSyn and its variants. Similar experiments have been carried out previously by Lyubchenko and colleagues[316,318-320] which has yielded information about the dimeric interactions of αSyn discussed in Section 1.6.4.2. These studies have shown that dimerisation events physiological-like conditions are not observed unless the the solution conditions are perturbed by acidification[315,316], addition of metal cations (Zn^{2+} and Al^{2+})[318] or addition of the polyamine spermidine[319]. In these studies, however, the aggregation/ fibril formation rates of the proteins are not analysed. Additionally, the effects of linkers or tip geometry on L_C[306], as discussed later in this chapter, are completely neglected.

In this chapter, we have demonstrated the application of simulating L_C distributions and comparing these with experimentally observed values in order to gain insight into the

interacting regions of the proteins. We have also been able to observe and analyse interactions at close to physiological conditions that have not been reported previously in SMFS studies.

4.3 Engineering recombinant αSyn

4.3.1 Molecular biology

pET23a plasmids encoding αSyn were kindly provided by Prof Jean Baum (Department of Chemistry and Chemical Biology, Rutgers University, NJ, USA). Variants containing a single engineered Cys residue at various positions were engineered to allow for thiol chemical immobilisation in SMFS experiments. Mutagenesis was carried out using the New England Biolabs Q5© mutagenesis protocol (Section 2.2.1.2). In this chapter, the cysteine was inserted at the C-terminal residue (A140C), a region of the protein which is highly disordered and not proposed to associate in the aggregation of αSyn. PD E46K and A30P mutants as well as synuclein homologues, β- and γSyn were generated using the same method.

4.3.2 Protein expression and purification

αSyn and all variants and homologues discussed in this thesis were expressed in the same manner, the purification of WT and familial PD mutated αSyn are shown as examples in Figure 4-1, this however is not extensive of all the constructs used in this thesis. αSyn expressing plasmids were transformed in *E. coli* BL21 (DE3) cells and protein expression was carried out as described in Section 2.2.1.7. The purification was carried out by performing an ammonium sulphate precipitation in which αSyn was salted out of solution with a high concentrations of ammonium sulphate before being washed and resolubilised in 25 mM Tris, pH 8.0. This protein solution was then subjected to anion exchange (AEX) chromatography followed by size exclusion chromatography (SEC). The purification process is shown in Figure 4-1. The presence of αSyn (~15 kDa) was confirmed by SDS-PAGE (Figure 4-1 A). The purity of αSyn and other synuclein variants was confirmed by SDS-PAGE and MS.

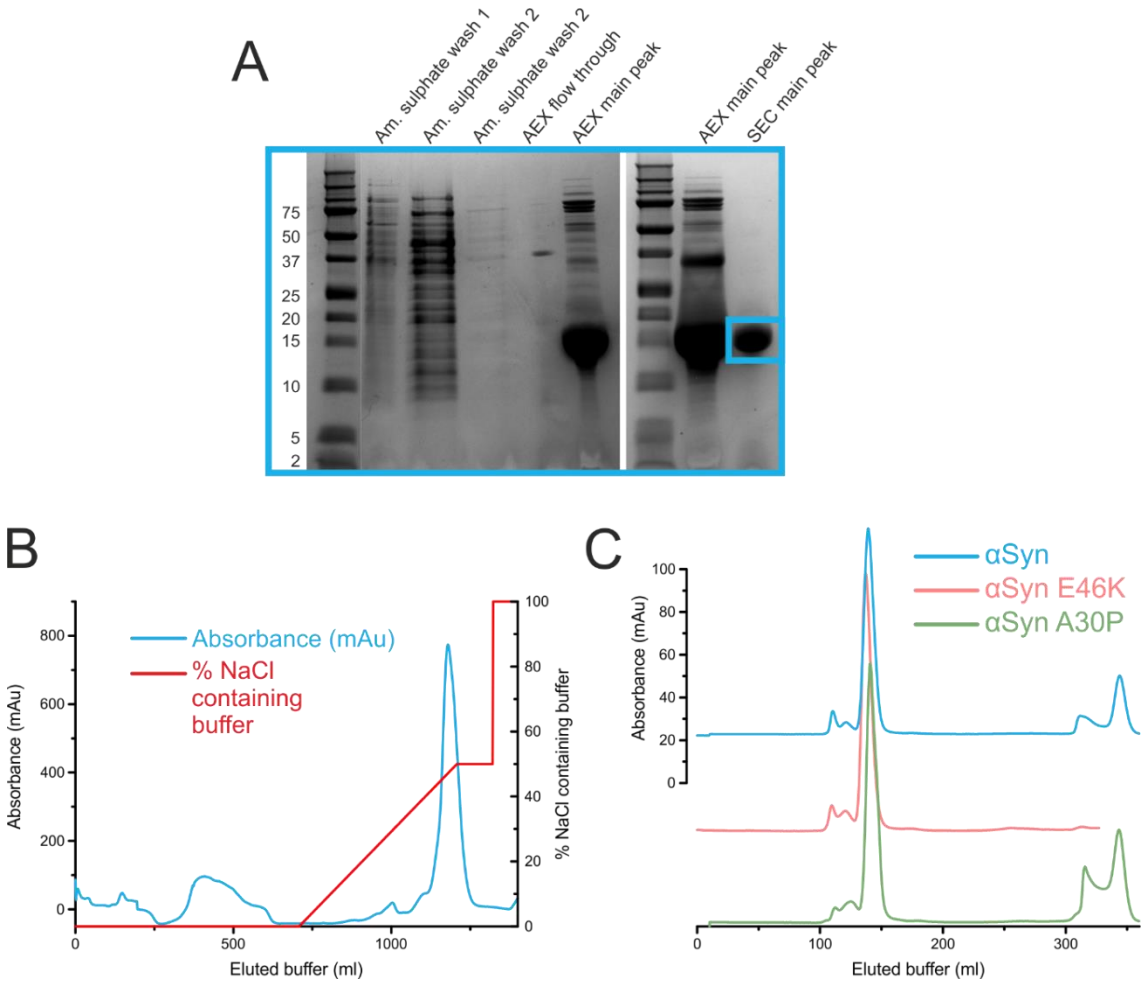


Figure 4-1. Overview of the purification process of selected number of full length αSyn variants. A) SDS-PAGE gel highlighting steps in the purification process. The final purified protein after SEC is shown as an intense band without obvious impurities in the right most lane of the gels highlighted with a blue box. B) Example of the initial affinity capture chromatography step: AEX (anion exchange). The protein was eluted via a continuous gradient of 0-50 % 1 M NaCl containing buffer (red trace) over 500 ml. The main peak was taken forward for further purification. C) The second chromatographic step of SEC. The main peak fractions were taken, these correspond to the final lane in A).

4.4 Biophysical characterisation of αSyn

αSyn is generally considered to exist as a monomeric IDP[175] but, it cannot simply be described as a random coil[210,211,261] as CD data have shown that αSyn lacks structural features typical of an unfolded protein. Additionally, SAXS previously showed that the R_g of αSyn is smaller than that of a theoretical random coil of 140 amino acids[210] and so can be considered to more compact than a classic IDP.

Different biophysical techniques were used in order to confirm these previous findings for our recombinantly engineered synuclein variants.

Analytical SEC was carried out on αSyn and its human homologues β and γSyn as these proteins are also IDPs but are slightly different molecular weights (14.49, 14.32, and 13.32 kDa for α-, β- and γSyn C-terminal Cys mutants respectively). Protein calibrants of various different masses were ran on a SEC column and a linear regression was fitted, synuclein variants lie off the calibrated line indicating that they run at a different mass than expected of a globular protein. This result shows that synuclein proteins are indeed more unfolded than classic globular proteins as they elute sooner than would be expected for a globular protein of their mass.

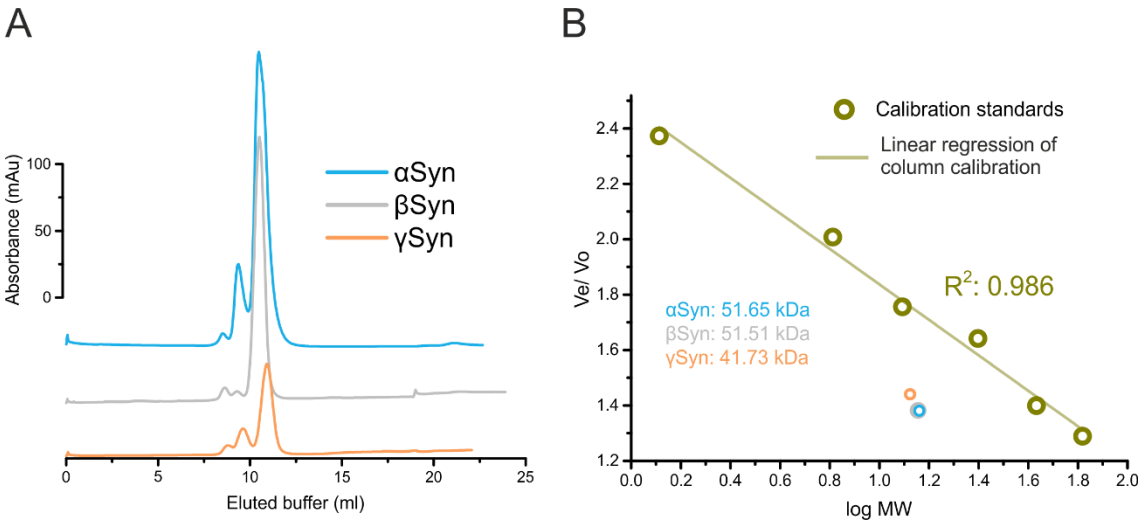


Figure 4-2. Analytical SEC of full length αSyn and homologues. A) Analytical SEC of αSyn and homologues studied in this thesis. Blue, grey and orange traces show α, β and γSyn respectively. B) Column calibration plot of the molecular standards BSA, ovalbumin, chymotrypsin A, chytochrome C, aprotinin and vitamin B_{12} (66, 43, 25, 12.4 6.5 1.3 kDa, respectively) fitted with a linear regression. On the y axis, Ve/Vo denotes the eluted volume/ void volume (calculated to be 7.6 ml from the elution of blue dextran with a MW of ~2000 kDa). Synuclein variants are shown on the plot using the same colour as A. These proteins do not fall near the line indicating that they are expanded IDPs. Apparent MWs were calculated for synuclein variants from their elution profile and the column calibration. These are shown inset and are all much larger than the actual MW (~14k Da) a characteristic typical of IDPs.

4.4.1 Circular Dichroism analyses

Far-UV CD analysis was carried out on WT αSyn (blue) and the PD familial mutant E46K (red) (Figure 4-3). These spectra are very similar, characteristic of essentially unfolded polypeptide chains with an absence of bands in the 210-230 nm region as has been shown previously for αSyn[210]. CD experiments were carried out in 20 mM Tris, 200 mM NaCl, pH 7.5 (to mirror SMFS conditions used in this chapter, see below) and so the spectra don't include wavelengths less than 205 nm due to significant noise in this region due to a high concentration of Cl^- ions. There is a slight increase in negative intensity at wavelengths around 222 nm for E46K

compared to WT. A similar effect has previously been observed with decreasing pH[210] (Figure 1-15) and was proposed to indicate an increase in pH-induced structure, providing evidence for a partially folded intermediate. It may therefore be the case that altering the charge of the protein in the E46K mutation causes subtle increases in protein secondary structure.

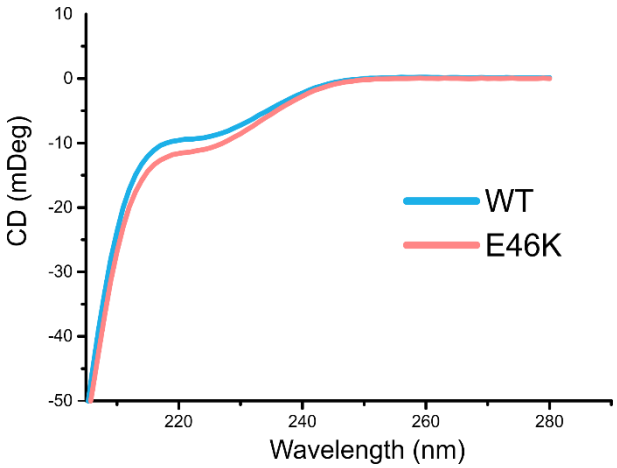


Figure 4-3. Far-UV CD spectra of WT and E46K αSyn at pH 7.5. The data show similar spectra for WT αSyn (blue) and the PD familial mutant E46K (red). The spectra are characteristic of a random coil structure typical of an IDP. There is a slight difference between WT and E46K that could suggest subtle differences in structure between the two proteins. The proteins were at 100 µM in 20mM Tris 200mM NaCl pH 7.5, the same conditions SMFS experiments was carried out (see below). The spectra are displayed from 205 nm onwards as at shorter wavelengths there is significant noise attributable to the high concentrations of salt in the buffer used.

4.4.2 MS

Native ESI-MS analysis was also carried in order to confirm the proteins were the expected mass and also to observe any quaternary interactions between αSyn monomers. Figure 4-4 shows native MS spectra of α- (blue), β- (grey) and γSyn (orange) as well as αSyn E46K (red). The mass of each protein was found be within error of the estimated masses indicating an absence of unexpected modifications or truncations in the recombinantly expressed synuclein variants. All spectra show that the synuclein variants are primarily present as monomeric proteins with very similar charge state distributions. Some of the charge states are high (up to 13+ is denoted for monomeric proteins but up to 18+ can be observed) typical of disordered proteins[358,359].

A small amount of dimers can be observed in spectra from all the synuclein variants indicating that all the proteins studied here have the ability to dimerise as observed previously[276]. It was also shown in the same study that the dimerisation propensities of the synucleins are not predictive of aggregation[276]. The data presented in Figure 4-4 may point to a similar conclusion

as there is an absence in correlation between dimer intensity and the aggregation propensity of the synuclein variants.

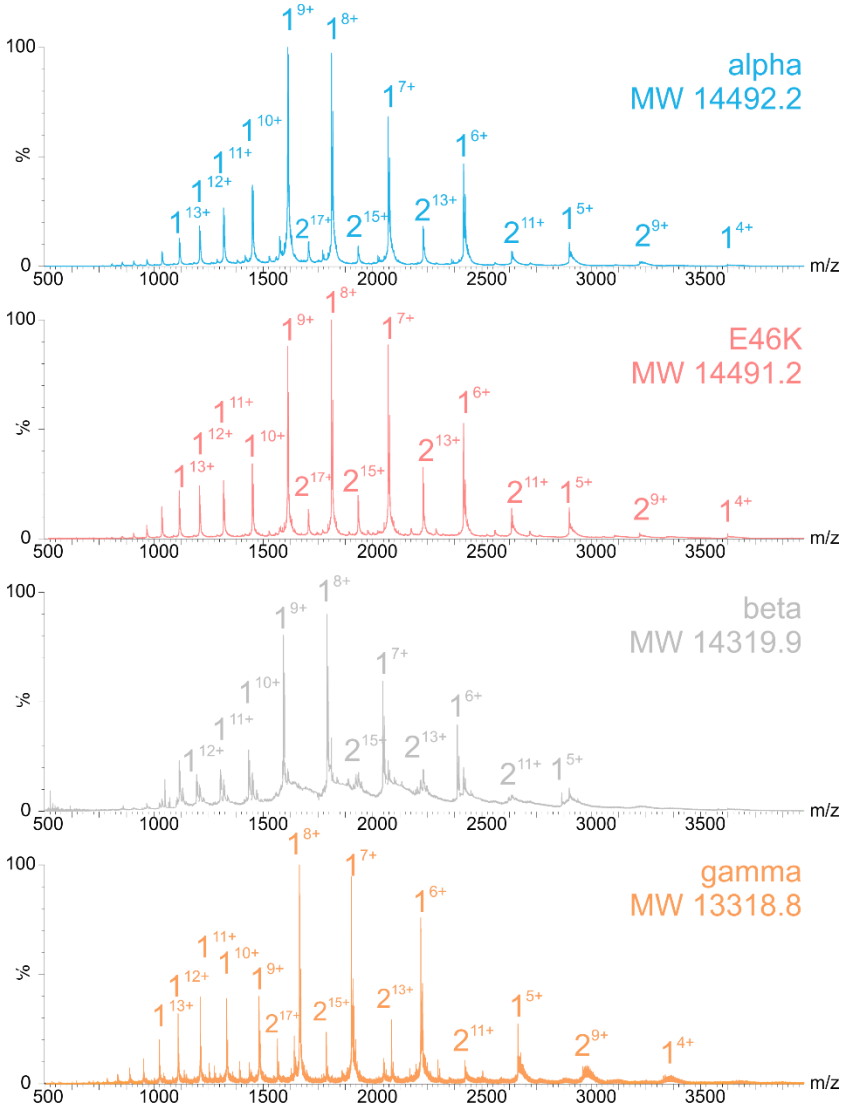


Figure 4-4. Native ESI-MS of αSyn, synuclein homologues and PD variant αSyn E46K. Mass spectra for WT, E46K, βSyn and γSyn are shown (blue, red grey and orange spectra respectively). Proteins were sprayed at 100 µM in 100mM acetate buffer pH 6.8. The proteins are present in a primarily monomeric state although all variants show dimer formation. Mass spectra show mass to charge on the x-axis (m/z) as a function of intensity. The numbers above the peaks denote the oligomer order, with the positive-charge state of ions in superscript.

ESI-mass spectra were also recorded of protein mixtures of αSyn and its homologues β and γSyn (Figure 4-5). In mixed samples, the individual spectrum of each homologue can be assigned to a particular protein variant due to the differences in MW. Mass spectra from mixed α/β (top) and α/γ (bottom) samples are shown in Figure 4-5. Numbers above peaks in blue, grey, orange and dark blue indicate a peak attributable to α-, β-, γSyn or mixed samples,

respectively. The data shows that both γ- and βSyn are able to form heterodimeric species with αSyn. This data may suggest that the association interaction observed is generic between synucleins and not dependent on aggregation propensity.

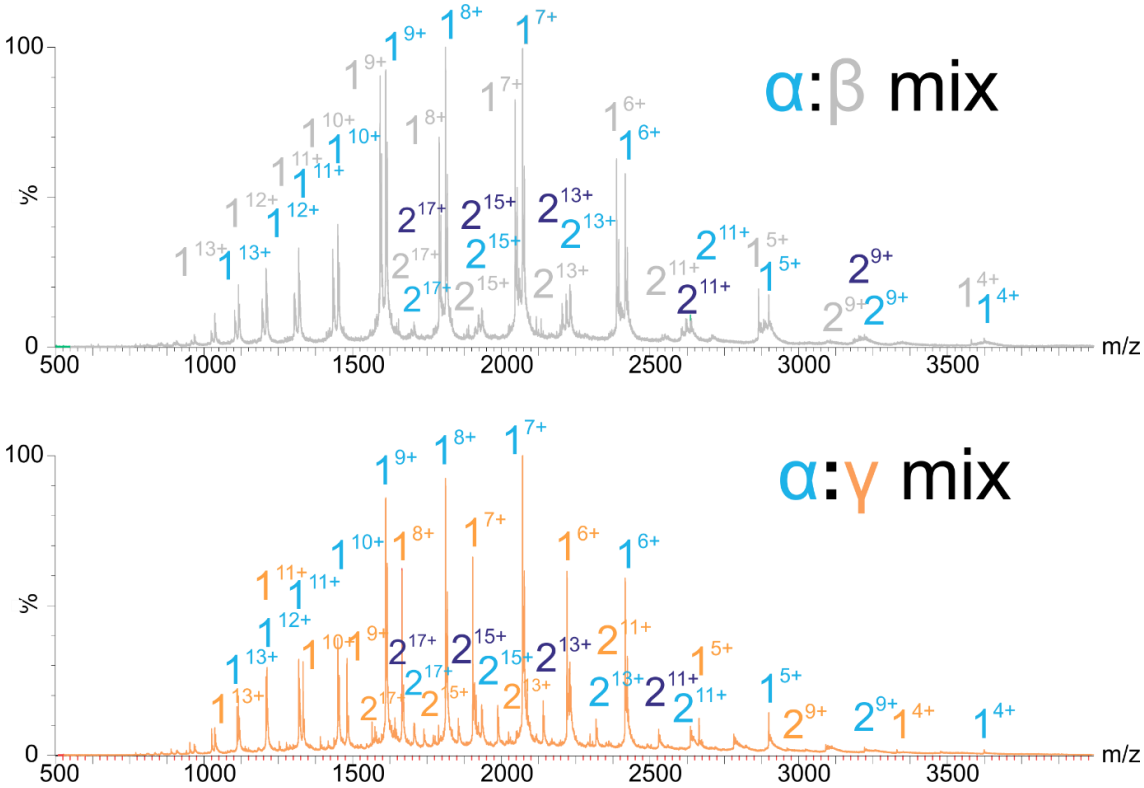


Figure 4-5. Native ESI-MS of heterogeneous mixtures of αSyn with either β or γSyn. Mass spectra for α/β and α/γ mixed samples are shown on the top and bottom panels respectively. Numbers above peaks in blue, grey, orange and dark blue indicate a peak attributable to α, β, γ or mixed samples respectively with the positive-charge state of ions in superscript.

To investigate the stability of each homodimer, ESI-mass spectrometry experiments were carried out in tandem with collision induced dissociation (CID) experiments. In this experiment homodimers were accelerated at increasing electrical potentials, causing the proteins to collide with neutral gas molecules in the collision cell triggering dissociation. The relative proportions of dimeric and monomeric species were monitored at different voltages. These experiments were carried out on synuclein variants (α-, β, γSyn and E46K) and the data shown in Figure 4-6. Overall, the voltages at which dimeric synuclein species dissociate are similar between all variants indicating similar affinities for self-dimerisation. Closer examination of the of the CID profiles reveal that E46K dissociates at a higher voltage than WT, the non-aggregation-prone synuclein homologues β- and γSyn dissociate at lower voltages than WT.

This may indicate a higher affinity dimeric interaction for the familial PD mutant αSyn E46K than that of WT αSyn or the αSyn homologues.

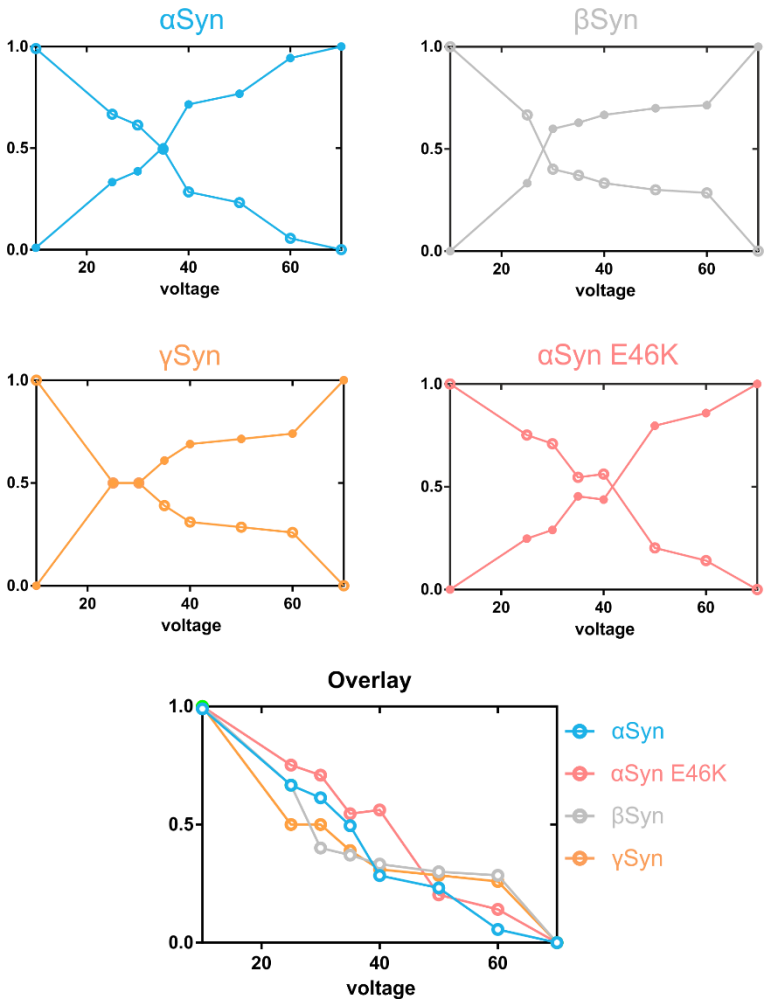


Figure 4-6. CID voltage dependence of dimer dissociation of synuclein variants. WT αSyn, βSyn, γSyn and αSyn E46K are shown in blue, grey, orange and red respectively. Open and closed circles show the populations of dimeric and monomeric species respectively. An overlay of the dimeric population is shown in the bottom panel. The dissociation profiles of all the synuclein species are similar. E46K dissociates at a slightly increased voltage than that of the other variants.

4.4.3 SAXS

SEC in combination with small angle X-ray scattering (SEC-SAXS) was used in order to analyse the structure and the conformation of monomeric αSyn as described (Section 2.2.3.3). For comparison, a familial PD mutated variant: E46K was also analysed using this method.

SEC was required in order to eliminate any contributions of higher order species that may occur thus isolating scattering data from monomeric αSyn. SEC-SAXS experiments were carried out in 20 mM Tris, 200mM NaCl pH 7.5 (to mirror SMFS conditions used in this chapter, see

below). Similar scattering profiles were observed for both WT and E46K αSyn (Figure 4-7). Kratky plots of the data (Figure 4-7 B) show features typical of a disordered protein as previously reported[210] with a lack of an obvious peak and a continuous rise in the profiles. By performing a Guinier approximation, the radius of gyration was found to be 41.1 and 41.5 Å for WT and E46K respectively. A similar R_g (40 Å) has been reported previously for αSyn at neutral pH[210]. The average R_g for a globular protein of the same size as αSyn would be 15.1 Å whereas a completely disorderd chain would have an R_g of approximately 52 Å[210]. αSyn has been shown previously, and here to be disordered, however, the experimental R_g reported here is smaller than that of a theoretically completely disordered polypeptide as has been reported previously[210]. This suggests that αSyn is more compact than a random coil and may have some structure in these conditions.

The conformation of αSyn has previously been shown to display a high degree of plasticity[343,360]. To further analyse the different conformational contributions of αSyn in the ensemble scattering data an ensemble optimisation method (EOM) was used (Section 2.2.3.4). EOM generates a large pool of random structures using sequence and structural information. A genetic algorithm then compares the average theoretical scattering intensity from *n* independent ensembles of conformations against the SAXS data and the conformations that best fit the experimental data (Figure 4-8 A) are selected. The genetic algorithm carries out this process in an iterative manner using various rounds of hierarchical selection in a fitness function process. The distributions of the EOM selected ensemble show a distinct bimodal distribution of collapsed and expanded states depicting the conformational heterogeneity of αSyn.

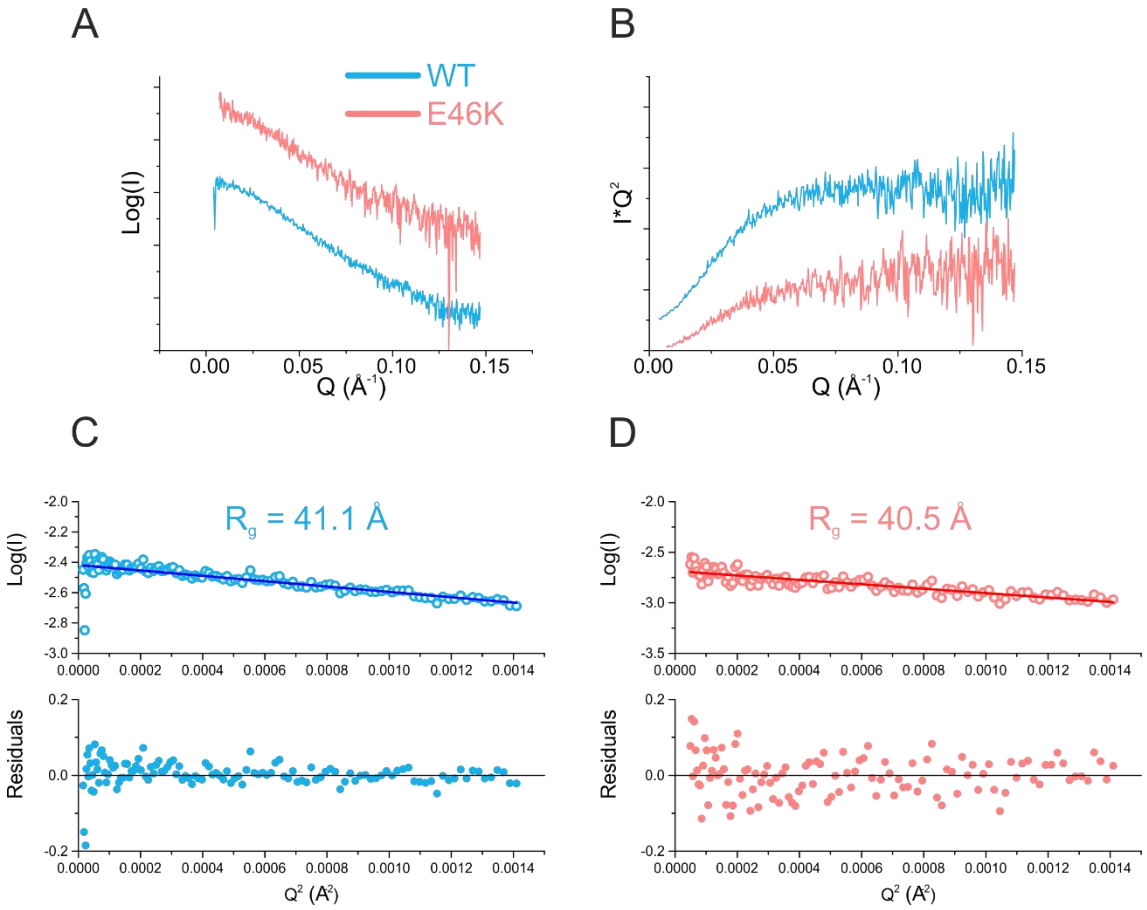


Figure 4-7. SEC-SAXS analysis of αSyn variants WT and E46K αSyn at pH 7.5. A) Scattering data of WT (blue) and the PD familial mutant E46K (red) displayed in a conventional semi-logarithmic plot. B) Kratky plot of experimental data containing features which are typical of an IDP - a continuous rise and an absence of a clear maximum peak. C) and D) Guinier plots of WT and E46K αSyn, respectively fitted with a linear regression. R_g values were estimated from a Guinier approximation to be 41.1 and 40.5 Å for WT and E46K, respectively. Scattering profiles in A and B are offset for clarity.

In agreement with these data, a previous study has also observed a bimodal conformational distribution for the αSyn monomer using SAXS and[343]. The R_g values for the collapsed and expanded distributions are shown in Table 4-1 calculated from Gaussian fittings from individual peak populations. The ratio of each population is different between WT and E46K with the PD mutant containing a higher proportion of expanded species (42.1 % compared to 25.3 %). These data show the same trend as with a previous EOM study in which the expanded species of E46K also increased in proportion relative to WT[343]. This difference between these variants could play an important role in the aggregation of the proteins which are otherwise very similar.

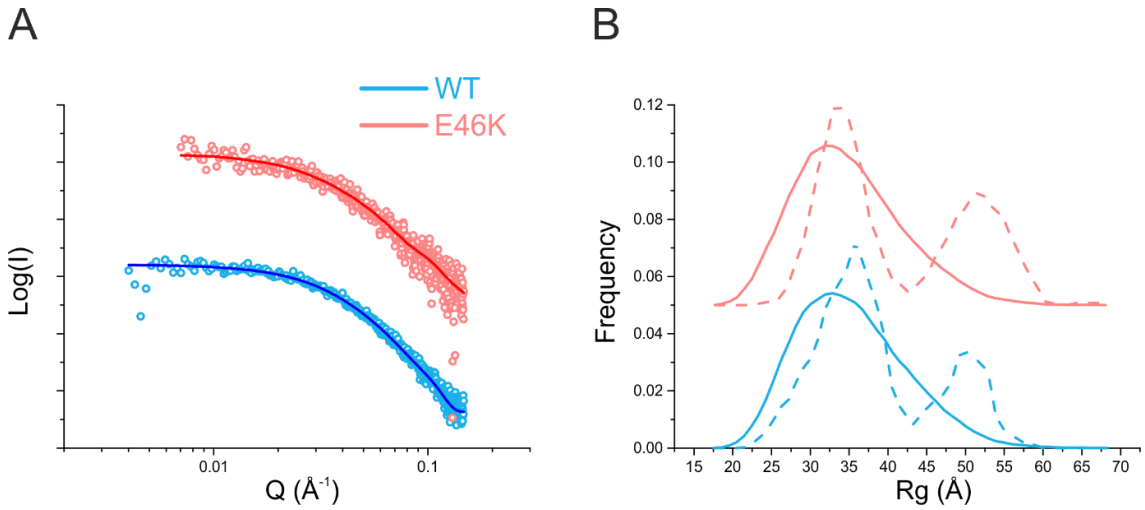


Figure 4-8. EOM analysis of SEC-SAXS data at pH 7.5. A) log-log plot of scattering data with EOM fits (solid lines) for WT and E46K αSyn (blue and red respectively). B) EOM R_g distributions. The distributions for the random pool (solid line) and those from the EOM-selected ensemble (dashed line) are shown. The EOM-selected ensemble structures show a double distribution for both constructs showing both expanded and collapsed species (see Table 4-1)

Protein variant	R_g (Å) collapsed peak	R_g (Å) expanded peak	Population of collapsed peak %	Population of expanded peak %
WT	35.0	50.3	74.7	25.3
E46K	33.8	51.7	57.9	42.1

Table 4-1. Data from EOM selected ensemble R_g distributions based on Gaussian fits of the individual peak populations observed in EOM data. % peaks are calculated from the area of Gaussian fits.

4.5 Aggregation of αSyn

4.5.1 Effect of solution conditions on αSyn aggregation

Fibril growth assays were carried out as described (Section 2.2.2.2) in order to characterise the aggregation of αSyn in different conditions and also to choose conditions in which SMFS experiments would be carried out (Figure 4-9). The amyloid-binding small molecule ThT was used to monitor the time dependence of aggregation in these experiments. Upon binding to β-sheet rich amyloid structures, the molecules fluorescence emission spectrum displays a characteristic red shift [361]. The dye is exited at 440 nm and increasing fluorescence in the presence of amyloid is monitored at 480 nm in these experiments.

The aggregation of αSyn in different solution conditions is shown in Figure 4-9 and Figure 4-10. The contributions of both NaCl (Figure 4-9) and pH (Figure 4-10) on the aggregation rate of

αSyn are highlighted. Increasing concentrations of NaCl increases the rate of aggregation (Figure 4-9) as has been shown previously[330,362] with a dramatic reduction of the lag time at 200 mM NaCl.

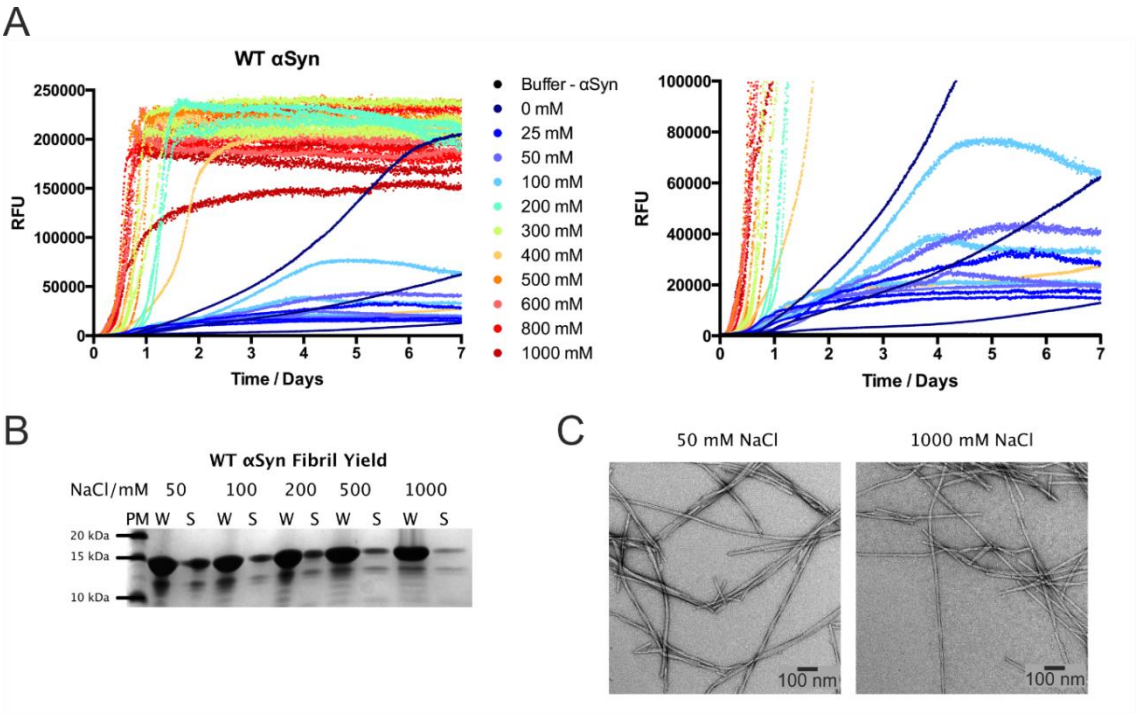


Figure 4-9. Aggregation of WT αSyn as a function of NaCl. A) ThT fibril growth assay of 80 µM αSyn in 20 mM Tris pH 7.5 as a function of varying NaCl concentration. Increasing NaCl increases the aggregation rate of αSyn with a dramatic reduction of lag time in the presence of 200 mM NaCl. The data are replotted with an expanded y-axis on the right. B) Fibril yields as assessed by SDS-PAGE with the whole sample (W) and the supernatant sample (S) shown for each NaCl concentration. The protein marker (PM) is shown in the left-most lane. C) Example TEM images of αSyn at 50 and 1000 mM NaCl showing fibrillar morphologies which aren't significantly different with different NaCl concentrations.

The rate of aggregation also increases with decreasing pH with observable changes in end point fibril morphology. The increase in aggregation upon acidification has been shown in previous studies[330,363]. The pI of αSyn is 4.67, the increase in aggregation at acidic pH values may therefore arise from the loss of negative charges and neutralisation of the protein molecule. In αSyn, acidic residues are concentrated at the C-terminus, loss of negative charges in this region may play a role in aggregation. The C terminus supplies flexibility to the protein and has been shown to play a role in preventing aggregation[208,209]. Indeed, as discussed previously, studies have shown that at more acidic conditions, the C-terminal region becomes more collapsed[221] and its proposed capping ability of the central highly hydrophobic regions is reduced[215,216,220].

Similar to the protonation of acidic residues, increasing NaCl concentrations may act to neutralise negative charges in αSyn via electrostatic shielding and act to increase the rate of aggregation via a similar mechanism.

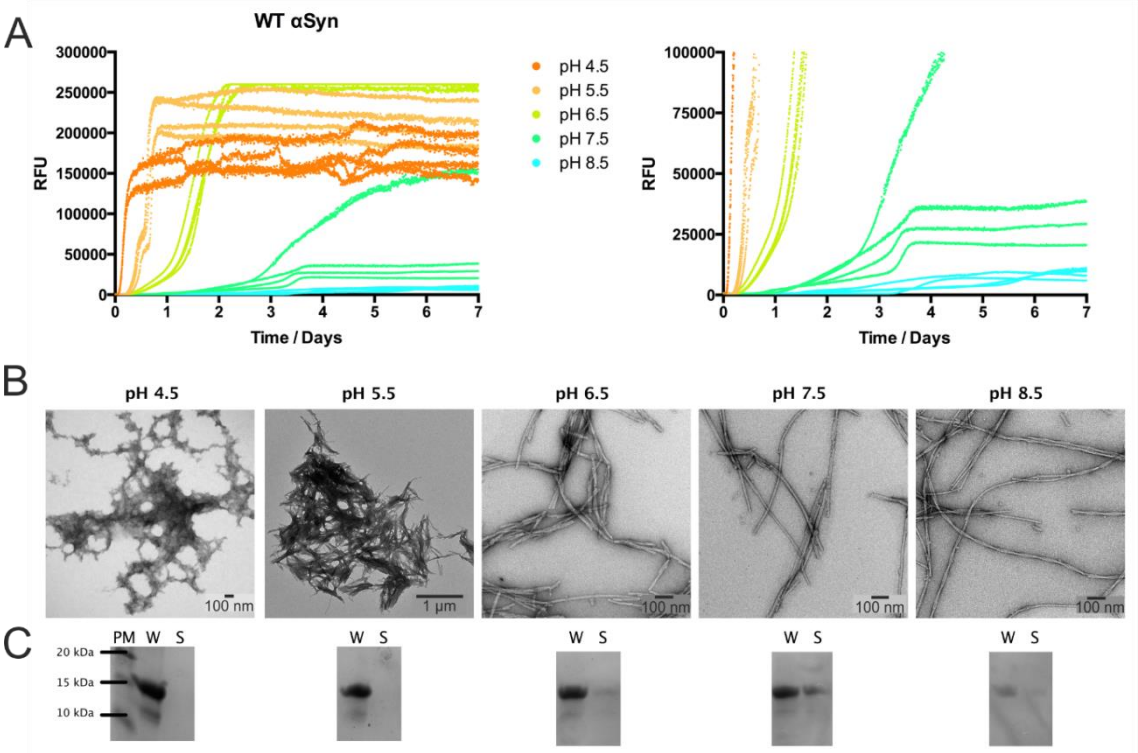


Figure 4-10. Aggregation of αSyn as a function of pH. A) ThT fibril growth assay of 80 µM αSyn in 20m M MES/Tris/Acetate buffer as a function of pH. Lowering of the pH increases the rate of aggregation of αSyn. The data are replotted with an expanded y-axis on the right. B) TEM images of αSyn incubated at different pH values. The morphology of aggregates changes with pH. C) Fibril yields as assessed by SDS-PAGE in which the whole sample (W) and the supernatant (S) sample after incubation is shown for each pH value assessed. Overall fibril yields appear to increase with decreasing pH.

4.5.2 The effect of the PD mutation E46K on aggregation

Six familial mutations of αSyn have been discovered to be associated with early onset PD. Of these, E46K is discussed extensively in this thesis. As for the other mutants, E46K is located in the N-terminal region of αSyn and has been shown to increase the rate of aggregation relative to WT[262]. In the previous section it was demonstrated that changes in charge due to alterations in salt concentrations and pH play an important role in the aggregation of αSyn. E46K was, therefore, chosen to complement WT in SMFS experiments as this mutant has the largest change in charge of the PD mutants (a negative to a positive residue).

The aggregation of E46K was monitored and the data are shown in Figure 4-11. E46K aggregates significantly faster than WT under the conditions studied. The lag times were 52.4 ± 7.5 hours and 16.3 ± 0.6 hours for WT and E46K, respectively. Negative stain TEM also revealed

the end-point morphology of the fibrils were also different between WT and E46K, which both formed fibrillar structures, those for E46K are twisted compared to the characteristic straight morphology observed for WT. These data show that a single charge swap mutation can cause significant alterations in the aggregation propensity of αSyn. This again indicates the importance and the fine balance that charge plays in αSyn conformation and aggregation.

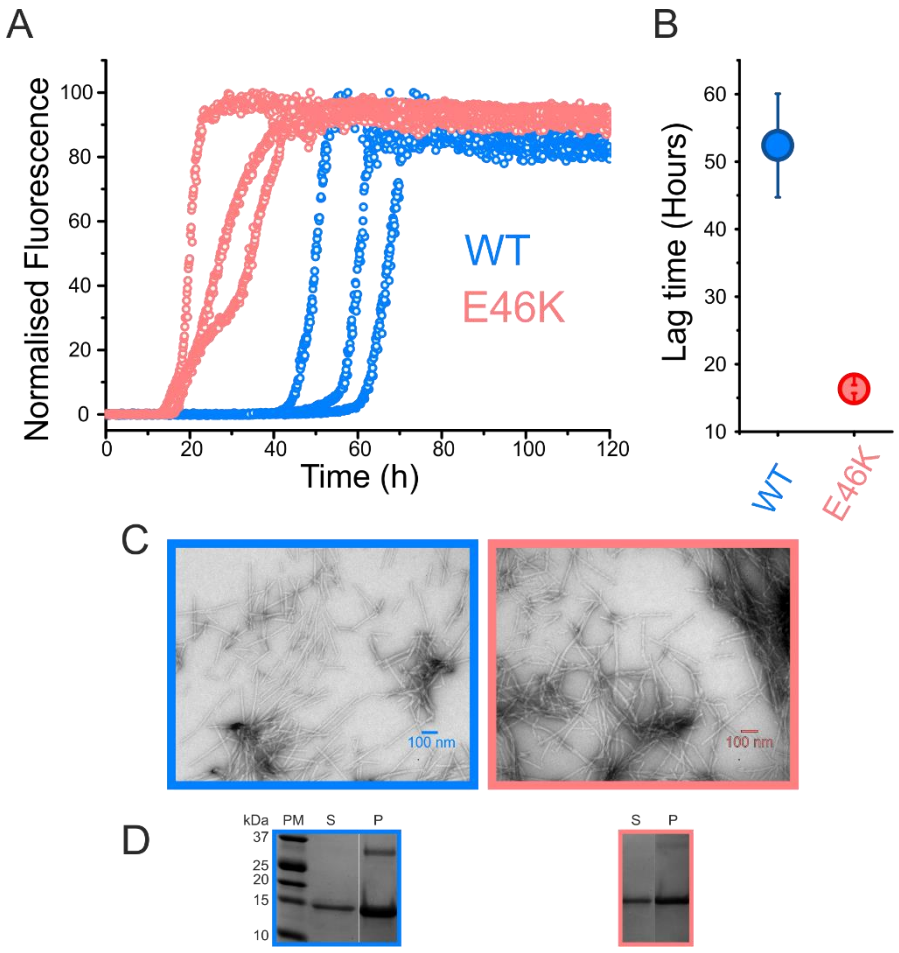


Figure 4-11. Aggregation of WT and E46K αSyn. A) Growth assay of 100 µM WT (blue) and E46K (red) in 20 mM Tris 200 mM NaCl pH 7.5 at 37°C and shaking at 600 rpm monitored by ThT fluorescence. B) Lag times of WT and E46K calculated from ThT data showing faster aggregation for the E46K. Average lag times are represented as filled circles, the mean lag time is at the centre of each circle, error bars show SD. C) Negative stain TEM of aggregates at the end point of the incubation, fibrillar structures can be observed for both WT and E46K, the morphology of which differs between the two variants, straight fibrils are observed for WT and twisted fibrils are observed for E46K. D) Fibril yields as assessed by SDS-PAGE. The protein marker (PM) soluble (S) and pelleted (P) samples are shown. The fibril yields are 76 % and 85 % for WT and E46K respectively. αSyn variants appear as a monomer in both the supernatant (S) and pelleted (P) sample, dimeric species are also present in the pelleted sample.

4.5.3 Aggregation propensity of the human homologues of αSyn

β- and γSyn, the human homologues of αSyn, are less aggregation-prone and not linked to PD[273-275]. The aggregation propensity of these proteins were analysed in the conditions used

for SMFS experiments (Figure 4-12). The data indeed show that whereas αSyn forms ThT positive material over time, there is an absence of an increased ThT signal for β- and γSyn in the same conditions up to 150 hours. The self-association of synuclein observed in MS studies described in Section 4.4.2, is therefore not likely to be on pathway to aggregation.

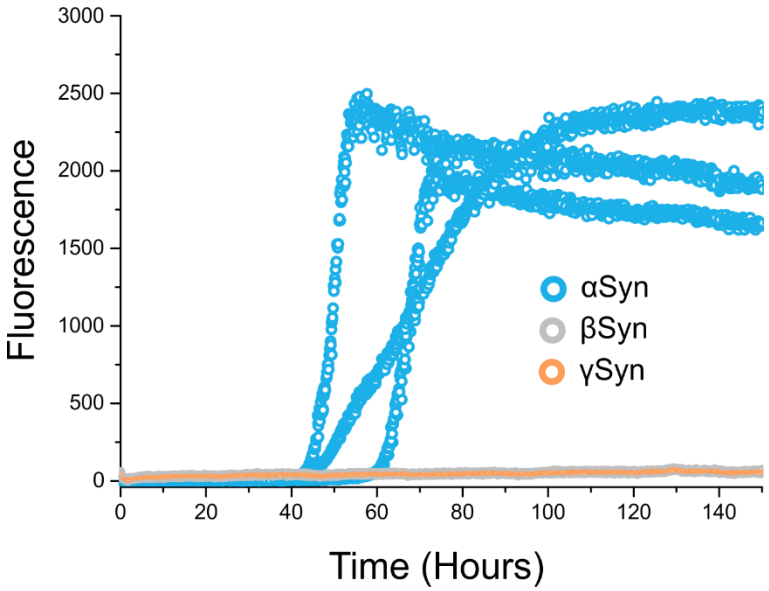


Figure 4-12. ThT aggregation assay of αSyn and its human homologues. ThT fibril growth assays of α, β and γSyn in blue, grey and orange respectively. The data shows that αSyn aggregated to form ThT positive material after around 40 hours whereas β and γSyn solutions don't show an increase in ThT fluorescence after 150 hours at least.

4.6 SMFS experiments

One of the main aims of this thesis is to analyse the dimerisation interaction between monomers of αSyn on a single molecule scale. A schematic of the SMFS experimental setup used to achieve this is shown in Figure 4-13. In this design, variants containing single Cys residues are covalently immobilised onto silicon nitride AFM probes and surfaces using flexible heterofunctional PEG linkers (SM PEG_{24}). The functionalised tip then approaches a similarly functionalised surface in buffered solutions allowing αSyn to form dimerisation interactions (trigger: 300 pN with no dwell). The interaction is then pulled apart by withdrawing the probe from the surface at 1000 nms^{-1}. Force maps consisting of 500 approach retract cycles taken over an area of 20 μm^2 were taken and the cantilever moved to a new area of the functionalised silicon nitride wafer between force maps (to ensure adequate surface coverage and to avoid contributions of high local protein concentration which may occur). All detected events (that meet filtering criteria, Section 2.2.4.6) are then fit to a WLC model (Equation 1-1) with a fixed persistence length (0.4 nm), using the hard contact to identify the zero distance and the retraction baseline to zero the force.

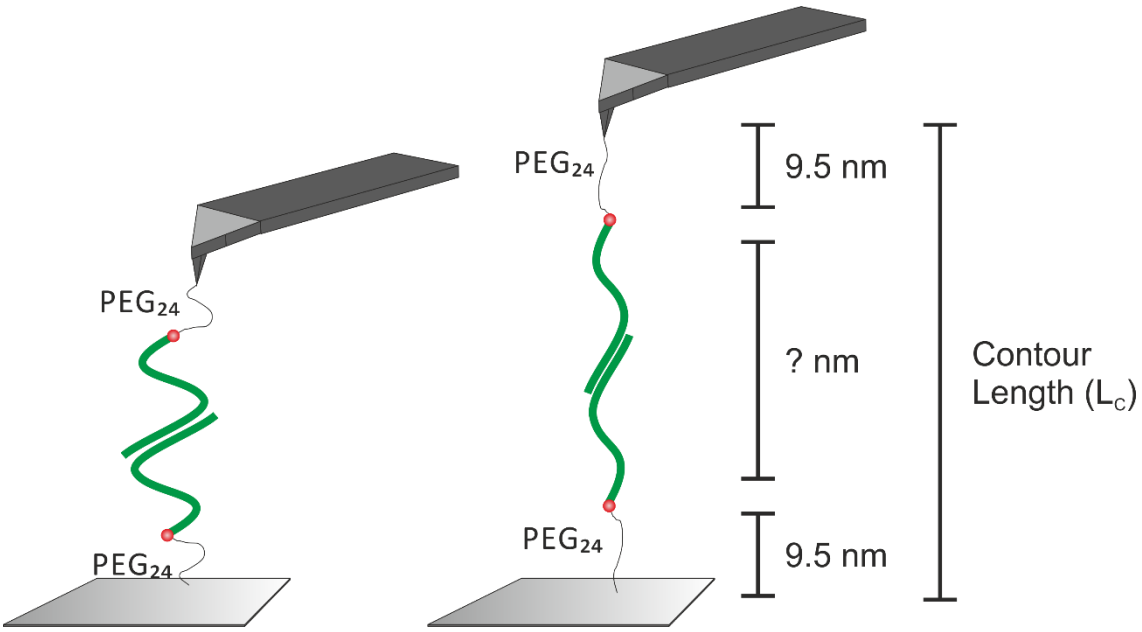


Figure 4-13. Schematic of SMFS experimental setup to investigate dimerisation of full length αSyn. αSyn Cys mutants (represented as green chains lacking structure with red dots showing the position of a unique Cys residue) are covalently immobilised onto silicon nitride AFM probes and surfaces via hetrofunctional PEG linkers with 24 repeating units with a defined length of 9.5 nm. Functionalised cantilevers are brought into contact with similarly functionalised surfaces enabling αSyn molecules to interact. Upon interaction, withdrawing the AFM probe will propagate force across the dimers. The force increases until the protein complex ruptures and at this point the F_R can be found and L_C can be calculated.

4.6.1 Single molecule dimerisation of αSyn

4.6.1.1 Optimising SMFS experiments

In order to obtain conditions in which single molecule events can be observed reliably in force experiments, different concentrations of proteins and or PEG linkers were used. It is especially important in force experiments of aggregating systems that the surface density of protein is low enough so that contributions from higher order species are limited. PEGylated silicon nitride surfaces and cantilevers were first incubated with 50 µM A140C αSyn (Cys mutant at the C-terminal residue) and the data are shown in Figure 4-14. The data are presented in scatterplots of L_C as a function of F_R, the deeper colour denoting a higher frequency of events. Combining the data in a scatter plot allows for the specificity of the interaction to be assessed by the presence (specific interactions) or absence (non-specific) of a 'hotspot' in the data indicative of correlated forces and distances (Figure 4-14 B). Individual force events were classified into single events in which only one event is observed in a force-extension trace; or multiple events when more than one event is observed. In the data using 50 µM protein solution for the immobilisation step (Figure 4-14) there is an absence of obvious hotspot in the

total, single and multiple type events. The hit rate for the total combined single and multiple events is 17.9% which is high in experiments aiming to interrogate single molecule interactions. Typically, the hit rate for SMFS experiments should be low at around 5% to ensure a single molecules regime[364]. The ratio of multiple: single events is also high, this again suggests that interactions of a higher order than single molecule are being probed in this experiment. Overall, the data suggests that in these conditions, it is not possible to reliably analyse specific, single molecule interactions.

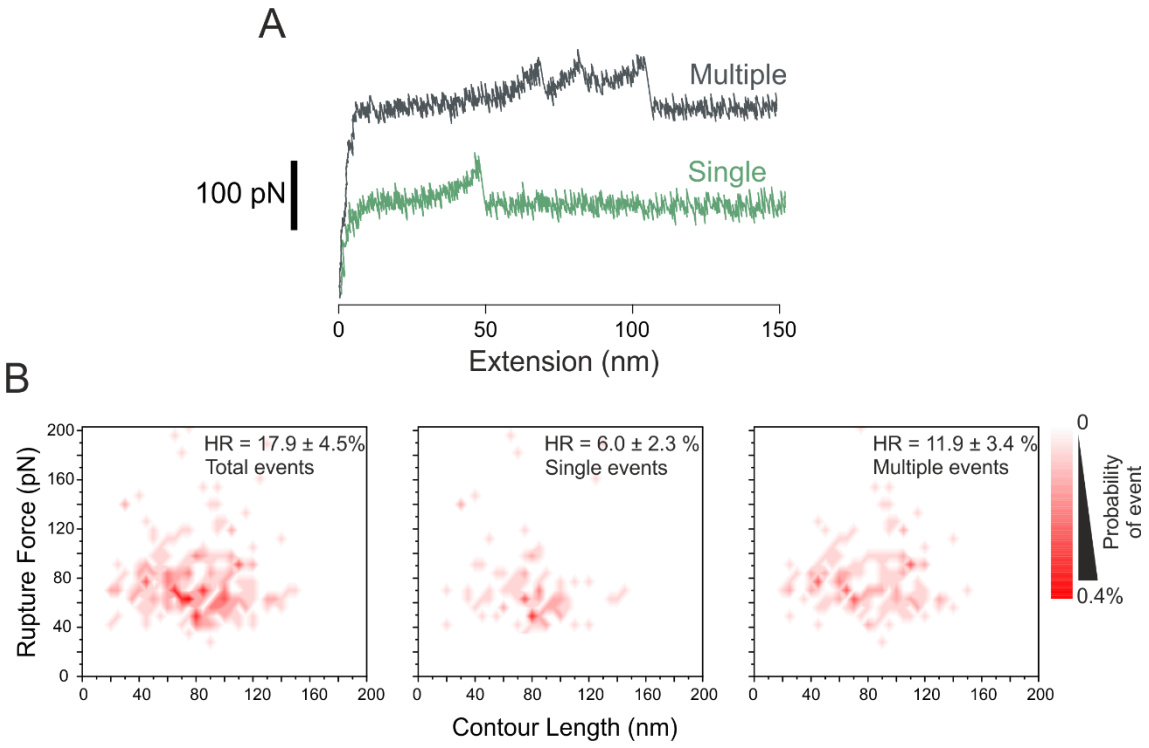


Figure 4-14. SMFS data during optimisation experiments using 50 µM αSyn A140C protein solution for the immobilisation step. A) Raw force-extension traces with examples of a multiple (dark green) and single event (light green). B) Scatter plots of SMFS displaying contour length as a function of force. The points were binned by increments of in L_C of 5 nm and F_R of 7 pN. The data were normalised to the total number of approach-retract cycles in the experiment. Deeper colours denote higher frequency of events. From left to right scatter plots for total events, single events, and multiple events observed during 1500 approach-retract cycles. Hit rates (HR) are shown inset for each type of event. Errors are SD of HR between force maps. A clear 'hotspot' of events is absent in all of the plots indicating non-specific interactions. All experiments were carried out at 20 mM Tris, 200 mM NaCl pH 7.5.

The SMFS experiments as outlined in Figure 4-13 and Figure 4-14 were repeated using 5 µM of αSyn A140C for immobilisation. The multiple and single events were analysed as before (Figure 4-14). The data shows that at this immobilisation density, single events occur at a 3.4 % hit

rate, whereas reducing the concentration of immobilised protein has reduced to incidence of multiple events to 0.6 %, this is within the level of noise from those observed in control experiments (Figure 4-16). Moreover, a clear hotspot is present in the single event analysis suggesting that distinctive single molecule interactions are being observed. These conditions, therefore, were used in all subsequent SMFS experiments on full length synuclein variants.

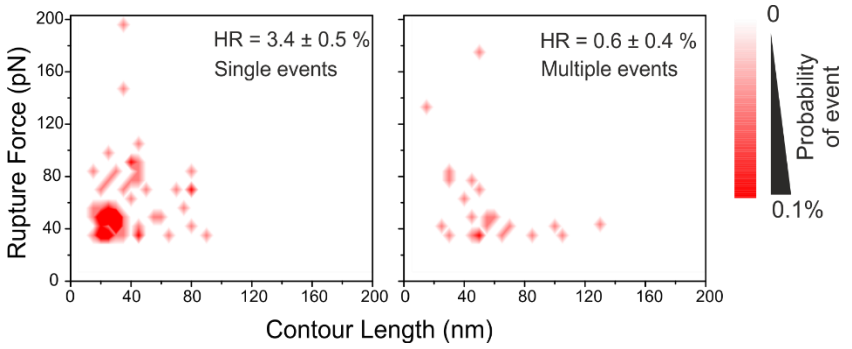


Figure 4-15. SMFS data during optimisation experiments using 5 µM αSyn A140C protein solution for the immobilisation step. Scatter plots of SMFS displaying contour length as a function of force. The points were binned by increments in L_C of 5 nm and F_R of 7 pN. The data were normalised to the total number of approach-retract cycles in the experiment. Deeper colours denote higher frequency of events. From left to right scatter plots for total events, single events, and multiple events observed during 2000 approach-retract cycles. Hit rates (HR) are shown inset for each type of event. A clear 'hotspot' of events is present in for analysed single molecule events (left) and absent for multiple events (right). All experiments were carried out at 20 mM Tris, 200 mM NaCl pH 7.5.

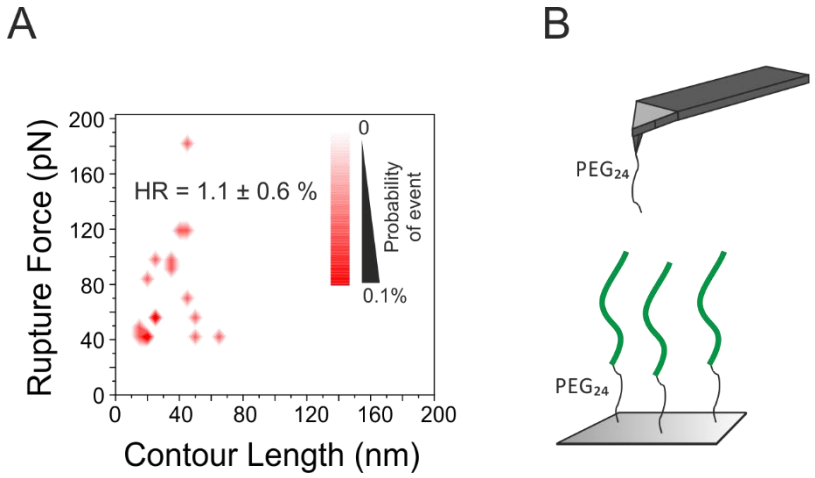


Figure 4-16. Control SMFS experiment. A) Scatter plot of SMFS data displaying contour length as a function of force. The points were binned by increments in L_C of 5 nm and F_R of 7 pN. The data were normalised to the total number of approach-retract cycles in the experiment (total 1,500). Deeper colours denote a higher frequency of events. A 'hotspot' is absent in the data suggesting a lack of a specific single molecule interaction. B) Schematic of the control experiment carried out with Si_3N_4 surfaces immobilised with 5 µM αSyn A140C and cantilevers in absence of protein.

Control experiments showed an absence of a hot spot of data when only the surface was functionalised with protein (Figure 4-16). This confirms that the SMFS data shows distinctive interactions between αSyn molecules immobilised on AFM tips with those immobilised on Si_3N_4 surfaces.

SMFS experiments using optimised conditions described above are shown in Figure 4-17. Characteristic single molecule events shown in Figure 4-17 A were fitted to the WLC model in order to obtain L_C and F_R. The experiment was carried out in triplicate with each single experiment consisting of at least 4 force maps of 500 approach retract cycles. The overall number of approach retract cycles for the experiment in triplicate was 10,000. The hit rate is lower (3.0 %), increasing the probability that true single molecule complexes are being observed. There is an obvious hotspot of data as displayed in the scatter plot (Figure 4-17 B). Histograms of L_C and F_R are shown and the modal Gaussian value is denoted (L_C = 26.5 nm, F_R = 48.9 pN). Overall the data suggests that we are able to observe specific single molecule interactions in these conditions.

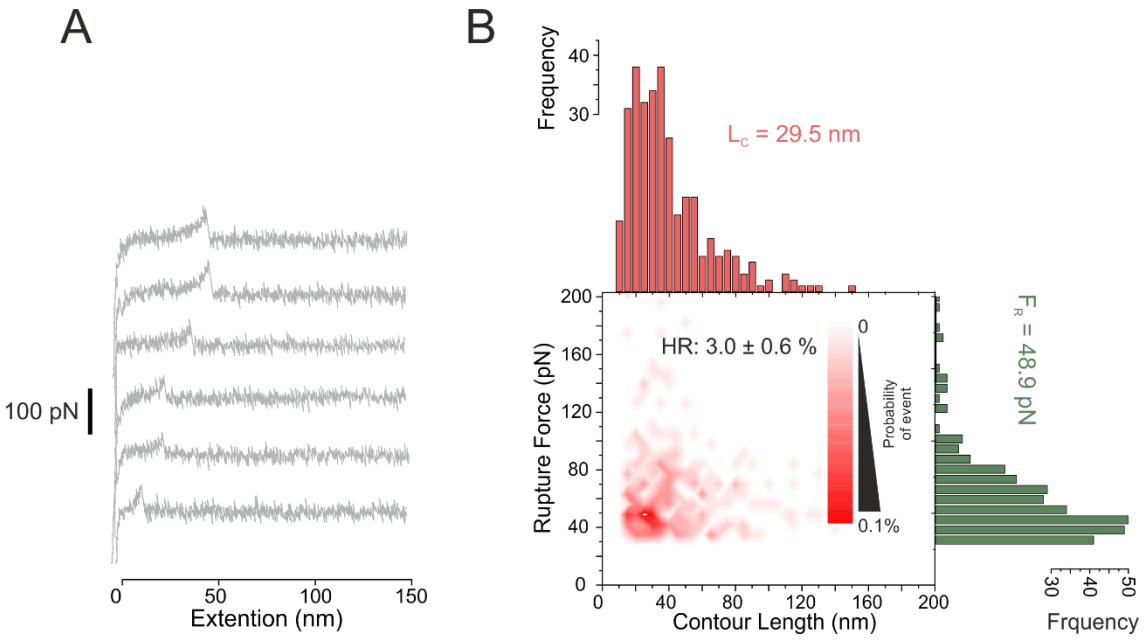


Figure 4-17. SMFS data from combines triplicate experiments for αSyn A140C. A) Example raw force-distance curves for αSyn A140C showing characteristic single events. B) Scatter plot of SMFS data displaying contour length as a function of force. The points were binned by a contour length of 5 nm and a rupture force of 7 pN. The data was normalised to the total number of approach-retract cycles in the experiment (total 10,000). Deeper colours denote higher frequency of events. A 'hotspot' is clearly present in the data suggesting a specific single molecule interaction. The hit rate (HR) is shown inset for the combined triplicate experiments. Histograms of both L_C (red) and F_R (green) distributions are also shown with the modal value from Gaussian fittings denoted. All experiments were carried out at 20 mM Tris, 200 mM NaCl pH 7.5.

SMFS on αSyn A140C was carried out in triplicate and the reproducibility of data using SMFS was assessed between triplicate data sets is displayed in Figure 4-18. The distributions of L_C (red) and F_R (green) of individual triplicates for αSyn A140C as shown in Figure 4-17. Histograms are shown in the top panes and the modal value from Gaussian fits are denoted, these values are similar between the triplicate data sets. It is common in SMFS to analyse data in this way, however, the Gaussian fits of the L_C and F_R distributions may not be sufficient as it is unclear whether the observed data represents multiple populations as it is not normally distributed. Box plots are, therefore also plotted and median values taken as another comparable value to the modal L_C. Box plots show that there is some asymmetry in the data and minor variations between triplicates that is mainly due to differences in higher values of L_C and F_R, this is most likely due to noise as there is no correlation or hotspot in the data at these higher values.

In contrast to the data presented here, previous work by Lyubchenko and colleagues failed to observe dimerisation of αSyn at close to physiological pH[316,318,319]. This is most likely because of the differences in protein concentrations used in the immobilisation step (5 µM in this study compared to 19 nM in previous studies). The hit rate for αSyn dimerisation at pH 7.0 was reported to be 0.8 % by Lyubchenko and colleagues[318] and not further analysed. It is likely, therefore, that the interactions of αSyn at neutral pH were not analysed in previous studies. We have, therefore, presented the first SMFS of αSyn dimerisation under physiological like conditions here.

The L_C from SMFS experiments will be discussed in detail in the subsequent sections and throughout this thesis. The F_R of αSyn dimeric interactions observed in SMFS experiments will not be the main focus of attention in this thesis, however, the F_R observed in SMFS (Figure 4-17 and Figure 4-18) are surprisingly high for a non-evolved interaction of two IDPs. The cognate interaction between the *E. coli* proteins E9 and Im9 as studied by this group[305] dissociates at a similar F_R of ≈40 pN at 1000 nm/s (comparable to pulling velocity used here).

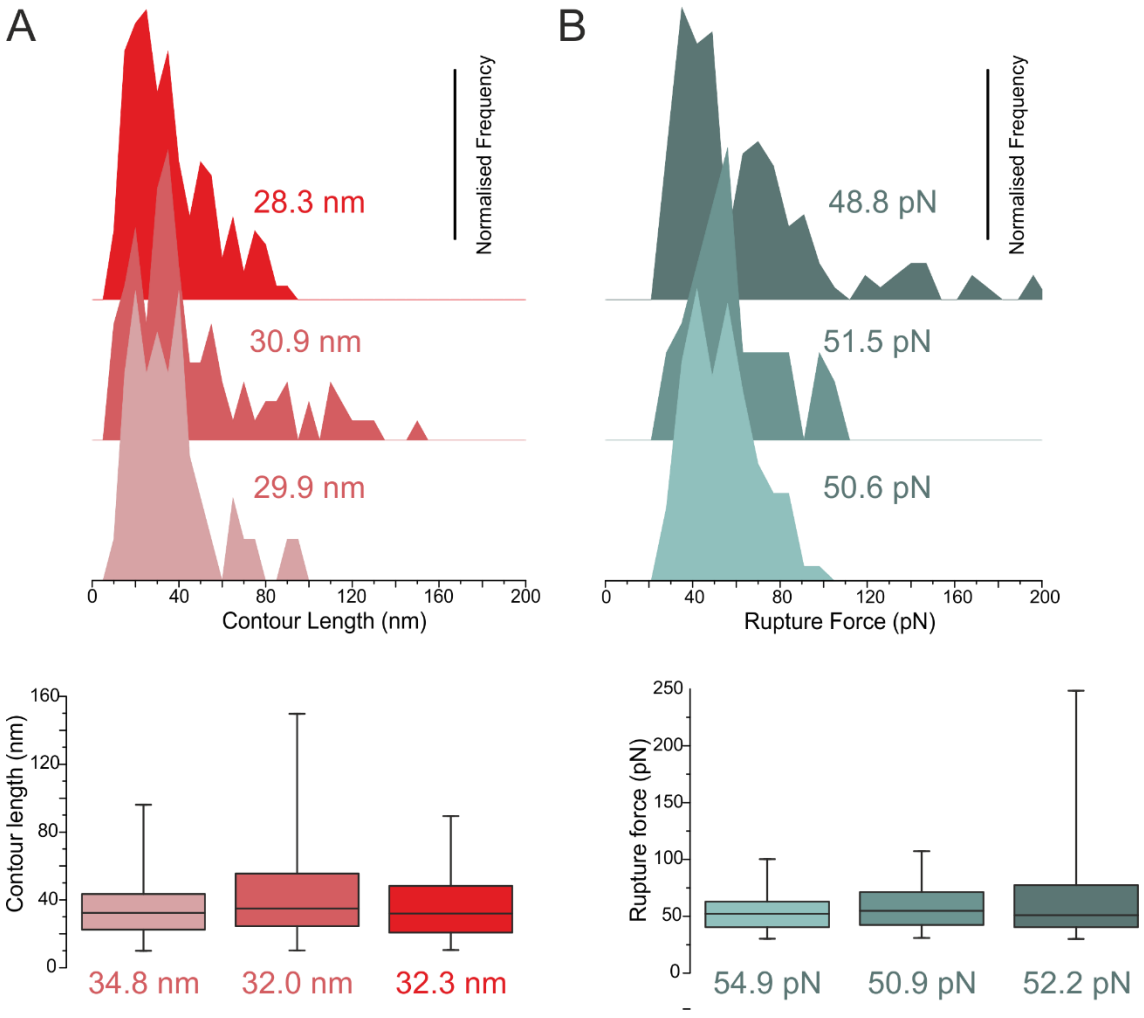


Figure 4-18. L_C and F_R distributions for each triplicate SMFS experiment of αSyn A140C. A) Top, L_C distribution histograms (red, bin size 5 nm) displayed as area plots with modal values from Gaussian fittings denoted. Frequency is normalised to the maximum frequency of each triplicate. Box plots of L_C distributions with the same colour coding as above showing some asymmetry of the data sets. The median L_C values are 34.8, 32.0 and 32.3 nm and for box plots from left to right. Whiskers on box plots are minimum and maximum values B) F_R distributions are shown in the same manner as L_C (histograms binned at 7pN). The median values for box plots from left to right are 54.9, 50.9 and 52.2 pN. Whiskers represent minimum and maximum values. The data shows that similar distributions of events are observed between triplicate experiments. All experiments were carried out at 20mM Tris, 200mM NaCl pH 7.5.

4.6.2 Analysis of SMFS experiments

4.6.3 Contour length simulations

The L_C data reveal a great deal of information about the dimerisation interaction of αSyn. As discussed (Section 1.6.2), the L_C can be used to interrogate different aspects of dimeric structure such as the interaction region or the conformation of the proteins in the dimer. The L_C data cannot, however, simply be interpreted as the length of linkers and protein. This is due

to the important effects that linker length and AFM tip geometry have on the apparent L_C values measured experimentally (depicted in Figure 4-19). These effects have largely been neglected previous SMFS studies of proteins leading to incorrect estimations of L_C. This issue has been addressed by our group previously[306]. Brockwell and colleagues have derived a model that predicts L_C for SMFS experiments. This model was further adapted by Dr Yun Chen (Brockwell Group, University of Leeds). An additional parameter was added to the published model which uses as a Monte Carlo procedure which accounts for the stochastic rupture of protein complexes which can occur anywhere along the theoretical WLC parabolic curve. This property has an effect on the L_C values obtained.

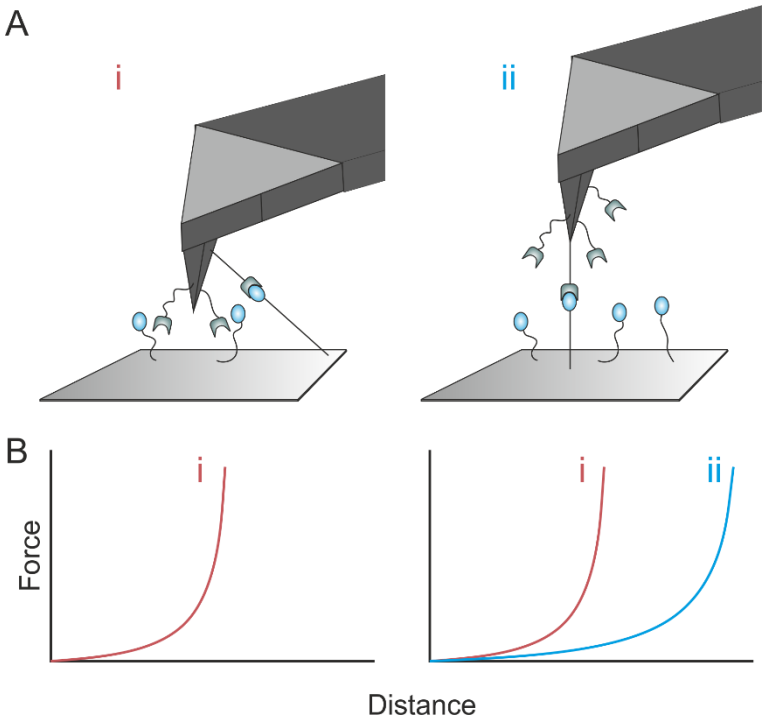


Figure 4-19. Schematic representation of considerations for determining the observed L_C in SMFS experiments. The observed L_C is dependent on the immobilisation location of protein partners on both the tip and surface. A) Schematic of different possible attachment positions on AFM tips and surfaces. B) Corresponding schematic force curve that describe the events in A. i) The interaction of proteins distal to the apex of the AFM tip and not directly below the tip results in shorter observed L_C values than expected (red line). ii) The maximum observable L_C corresponds to an interaction when a protein immobilised onto the tip apex interacts with a protein partner directly below it on the substrate (blue line). Figure redrawn and adapted from Farrance *et al* 2013[305].

It was assumed that αSyn behaves as an IDP in SMFS experiments and that αSyn dimers have a single interaction site. Given the results from the previous chapter, and previous literature, this interaction region was hypothesised to exist in the NAC region.

To locate the interaction interface, the L_C model simulation described above was carried out and compared to experimentally determined L_C values. In all, 14 separate L_C simulations were carried out, each modelled a 10 residue sequence of αSyn interacting with the corresponding sequence in the second monomer in a dimeric interaction. This was carried out throughout the protein sequence in sequential, non-overlapping, 10 residue sequences. The modelled L_C data is shown in Figure 4-20 as coloured histograms, the experimentally observed L_C data for αSyn is overlaid as a black line. The data suggests that the dimerisation interaction of αSyn occurs at the C-terminus of the protein primarily in residues 100-140. This is in contrast to the hypothesised interaction interface in the highly aggregation-prone NAC region. The experimental distribution of data has a wide range which may indicate some heterogeneity of the interaction region.

It is however important to consider the caveats with modelling the data in this manner. The model treats the αSyn monomers as essentially unstructured polypeptide chains. As discussed previously and highlighted in data in this chapter, αSyn does exist as an IDP[210,211], but, it has been shown to form long range transient interactions[212-214,216-218] and to exhibit some structure in the monomer[210]. We may not therefore be able to treat the proteins as essentially disordered chains in the modelled data. If there were indeed some structure in the dimer, the modelled approach would locate an interaction interface closer to the immobilisation region. To our best knowledge, there are no studies in the literature of the conformational arrangement of dimeric αSyn.

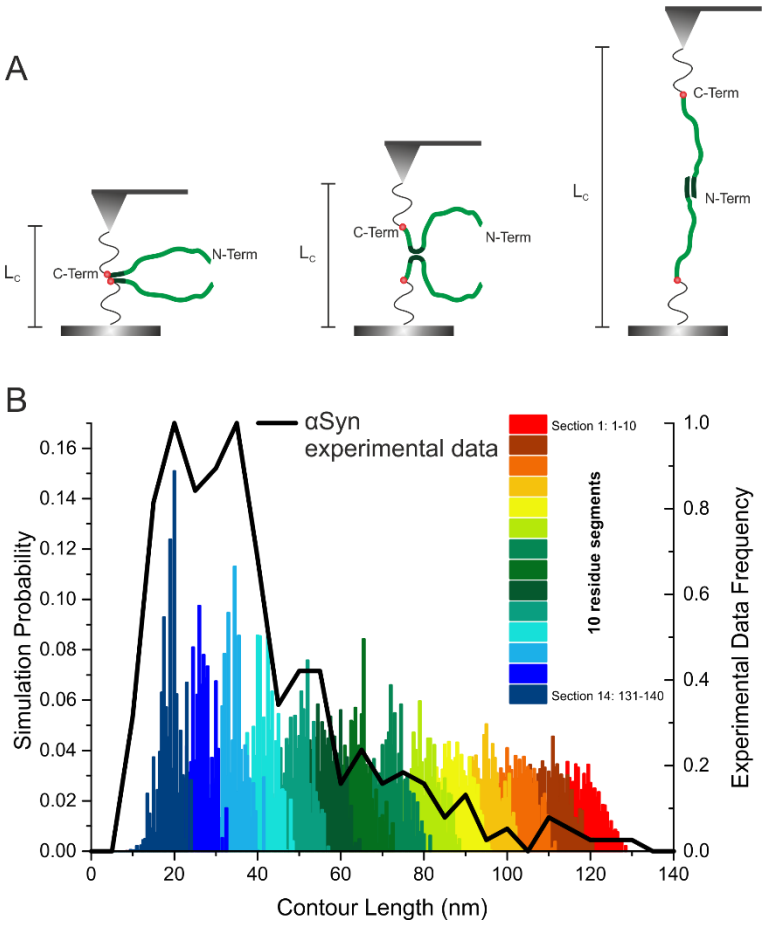


Figure 4-20. L_C modelled data for αSyn A140C SMFS experiments with overlaid experimentally observed data at 20 mM Tris 200 mM NaCl pH 7.5. A) Schematic of approach in modelling the L_C arising from different interaction interfaces. αSyn is represented by light green chains lacking structure. Interaction regions are shown in dark green. The immobilisation point (in this case C-terminal A140C). the L_C increases with increasing distance of the interaction region from the immobilisation point. B) Modelled L_C data is shown in histograms coloured from blue to red in order of increasing L_C. Histograms of modelled data simulate a 10 residue sequence of αSyn interacting with the corresponding sequence in the second monomer. This was carried out throughout the protein sequence in sequential, non-overlapping, 10 residue sequences. The simulation probability is shown on the left y-axis. Experimentally observe L_C data is shown as a black line depicting the outline of a histogram (bin size 5 nm). The normalised frequency of experimental data is shown on the right hand y-axis.

4.7 Single molecule dimerisation of a PD familial mutant

To further validate the approach, we next examined the interaction of the PD familial mutant, E46K dimers as the protein has distinct aggregation kinetics to WT yet is highly similar in sequence and conformation, a change in the scatter plots would reinforce the idea that real interactions are being interrogated. E46K was immobilised in the same manner as WT at the C-terminus (A140C). The SMFS data for E46K is shown in Figure 4-21 (red) and is compared to WT data (blue). The data for E46K, as with WT is the sum of triplicate experiments. Each individual experiment for E46K consisted of at least 5 force maps of 500 approach-retract

cycles. The overall number of approach-retract cycles for the experiment in triplicate was 9,000. The scatter plots in Figure 4-21 A show a different scatter profile for the E46K when compared to WT. L_C distributions are displayed in Figure 4-21 B and C, the data show that E46K has a longer L_C than that for WT (Gaussian modal values of 29.5 and 48.4 nm for WT and E46K respectively) suggesting that the conformation of the E46K and/or the interaction interface is different in E46K relative to WT. the longer L_C may therefore be reporting on a dimeric interface that is further away from the C-terminus (immobilisation point) or/and in fact the protein conformations are more expanded. SAXS data (Section 4.4.3, Figure 4-8) indeed indicates that a higher proportion of monomeric E46K exists in an expanded conformation (42 %, 51.7 Å) than WT (25 %, 50.3 Å).

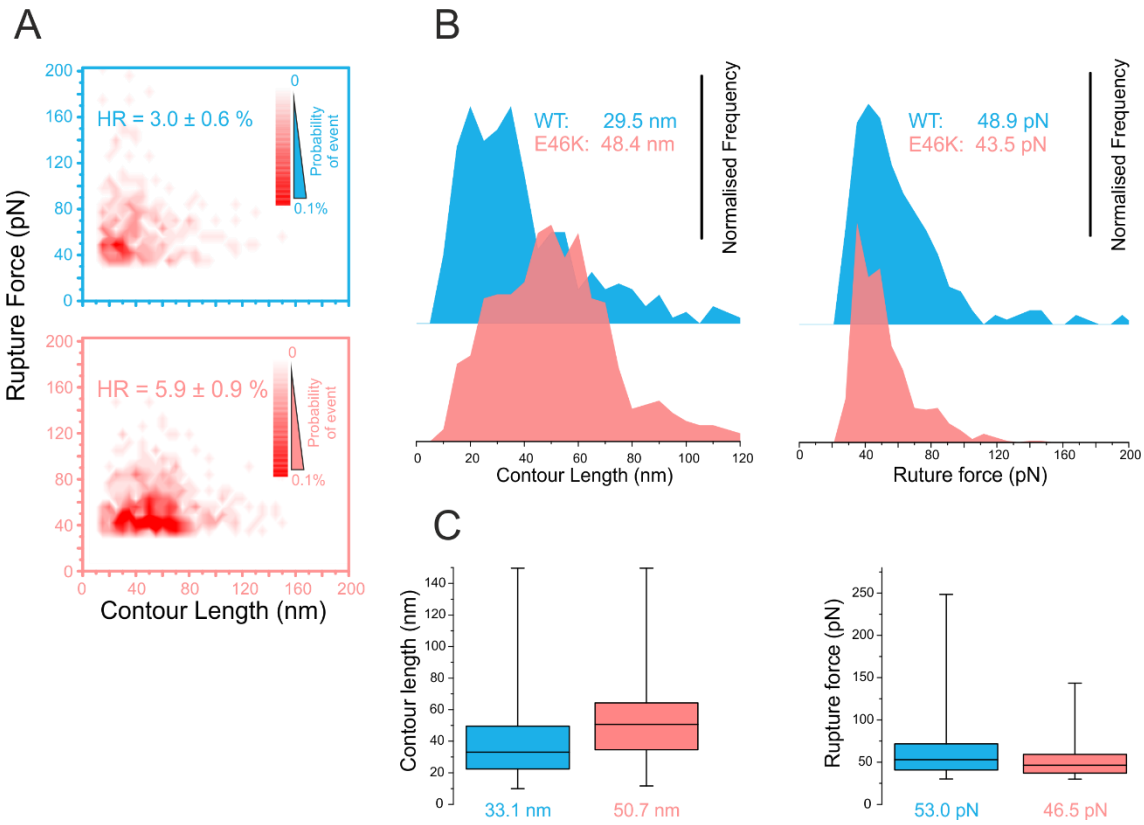


Figure 4-21. SMFS experiments of WT and E46K. A) Scatter plots of SMFS triplicate experiments for WT (blue), and E46K (red). The same colours are used throughout. The points were binned by increments of in L_C of 5 nm and F_R of 7 pN. The data were normalised to the total number of approach-retract cycles in the experiment (total 9,000). Deeper colours denote higher frequency of events. The hit rate (HR) is denoted for the combined triplicate experiments. B) L_C (left) and F_R (right) distributions for WT and E46K. The modal L_C values as calculated from Gaussian fits of the data are shown inset. E46K has a longer modal L_C (48.4 nm) than that of WT (29.5 nm). Modal F_R values were calculated in the same way yielding similar F_R values for WT (48.9 pN) and E46K (43.5 pN). C) Box plots showing the distributions of L_C (left) and F_R (right). Median values are shown below the plot. Whiskers represent minimum and maximum values. All experiments were carried out at 20 mM Tris, 200 mM NaCl pH 7.5.

The L_C distribution for E46K is also much wider which may suggest the presence of more than one interaction site and/or more than one dimeric conformation. The hit rate is also different between the proteins variants, 3.0 and 5.9 % for WT and E46K respectively which may indicate that E46K has a higher dimerisation affinity than that of WT. The F_R of E46K (48.9 pN) is similar to that of WT (43.5 pN).

The single point mutation E46K, which only causes subtle changes in the monomeric conformation (see Section 4.4), causes dramatic changes in the ensemble aggregation assays (Figure 4-11) as well as causing early onset PD *in vivo*. The mutation causes significant changes in the dimerisation of the αSyn as revealed by SMFS experiments here, possibly showing that the conformation or dimeric interaction interface is different in E46K relative to WT. This could indicate that dimerisation is a key step for this PD mutant and could provide a selective target for therapeutics aimed at this mutant.

4.7.1 Contour length simulations

As discussed in section 4.6.3, experimentally observed L_C values can be compared to modelled values in order to gain more information about the interaction region in the dimer. As discussed above, the SMFS data for E46K differs from that of WT. Dimerisation interactions for E46K display a longer contour length than that of WT. Both the modelled (coloured histograms) and experimentally observed L_C values (black line) are shown in Figure 4-22. The experimental L_C distribution of E46K is wide relative to WT indicating that the interactions for E46K are more heterogeneous than that of WT. Comparisons to the modelled data suggest an interaction region in the dimer is primarily in residues 60-130, from the NAC region extending into the C-terminal region (assuming that αSyn adopts an intrinsically disordered conformation).

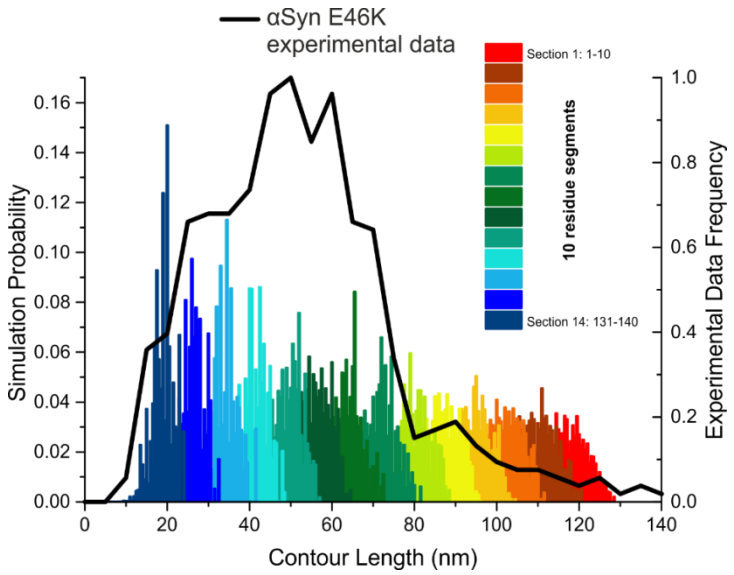


Figure 4-22. L_C modelled data for αSyn E46K A140C SMFS experiments with overlaid experimentally observed data at 20 mM Tris 200 mM NaCl pH 7.5. Modelled L_C data is shown in histograms coloured from blue to red in order of increasing L_C. Each individual histogram of modelled data simulated a 10 residue sequence of αSyn E46K interacting with the corresponding sequence in the second monomer in the dimeric interaction. This was carried out throughout the protein sequence in sequential, non-overlapping, 10 residue sequences. The simulation probability is shown on the left y-axis. Experimentally observe L_C data is shown as a black line depicting the outline of a histogram (bin size 5nm). The normalised frequency of experimental data is shown on the right hand y-axis.

As discussed above, this however, may not be the case as the model treats the proteins as essentially random coils absent of structure. As previously discussed, it is well established that αSyn cannot be treated in this way. I have also shown evidence here from SAX and CD data that the structural/ conformational properties of E46K and WT monomeric proteins show subtle difference which may lead to differences in the mechanism of the dimeric interactions in these proteins.

4.8 Single molecule dimerisation of synuclein homologues

Data above suggest that the interaction region detailed here is not the NAC. To investigate this surprising result, SMFS experiments were also performed on β- and γSyn. As discussed in Section 1.5.5, αSyn is a member of a small family of highly homologous proteins that include β- and γSyn. These proteins have a much lower aggregation propensity, if they aggregate at all[274,275], and are not linked with PD and other synucleopathies[273]. Importantly, both β- and γSyn have smaller, less aggregation-prone NAC regions than αSyn, the NAC region in βSyn is also smaller and lacks the key central 12 residue sequence (71-82, the subject of investigation in Chapter 3) that has been shown to be both necessary and sufficient for fibril formation[204].

SMFS was carried out on the synuclein homologues in the same way in which force experiments were carried out on αSyn, immobilising the proteins at their C-termini. The data for β and γSyn, as with WT is the sum of triplicate experiments. Each individual experiment for β and γSyn consisted of at least 5 force maps of 500 approach-retract cycles. The overall number of approach retract cycles for the experiment in triplicate was 9500 and 7500 for β- and γSyn respectively.

The data for these SMFS experiments are shown in Figure 4-23. Scatter plots are shown for each of α, β and γSyn (blue, grey and orange respectively) in Figure 4-23 A. The data shows that all members of the synuclein family dimerise in SMFS experiments (validated by previous MS experiments, Section 4.4.2) as each contains an obvious hotspot in the data. The interactions appear very similar when judged on the scatter plots, with similar hit rates of 3.0, 4.7 and 3.3 % for α, β and γSyn respectively. The L_C distributions are displayed as histograms in Figure 4-23 B with modal values displayed inset calculated from Gaussian fittings, plus median values displayed in box plots of the L_C distributions (Figure 4-23 D). The F_R distributions are also shown to be similar in modal values and spread of the data (Figure 4-23 and Figure 4-23 C and D). The similarity of the distributions suggest a common dimerisation interactions of each of the synuclein variants. Interestingly, this is despite the different aggregation propensities in the members of the synuclein family. It can, therefore, be proposed that the dimerisation interactions of αSyn detailed by SMFS are not pathological interactions on pathway to aggregation. It has previously been proposed that αSyn may exist *in vivo* as physiological multimers[46,255-258] that, when disrupted, increase toxicity. This hypothesis will be explored in more detail in subsequent chapters.

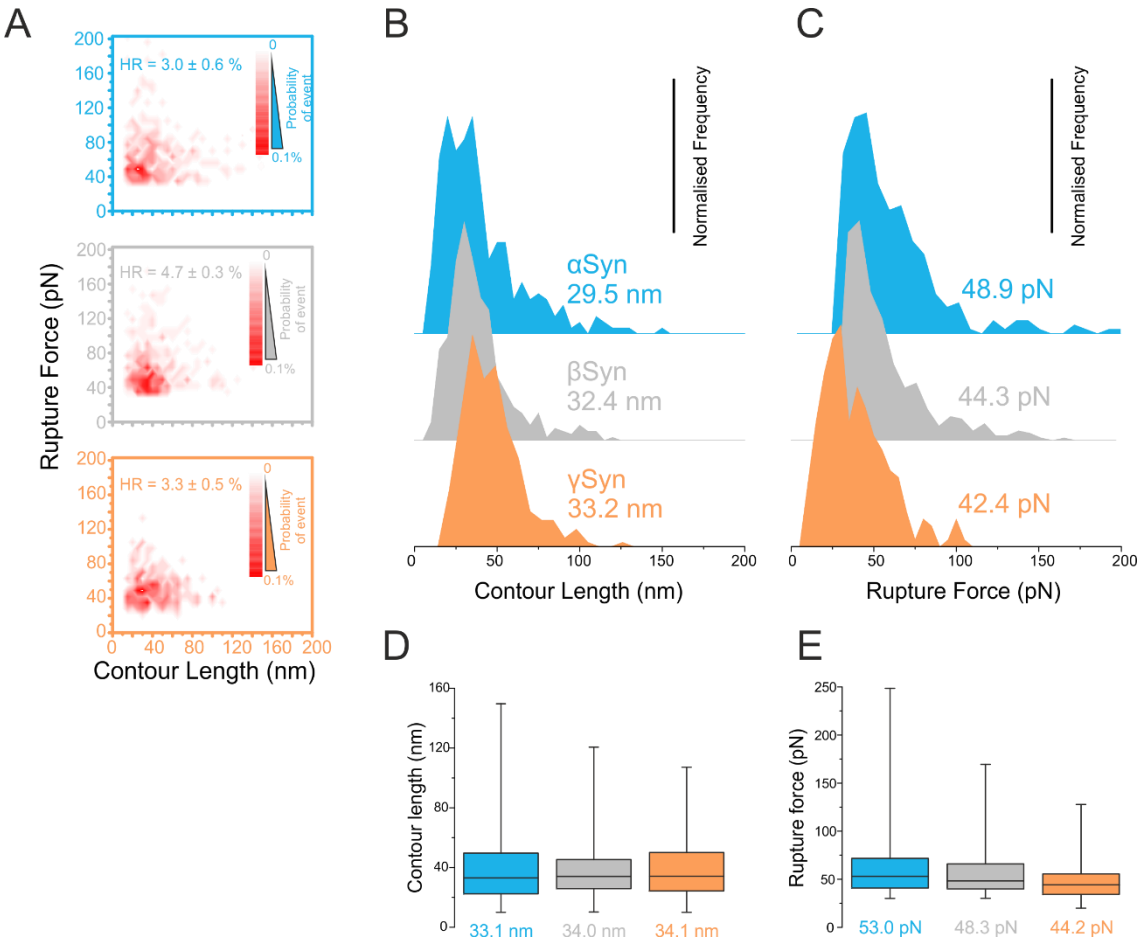


Figure 4-23. SMFS experiments of synuclein homologues. A) Scatter plots of SMFS triplicate experiments for αSyn A140C (blue) βSyn A134C and γSyn D127C showing an obvious hotspot of data in each construct. The points were binned by increments of in L_C of 5 nm and F_R of 7 pN. The data was normalised to the total number of approach-retract cycles in the experiment (total 10,000, 9500 and 7500 for α, β and γSyn respectively). Deeper colours denote higher frequency of events. The hit rate (HR) is denoted for the combined triplicate experiments. B) αSyn (blue), βSyn (grey) and γSyn (orange). The modal L_C values as calculated from Gaussian fits of the data are shown inset. D) and E) Box plots of L_C and F_R distributions, respectively. Medium values are shown below each plot. Whiskers show minimum and maximum values. All experiments were carried out at 20 mM Tris, 200 mM NaCl pH 7.5. Colours are the same throughout.

4.9 Discussion

In this chapter, the structural properties of monomeric αSyn and its aggregation propensity was explored. These data were also compared to that of an aggregation-prone, familial PD αSyn mutant αSyn E46K, as well as non-aggregation-prone human synuclein homologues β and γSyn. It was confirmed via SAXS, CD and analytical SEC studies that αSyn does exist primarily as a IDP in the conditions studies (20 mM Tris, 200 mM NaCl pH 7.5), as is well established in the literature[175]. However, the results from SAXS experiments confirmed that the R_g of αSyn (40.5 Å) is in fact significantly smaller than the calculated value for a completely unstructured

random coil of the same length (52 Å), as have been previously shown[210]. This indicates that αSyn may have some degree of structure in an essentially disordered protein. SAXS and CD studies showed that the PD familial mutant, αSyn E46K, is very similar to that of WT but indicates subtle structural and conformational differences.

EOM data analysis further analysed the ensemble SAXS data and showed that αSyn may exist in two distinct conformations, a more collapsed form, centred around 35 Å, and a more expanded conformation at around 50.3 Å. Similar results were obtained for αSyn E46K but differences in the relative proportions of collapsed and expanded conformations were observed. These subtle difference may be key to the intrinsically different aggregation propensity for these different variants and their subsequent links to disease.

Aggregation experiments revealed the importance of solution conditions on the aggregation propensity of αSyn. It is clear that the relative charge of αSyn plays a highly important role. Modulating the relative charge of αSyn with either changes in pH or salt concentration changes the aggregation rates significantly. The importance of charge is also evidenced by the increased aggregation propensity of αSyn E46K when compared to WTαSyn. A single charge mutation increases aggregation propensity by over 3 times based on lag time analysis. Moreover, SMFS experiments revealed that there is a significant difference in the dimerisation of E46K and WT αSyn at the single molecule scale. The modal L_C values increase significantly from 29.5 nm for WT to 48.4 nm for E46K which may report on a difference in conformation or interaction region of the dimer.

Importantly, it was observed from SMFS studies that the dimerisation interactions of homologues are very similar to that of WT αSyn. This is despite the differences in the NAC region between these proteins. The evidence, therefore, suggests that this is not the predominant interaction region. SMFS revealed, however, that the L_C distributions of E46K are significantly different to WT (Figure 4-21 and Figure 4-22, 29.5 nm and 48.4 nm for WT and E46K, respectively). This could suggest that the conformation of E46K dimers or/and the interaction interface are different for E46K relative to WT. The modelling approach would locate a proposed interaction interface of WT primarily in the C-terminus of the protein (residues 100-140). The interaction region for E46K, importantly, extends into the hydrophobic NAC region (60-130). It must be noted that these modelled interaction interfaces are estimates based on the assumption that the proteins interact as intrinsically dishoarded chains. The increased aggregation propensity in E46K may therefore be attributed to the face that the protein can form dimers via the NAC region, shown to be the core of amyloid fibrils[100].

The data presented in this chapter may therefore indicate that the dimerisation interaction of WT probed in these experiments is therefore not linked to aggregation and may therefore be physiological. Indeed it has been observed previously that αSyn may exist *in vivo* as structured multimers, proposed to be predominantly tetramer, but also showed significant amounts of dimer. Disruption of multimeric interactions by mutations in the N-terminus of the protein led to a decreased multimer: monomer ratio which correlated to an increase in toxicity[46,255-259]. This observation is further corroborated in our SMFS studies in that E46K changes the 'physiological' dimerisation of αSyn. This hypothesis will be further interrogated in subsequent chapters.

5 Novel dimeric conformations revealed by SMFS

5.1 Abstract

In the previous chapter, it was demonstrated that bone fide single molecule interactions can be accurately detected in SMFS studies. There is, however, uncertainty in the interpretation of L_C data as differences in these values may arise due to the presence of different interaction interfaces and/or different protein conformations. Here we address this issue by utilising different immobilisation regimes and interrogating these data using with L_C simulations. The data reveal that dimeric αSyn contains some degree of force-resistant structure. A similar SMFS approach was used to analyse the differences in dimerisation under acidic conditions. These experiments revealed a more collapsed conformation of the dimeric αSyn species. The data suggest that this collapse is localised in the acidic, C-terminal region of the protein.

5.2 Introduction

It has previously been shown that αSyn exists as an ensemble of conformations[34,35], whose structural organisation depends on the environmental conditions[210]. These, often subtle, differences in protein structure can have a significant effect on aggregation propensity[210] and therefore can modulate disease pathology. As discussed previously, αSyn has been shown to have a smaller R_g than that of a theoretical unstructured polypeptide chain of the same length[210], suggesting compaction. It is also well established that αSyn forms a variety of intramolecular, transient interactions[212-218] which may also play a role in the contribution of structure in the IDP. These studies were carried out on the monomeric protein and it is conceivable that upon dimerisation, there is an increase in structure as the conformational space and the internal free energy reduces. The data presented in this chapter describe the interrogation of force-resistant structure in the dimeric species of αSyn.

As noted in the previous chapter, pH changes significantly affect the aggregation propensity of αSyn, as expected for an IDP that is highly receptive to changes in the environmental conditions. Decreasing pH generally increases the rate of fibril formation[330,363] (Chapter 4). It has also previously been shown that αSyn becomes more compact and structured in acidic conditions[210]. It is important to consider how αSyn is affected by pH as different pH environments are present *in vivo* to which αSyn is exposed. Specifically, the acidic environment of the lysosome (pH 4.5) plays an important role in the cellular trafficking of αSyn in the endocytic pathway in the constitutive process of proteostasis. The chaperone mediated autophagy (CMA) arm of the endocytic pathway is particularly important in the cellular

processing of αSyn[331]. The process involves the interaction of a pentapeptide recognition motif, which is present in αSyn ([95]VKKDQ[99]), with a cytosolic chaperone and subsequent transport to the lysosome where the protein is degraded[332]. The importance of lysosomal impairment in PD is supported by the identification of genetic associations between PD and Gaucher's disease, the most common lysosomal storage disorder, which is caused by mutations in the gene encoding the lysosomal hydrolytic enzyme glucocerebrosidase[365]. Studies have shown that 5-10% of PD patients possess glucocerebrosidase mutations[365]. The self-association of αSyn at lysosomal pH is therefore highly relevant and important to study.

5.3 Contour length distributions suggest structure in the dimer

5.3.1 SMFS using different immobilisation regimes

The presence of structure in each monomer at dissociation in SMFS experiments makes identification of the interaction region difficult. In order to interrogate the dimeric interaction further, different αSyn Cys mutants were engineered using the New England Biolabs Q5© mutagenesis protocol (Section 2.2.1.2). Cys residues were substituted in each of the 3 domains of αSyn, the N-terminal (A18C), NAC region (A90C) and at the C-terminal (A140C, the immobilisation regime used in the previous chapter). Moving the point at which αSyn monomers are immobilised on AFM cantilever tips and surfaces, may reveal the interaction region together with the locations of any conformational changes.

SMFS experiments were carried out on each of the different αSyn Cys mutants (Figure 5-1) in the same experimental setup as described in the previous chapter and Section 2.2.4. The scatter plots (Figure 5-1 A) and contour length distributions (Figure 5-1 B) from SMFS experiments for all 3 immobilisation regimes at neutral pH (20mM Tris, 200mM NaCl, pH 7.5) are presented in Figure 5-1. Differences are observed in the L_C distributions of dimeric interactions of αSyn immobilised at different positions. It is intuitive that if the immobilisation point of αSyn changes, the L_C should also change. The change in L_C between the different cys mutants provides further validation that we are looking at a distinctive protein-protein interaction.

The data suggests that the interaction region is closer to the C-terminus than to the N-terminus as the L_C values are smaller for A90C (23.4 nm) and A140C (29.5 nm) than to A18C (40.1 nm). However, in order to get a more accurate picture of the interaction region, experimental data can be compared to simulated data as discussed in the previous chapter, these data are presented in subsequent sections here.

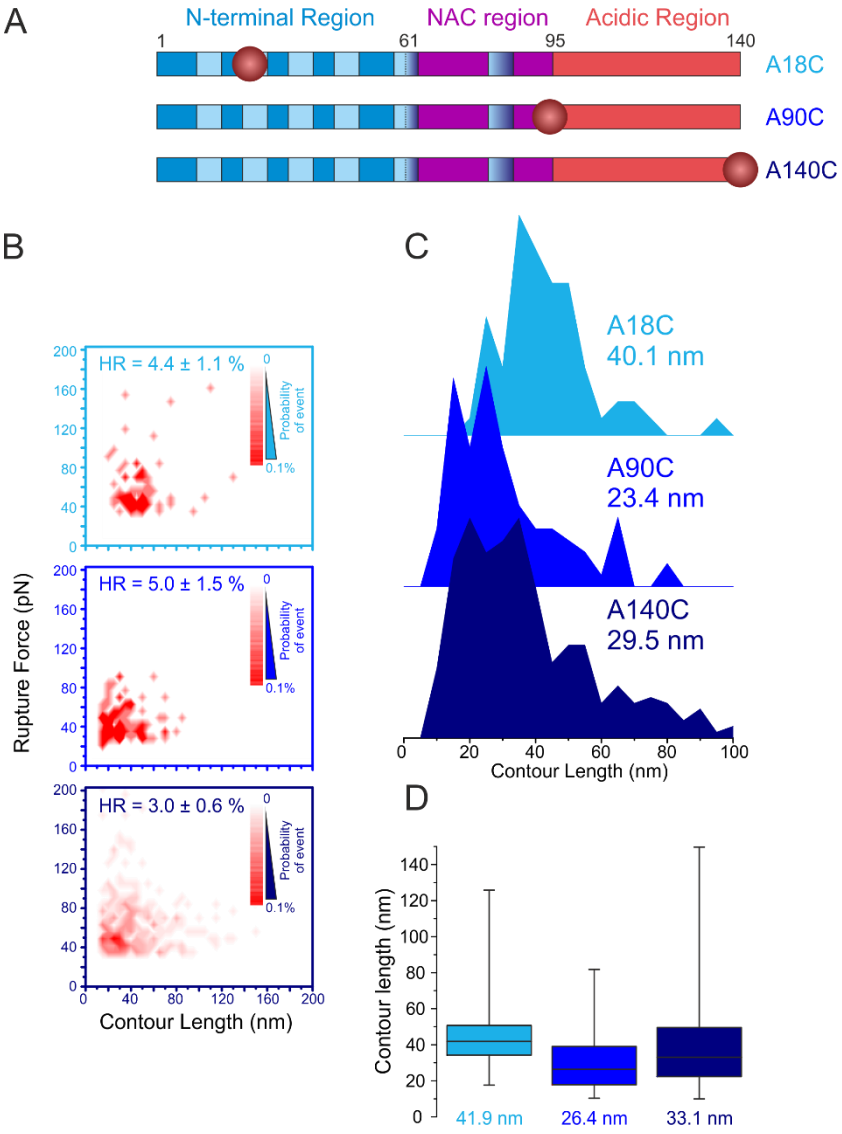


Figure 5-1. SMFS experiments of different αSyn cys mutants using different immobilisation regimes. A) Schematic representation of the αSyn sequence with primary regions shown (blue, pink and red represent the N-terminal region, NAC and acidic C-terminal region, respectively. Imperfect repeats are represented in light blue). Immobilisation positions are represented by red circles. B) Scatter plots of SMFS from A18C, A90C and A140C (lightest blue to darkest blue respectively). The points were binned by a contour length of 5 nm and a rupture force of 7 pN. The data was normalised to the total number of approach-retract cycles in the experiment (total events: A18C = 2000, A90C = 2000, A140C = 10,000). Deeper colours denote higher frequency of events. The hit rate (HR) is denoted inset. C) L_C distributions for A18C (light blue, top panel), A90C (mid blue, middle panel) and A140C (dark blue bottom panel). The modal L_C values as calculated from Gaussian fits of the data are shown inset. Each Cys mutant has a different L_C distribution with different modal values. Median values are shown below the plot D) Box plots representing the distribution of the L_C data. Whiskers show minimum and maximum values. All experiments were carried out at 20 mM Tris, 200 mM NaCl pH 7.5.

1.1.1.1 Contour length simulations

Contour length simulations were carried as discussed in the previous chapter. In all, 14 separate L_C simulations were performed; each modelled a 10 residue sequence of αSyn

interacting with the corresponding sequence in the second monomer in a dimeric interaction. This was carried out throughout the protein sequence in using non-overlapping, contiguous, 10-residue sequences. This was carried out for each of the three Cys mutants. The simulated L_C distributions are presented in Figure 5-2 and experimental distributions are overlaid (Figure 5-2 A) for SMFS experiments carried out at physiological and acidic pH. The SMFS data at pH 4.5 will be discussed in subsequent sections in this chapter. The same simulated L_C distributions are shown in Figure 5-2 B with offsets between individual simulated sections. The data are presented in this manner as there is considerable overlap of histograms from different simulated, interacting sections when the immobilisation point is at a central point as opposed to a proximal point in the protein sequence. This also acts to increase the ambiguity of experimental data as the same distributions of experimental data can occur from a much larger proportion of the protein interacting, as is the case for the data presented here.

For A140C, comparison of the experimentally determined data (at pH 7.5, light grey line in Figure 5-2 B) to the simulated data would suggest that the interacting region in the dimer is in the C-terminus at around residues 100-140. Making the same comparisons for A18C and A90C, however, suggests different interaction regions (residues ≈ 40-70 for A18C and ≈ 40-140 for A90C). All distributions are relatively wide, this may indicate that the interactions are heterogeneous and can 'slide' up and down the protein. Another possibility is that there is some conformational heterogeneity in the dimer observed in SMFS experiments. The increased ambiguity of the interaction region (more than 70% of the protein) in A90C is a result of the immobilisation point being in the middle of the protein as highlighted in Figure 5-2 B.

Comparing the simulated data to experimentally determined data should yield the same interacting regions for the different αSyn Cys mutants as in these experiments the protein remains the same, but is immobilised at different points. If A18C and A140C are taken as an example, this is not the case as the interacting regions as derived from the comparisons of simulated and experimental data, are different. The modelled data assumes that each monomer is devoid of structure at rupture. As discussed in the previous chapter, αSyn cannot be assumed to be a completely unstructured polypeptide (shown in data in this thesis and previously by other groups[210,212-215,217,218]). The discrepancy of the L_C data may indeed suggest that αSyn cannot be modelled in this way. The data indicates, therefore, that there is some force resistant structure in the dimer at physiological pH.

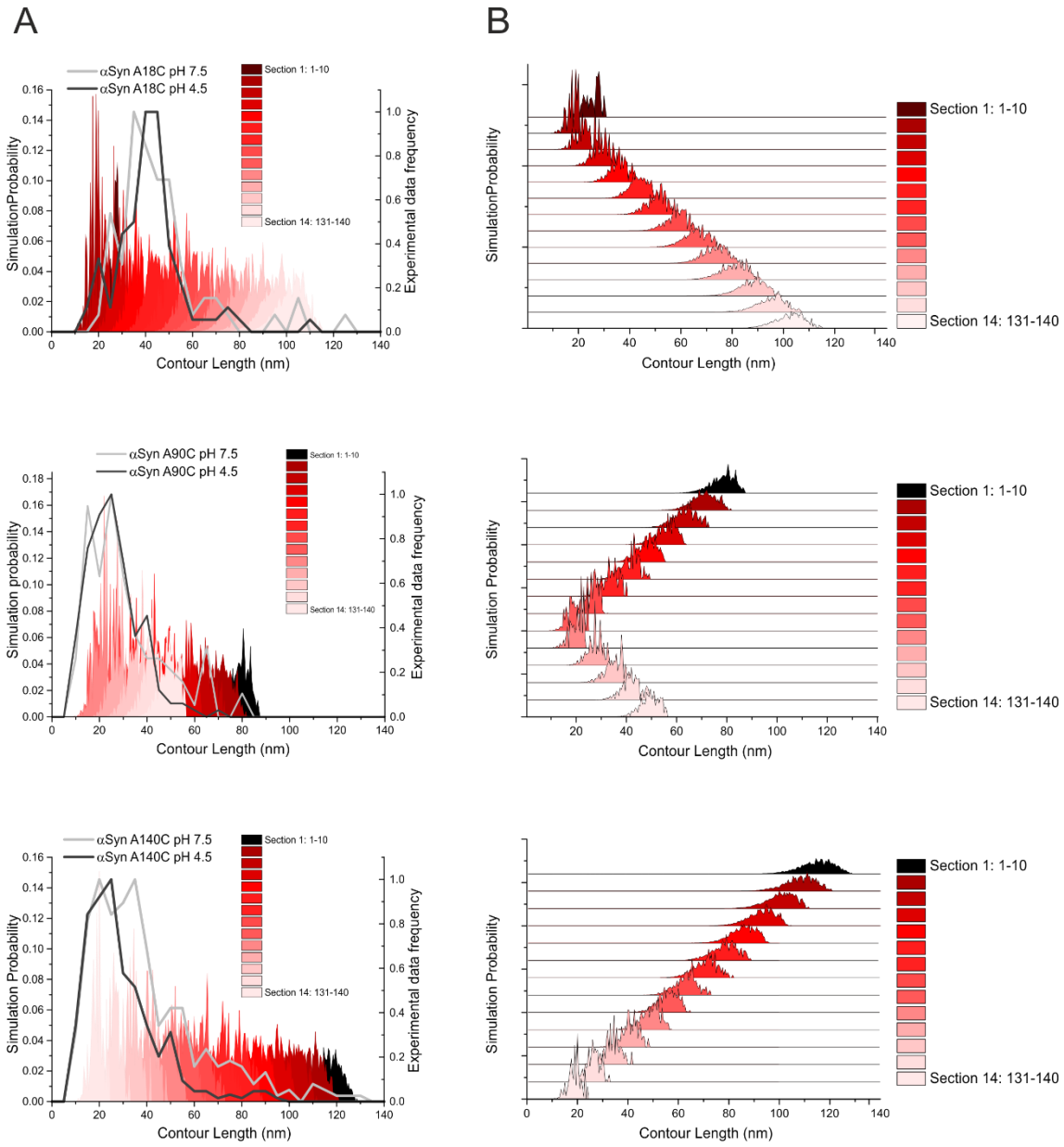


Figure 5-2. L_C modelled data of SMFS experiments carried out on immobilisation point variants with overlaid experimentally observed data. A) Modelled L_C data are shown in histograms coloured from dark red to light red from the N-terminal to the C-terminal of the protein respectively. Each individual histogram of modelled data simulated a 10 residue sequence of αSyn interacting with the corresponding sequence in the second monomer in the dimeric interaction. This was carried out throughout the protein sequence in sequential, non-overlapping, 10 residue sequences. The simulation probability is shown on the left-hand y-axis. Experimentally observed L_C data are shown as solid lines depicting the outline of a histogram (bin size 5nm). Dark grey and light grey denote experiments at acidic and neutral pH. The normalised frequency of experimental data is shown on the right hand y-axis. B) Offset simulation histograms in order to increase clarity with regards to overlapping interaction segments.

The difference in modal L_C values between A90C and A140C is 6.0 nm (Figure 5-1). This distance equates to 15 residues when taking a single residue as 0.4 nm. Therefore, there is a

35 residue discrepancy between these two immobilisations. It is possible that this can be attributed to the formation of structure in this region.

5.4 The key role of pH on αSyn and its dimeric interactions

5.4.1 Biophysical analysis upon acidification of pH

5.4.1.1 SAXS studies at acidic pH

In the same way as described in the previous chapter, SEC-SAXS was utilised to analyse the structure and the conformation of monomeric αSyn used in this study (Section 2.2.3.3). For comparison, a familial PD variant: E46K was also analysed using this method.

SEC-SAXS experiments were carried out in 20 mM acetate, 200 mM NaCl pH 4.5 (to mirror SMFS conditions used in this chapter) (Figure 5-3). As for SAXS data at pH 7.5 (see previous chapter), the scattering data for pH 4.5 are similar for both WT αSyn and E46K (Figure 5-3 A). The Kratky plots of both of these variants again show data typical of an IDP (as with pH 7.5) as there is an absence of an obvious peak and a continuous rise in the data (Figure 5-3 B). By performing a Guinier approximation (Figure 5-3 C and D), the R_g was found to be 36.5 and 34.9 Å for WT and E46K respectively. These R_g values are smaller than that recorded at pH 7.5 (41.1. and 41.5 Å for WT and E46K respectively), reflecting previous findings that acidification causes conformational changes in αSyn such as its compaction[210]. Previous SAXS studies indeed showed that αSyn had a smaller R_g at acidic pH of 30 Å, this is less than observed in the scattering data presented here, but, at a slightly lower pH (pH 3.0)[210]. Moreover, it has been shown that this compaction is primarily due to compaction of the highly acidic C-terminus of the protein[220,221]. This data suggests that at pH 4.5, αSyn remains primarily unstructured as previous studies have shown[220], it does however, show some compaction which may suggest a gain of some rudimentary structure. The E46K mutant possessed a similar R_g value to that of WT which suggests that a similar conformational rearrangement occurs in WT and E46K.

EOM analysis was subsequently carried out on ensemble scattering data in the same manner as discussed in the previous chapter. The optimised structures that best fit scattering data were selected (Figure 5-4 A solid lines). The R_g distributions for the EOM ensemble species are shown in Figure 5-4 B. The distributions at pH 7.5 showed an obvious bimodal distribution of both collapsed and expanded species for both WT and E46K. This is not the case, however, at acidic pH. There is some heterogeneity present in the R_g distributions of the synuclein variants at pH 4.5, which can be fit as separate species. However, these 'species' are much closer in size

and don't display the obvious collapsed or expanded conformations shown in the EOM data for synuclein variants at neutral pH.

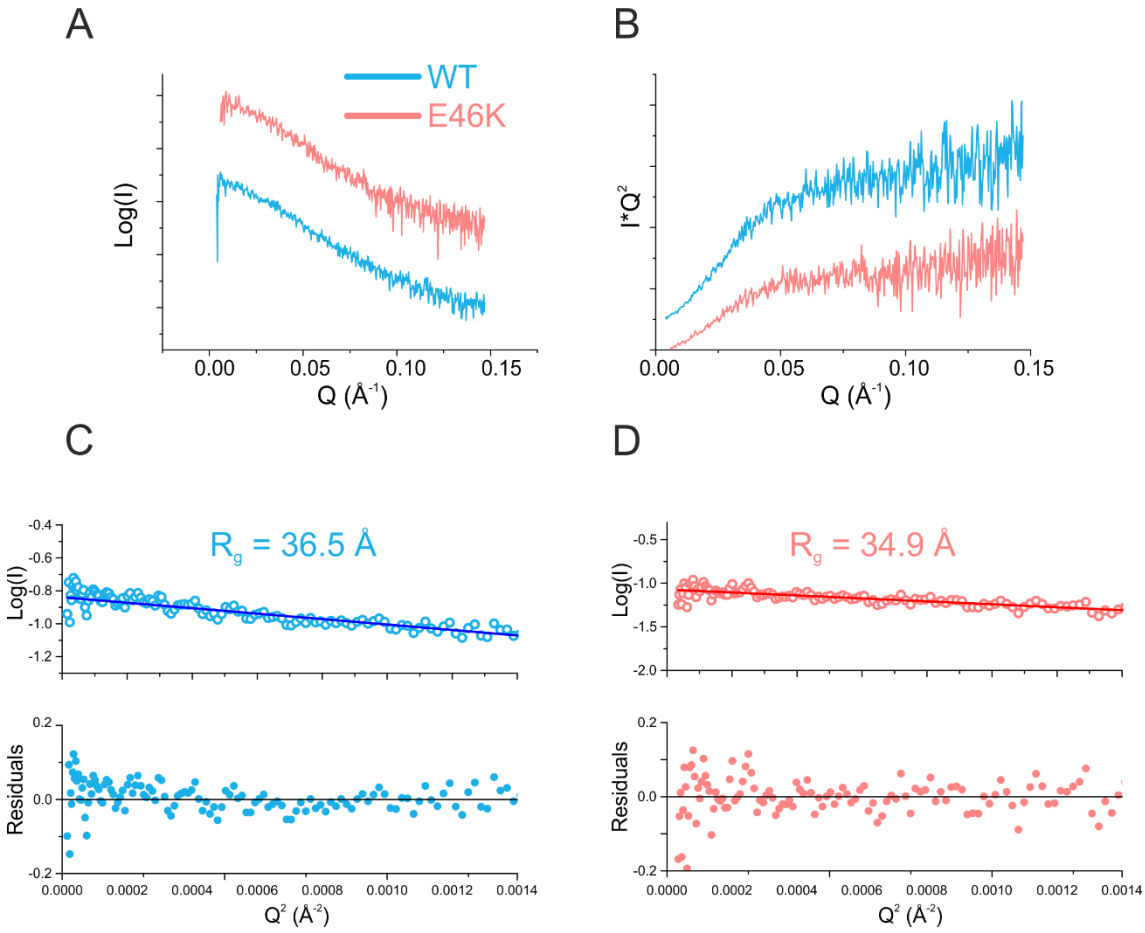


Figure 5-3. SEC-SAXS analysis of αSyn variants WT and E46K at pH 4.5. A) Scattering data of WT (blue) and PD familial mutant E46K (red) displayed in a conventional semi-logarithmic plot. B) Kratky plot of experimental data containing features of typical of an IDP with a continuous rise and an absence of a clear maximum peak. Plots in A) and B) are offset for clarity. C) and D) Guinier plots of E46K and WT αSyn fitted with a linear regression. R_g values were estimated from a Guinier approximation to be 36.5 and 34.9 Å for WT and E46K respectively.

The EOM R_g distributions were fit to multiple Gaussians and the modal R_g values calculated (Table 5-1). Both WT and E46K possess two similar R_g populations at around 30 and 38 Å. The WT protein has an extra expanded peak at 51.9 Å that is absent in E46K. This expanded peak is present at a higher population in both WT and E46K at pH 7.5. This data suggests that the collapse of synuclein upon acidification is due to the reduction in population of the expanded species.

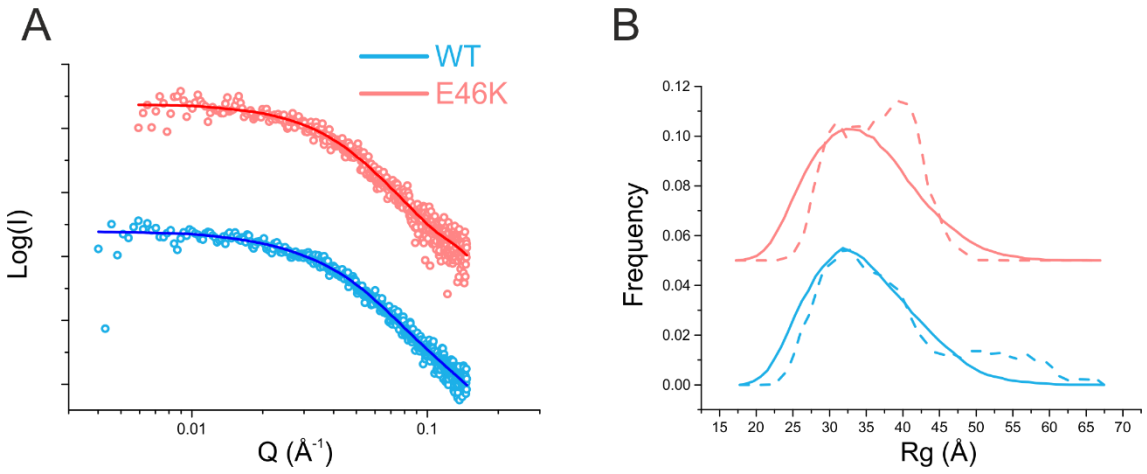


Figure 5-4. EOM analysis of SEC-SAXS data at pH 4.5. A) log-log plot of scattering data with EOM fits (solid lines) for WT and E46K αSyn (blue and red respectively). B) EOM R_g distributions. The distributions for the random pool (solid line) and those from the EOM-selected ensemble (dashed line) are shown.

Protein variant	**R_g (Å) Peak 1**	**R_g (Å) Peak 2**	**R_g (Å) Peak 3**	**Peak 1 %**	**Peak 2 %**	**Peak 3 %**
WT	30.8	38.0	51.9	44.0	31.7	24.2
E46K	31.0	39.3	-	38.0	62.0	0

Table 5-1. Data from EOM selected ensemble R_g distributions based on Gausian fits of EOM data. % peaks are calculated from the area of Gaussian fits.

1.1.1.2 CD analysis at acidic pH

Far-UV CD was carried out in order to interrogate structural changes of αSyn as a function of pH. CD spectra were recorded for both WT and E46K (Figure 5-5 A and B respectively. At both pH 4.5 and 7.5, similar characteristic spectra are observed for WT αSyn of mostly unfolded polypeptide chains with an absence of bands in the 210-230 nm region for both proteins. There are, however, differences in the intensity of the spectra between neutral and acidic pH values for WT αSyn. There is an increase in negative intensity at around 222 nm at pH 4.5 relative to pH 7.5, which may denote an increase in secondary structure at acidic pH. This has been reported previously for αSyn[210].

This gain in structure upon acidification was not observed for E46K (Figure 5-5 B). CD experiments were carried out in 20 mM Tris, 200 mM NaCl, pH 7.5 and 20 mM acetate, 200 mM NaCl pH 4.5 (to mirror SMFS conditions used in this chapter, see below) and so the spectra don't include wavelengths less than 205 nm due to significant noise in this region due to a high concentration of Cl^- ions. These results suggest that E46K is resistant to secondary structural changes upon acidification whereas WT is not. The CD results presented in the previous

chapter showed that E46K may have an increased structure relative to WT at neutral pH. Taken together, these results suggest that E46K forms some structure at pH 7.5, and it may be possible that this structure doesn't change with pH, unlike WT αSyn. These structural nuances may play a role in the difference in the aggregation propensities of the proteins and their corresponding association with disease progression/ severity.

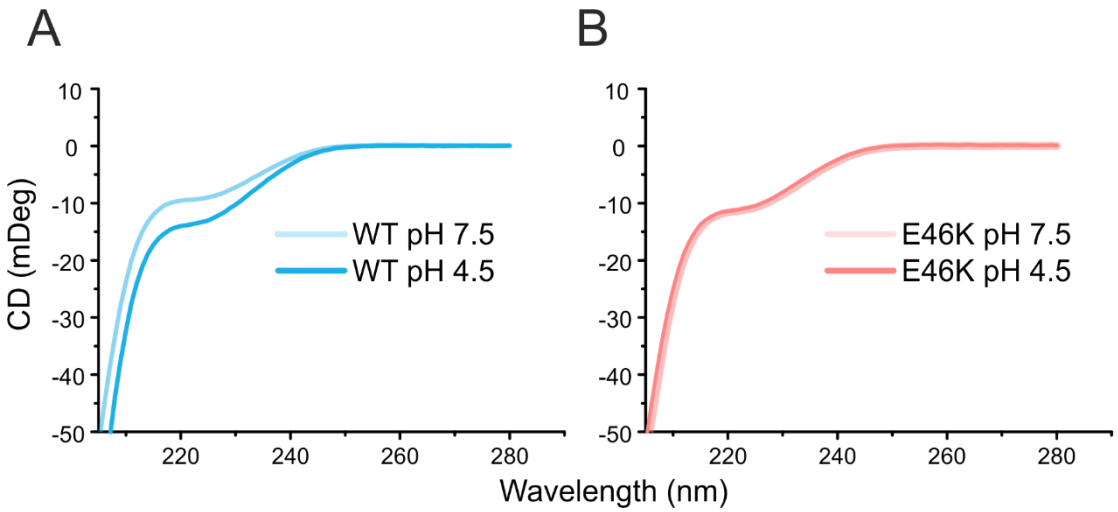


Figure 5-5. Far-UV CD spectra of WT αSyn and αSyn E46K at neutral and acidic pH. A) Spectra of WT αSyn at pH 7.5 (light blue) and 4.5 (dark blue). The spectra are characteristic of a random coil structure typical of an IDP. The WT spectra show slight differences at pH 4.5, there is an increase in negative intensity at wavelengths around 222 nm which may suggest a gain in secondary structure upon acidification. B) Spectra of αSyn E46K at pH 7.5 (light red) and 4.5 (dark red). The spectra are characteristic of a random coil structure typical of an IDP and are identical at the two pH values. The proteins were at 100 µM in 20 mM Tris 200 mM NaCl pH 7.5, the same conditions as SMFS experiments (see below). The spectra are displayed from 205 nm onwards as at lower wavelengths there is significant noise attributable to the high concentrations of salt in the buffer conditions.

5.4.2 Aggregation of αSyn at different pH values

Fibril growth assays were carried out on both WT and E46K αSyn under both neutral and acidic conditions (light and dark blue, respectively) (Figure 5-6). As discussed in the previous chapter, the aggregation of WT αSyn is significantly faster at pH 4.5 (lag time = 7.3 ± 0.5 hours) than pH 7.5 (lag time = 52.4 ± 7.7 hours) (Figure 5-6 A). This trend is less obvious for E46K (Figure 5-1 B). Although the aggregation propensity is higher in acidic conditions (dark red) (lag time = 13.2 ± 1.8 hours) relative to neutral conditions (light red) (lag time = 16.3 ± 0.6 hours), E46K is much less affected by a change in the pH than WT. These data reflects the same trend as observed in CD experiments shown above in that E46K is resistant to structural changes upon acidification of the pH whereas WT may gain some functional features upon this change in environment (Figure 5-5). It was also shown in the previous chapter that E46K possesses a

lower minimum value at around 222 nm suggesting more structure in the PD variant at neutral pH. This correlated with a higher aggregation propensity of E46K when compared to WT at pH 7.5. This suggests that a gain in structure as shown by CD experiments caused by a charge mutation or change in pH is linked with increased aggregation.

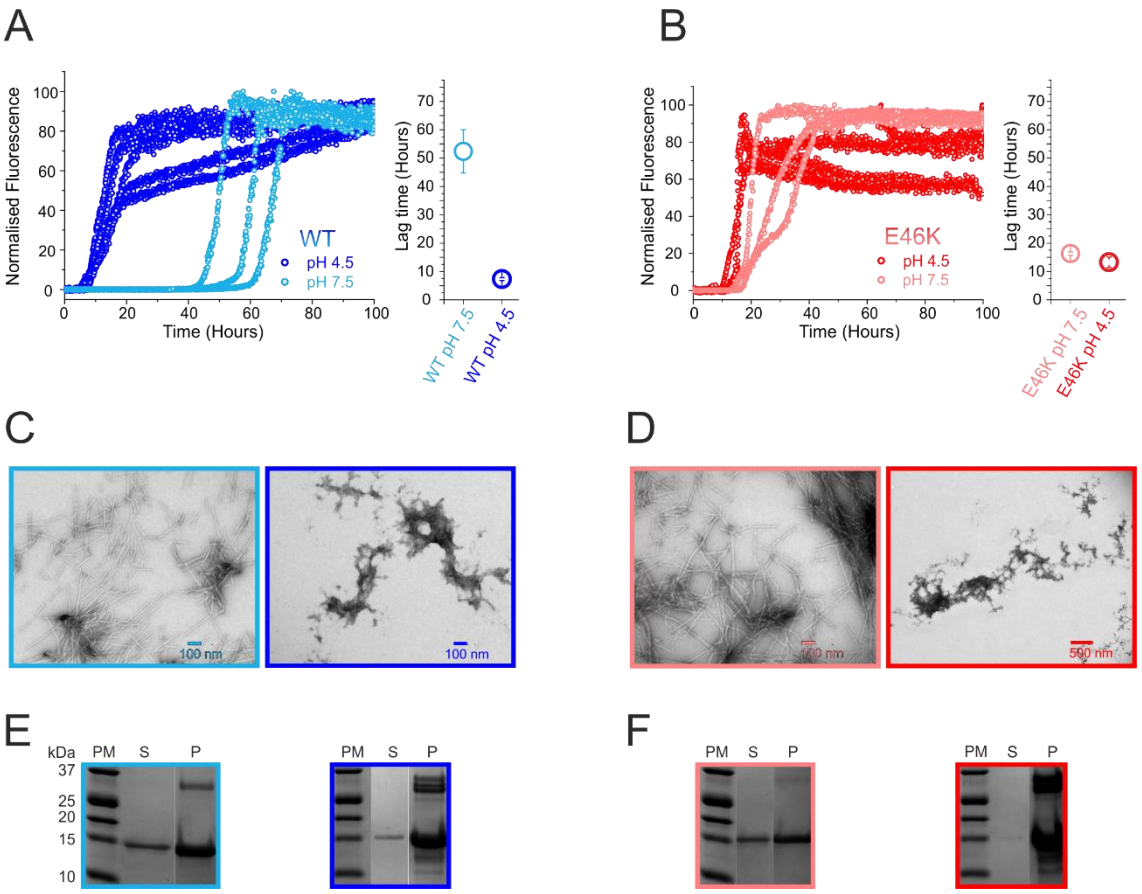


Figure 5-6. Aggregation of WT and E46K αSyn at neutral and acidic pH. A) and B) ThT fibril growth assays of 100 µM A) WT (blue) and B) E46K (red) in either 20mM Tris 200 mM NaCl pH 7.5 or 20 mM acetate, 200 mM NaCl pH 4.5, at 37°C and shaking at 600 rpm. Incubations at pH 7.5 are shown as a lighter colour whereas incubations at pH 4.5 are shown with a darker colour. Colour coding is consistent throughout. Lag times are shown to the right of ThT plots. Average lag times are represented as filled circles, the mean lag time is at the centre of each circle. Error bars show SD. The data show a much stronger lag time dependency on pH in WT than in E46K. C) and D) Negative stain TEM images of aggregated structures at the end-point of ThT incubations. For both variants, fibrillar structures are observed at neutral pH whereas amorphous aggregates are observed under acidic conditions. E) and F) Fibril yields as assessed by SDS-PAGE densitometry analysis. The protein marker (PM) supernatant (S) and pelleted (P) samples are shown. Fibril yields for WT are 79 and 95 % at pH 7.5 and 4.5 respectively. Fibril yields for E46K are 85 and 92 % at pH 7.5 and 4.5, respectively.

The fibril architecture as observed from negative stain TEM, shows fibrillar structures form at neutral pH for both WT and E46K, but, amorphous-like aggregates are predominantly observed for both protein variants at pH 4.5 (Figure 5-6 C and D). A higher proportion of protein is

present in the insoluble, pelleted fractions for both WT and E46K in acidic conditions (95 and 92 % for WT and E46K respectively) relative to neutral conditions (79 and 85 % for WT and E46K respectively) as assessed by SDS-PAGE densitometry analysis (Figure 5-6 C and D). Gel samples of the insoluble fractions reveal additional bands around monomer and dimeric molecular weight at acid pH which most likely are due to various αSyn truncations. The aggregation data taken together indicates that changing pH plays a significant role in the aggregation of both WT and E46K.

5.4.3 SMFS studies reveal structured dimer at acidic pH

5.4.3.1 Contour length collapse of WT at acidic pH

The effect of acidic pH on the dimerisation of αSyn was then studied by SMFS. Experiments were carried out in 20 mM acetate, 200 mM NaCl pH 4.5 in the same manner as SMFS experiments carried out at neutral pH. The effect of pH on the dimerization of αSyn A140C in SMFS experiments is highlighted in Figure 5-7. The modal L_C distribution reduces from 29.5 nm at neutral pH to 23.5 nm at acidic pH. This may suggest that either the interaction region is changing (closer to the C-terminus in acidic conditions) or there is a change in conformation/ structure of the protein.

As has been reported earlier in this chapter, the conformation and possibly structure of αSyn changes upon acidification of the pH. This is also well established in the literature as discussed in Section 1.5.2. Partial folding of αSyn upon a reduction in pH has previously been proposed[210]. It has also been shown using intermolecular PRE experiments that reduction of the pH induces a collapse of the highly acidic C-terminus[220,221]. This scenario would indeed fit with the pH dependent collapse of the L_C highlighted in the SMFS data here. Moreover the R_F for dimers are similar at both neutral and acidic pH, suggesting that the interaction is the same at both pH values, the collapse in the L_C may therefore be attributed to a conformational change in the dimer.

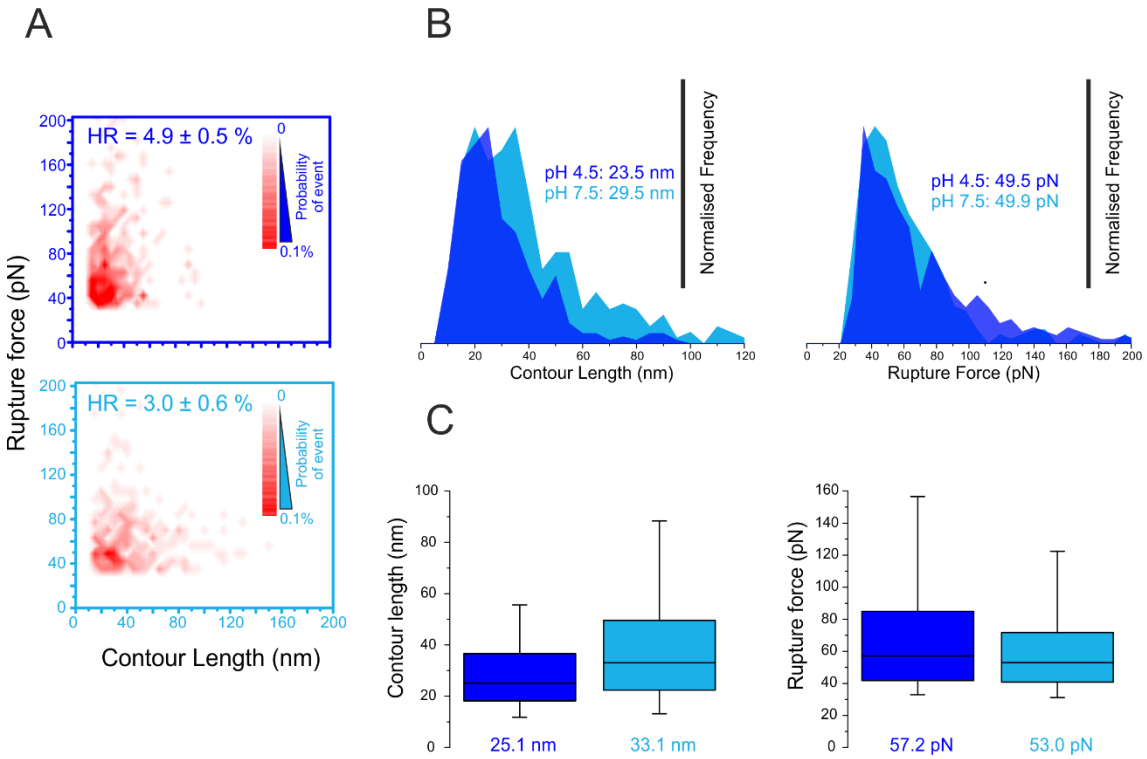


Figure 5-7. SMFS experiments of αSyn A140C at physiological and acidic pH. A) Scatter plots of SMFS from αSyn A140C at pH 4.5 (dark blue) and pH 7.5 (light blue). The points were binned by a contour length of 5 nm and a rupture force of 7 pN. The data were normalised to the total number of approach-retract cycles in the experiment (totals: pH 7.5 = 10,000 and pH 4.5 = 7500). Deeper colours denote higher frequency of events. A 'hotspot' is clearly present in the data for both conditions suggesting a specific single molecule interaction. The hit rate (HR) is shown inset for the combined triplicate experiments. B) L_C and F_R distribution histograms with the same colour coding as in A (bin size 5 nm and 7 pN for L_C and F_R, respectively) displayed as area plots with modal values from Gaussian fittings denoted inset. The data shows that the modal L_C in acidic conditions (23.5 nm) is smaller than that at neutral conditions (29.5 nm). Similar F_R values are observed at both pHs (49.5 and 49.9 pN for pH 4.5 and 7.5 respectively). C) Box plots of L_C (left) and F_R (right) distributions showing the spread of the data. Whiskers show 5th and 95th percentiles. Median values are shown below the plot. Experiments were carried out at either 20mM Tris, 200 mM NaCl pH 7.5 or 20 mM acetate, 200 mM NaCl, pH 4.5.

5.4.3.2 The effect on pH in different immobilisation regimes

SMFS experiments were carried out using different immobilisation regimes as discussed above at lower pH (Figure 5-8) to gain more insight into the region interaction interface and any conformational rearrangement that may occur. The pH dependent collapse of L_C (as discussed above) is only observed in the A140C immobilisation regime; A18C and A90C maintain the same modal L_C value at both neutral and acidic pH (41.1 and 23.5 nm respectively). This suggests that the interaction region of the dimer remains the same at the different pH conditions, but that the L_C collapse observed in A140C is due to a conformational change. Information can also be gained about the location of the proposed structure due to the fact that A18C and A90C are resistant to the pH induced collapse of the L_C. As outlined in the

schematics in Figure 5-8 B, it can be hypothesised that there is some degree of structure induced by a drop in pH in the dimer in the C-terminal region of the protein. The interaction region is closer to the N-terminal relative to the region of proposed structure. The fact that the L_C of A90C at pH 4.5 (23.5 nm) is the same as A140C in the same conditions (23.5 nm) suggests that almost the entirety of the C-terminal region of αSyn (residues 90-140) does not contribute to the L_C at acidic pH. These data suggest that at least these residues are collapsed and form some structure at pH 4.5. It should be noted however that this structure is force-resistant as it doesn't unfold when pulled from the C-terminus (A140C), it could be hypothesised, therefore, that the structure in this region is β-sheet structure.

There is precedent from previous studies by other groups that αSyn does indeed undergo a pH induced collapse of the C-terminus as discussed previously[220,221]. These studies have analysed the monomeric species of the protein. It should be noted that the collapse proposed here is novel in that it has been proposed to occur within the dimeric species of αSyn.

These experimental data can be compared to simulated L_C distributions as carried out previously. The data from different immobilisation regimes in acidic conditions is showed overlaid with L_C simulation data in Figure 5-2. The simulation data treats αSyn as a completely disordered chain, which, as discussed above, has been shown in this thesis not to be the case as there is evidence suggesting the formation of structure in the αSyn dimer. However, given that we are able to use the data here to propose the C-terminus as the location of structure in the dimer, we can interpret data from the A18C immobilisation as more accurately reporting on the interaction interface. This is highlighted in the schematic in Figure 5-8 B top panel. Given that the immobilisation is distal to the structured region, the observed L_C may be used in parallel with the corresponding simulations to predict an interaction interface. This suggests an interaction region in residues 40-70. This is a novel region of proposed dimeric interaction located mostly in the N-terminal and interestingly, mostly not in the highly hydrophobic NAC region of the protein (residues 65-90). It is also interesting that this region contains all but one of the six familial PD mutations. SMFS experiments of E46K presented in the previous chapter revealed different L_C distributions relative to WT which may also indicate the importance of this region. This novel interaction will be discussed in more detail in the next chapter in this thesis.

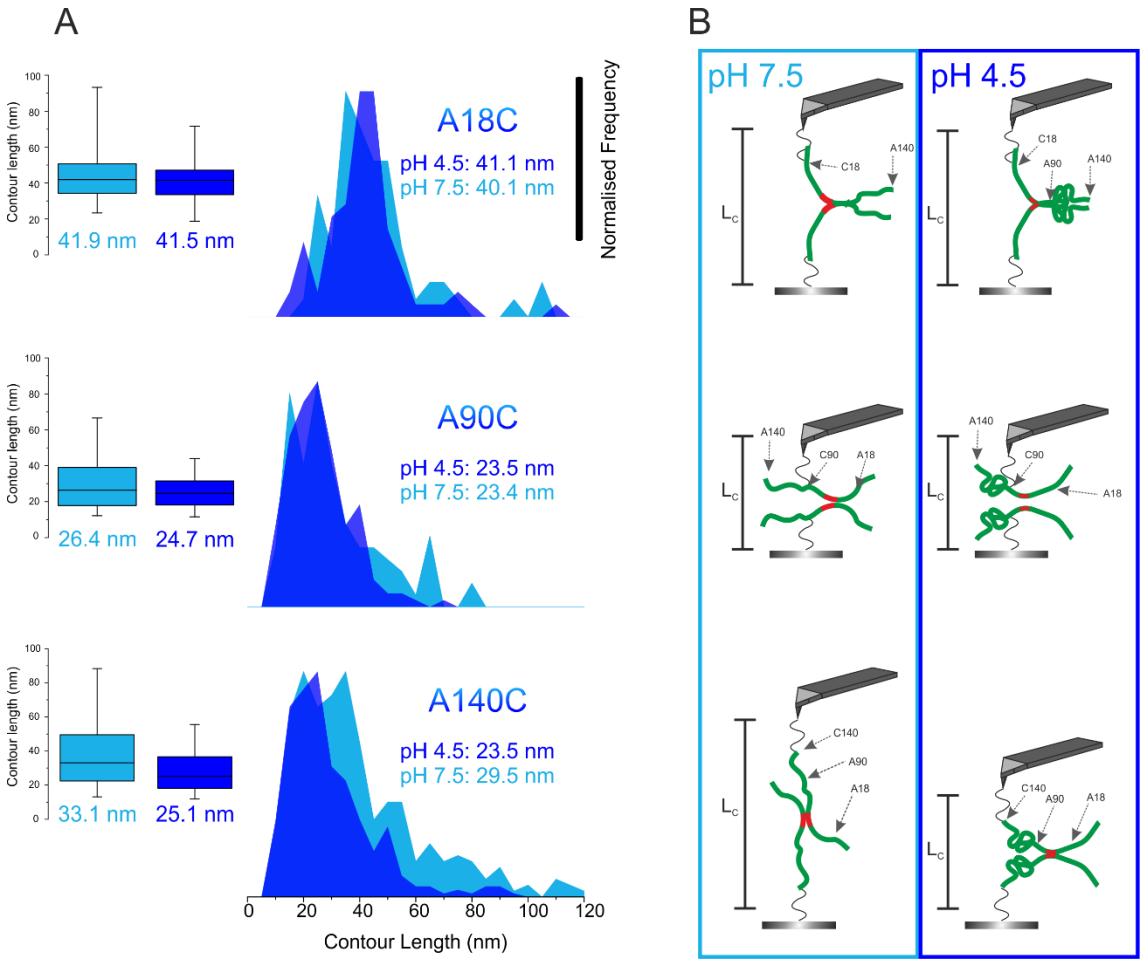


Figure 5-8. SMFS experiments of different αSyn Cys mutants using different immobilisation regimes in both neutral and acidic conditions. A) L_C distribution histograms presented as area plots of αSyn immobilised using different regimes and modal values from Gaussian fits are shown. Box plots of L_C distributions are shown inset. Whiskers indicate the 5th and 95th percentiles. Median values are shown below each plot. Data for conditions at pH 7.5 and 4.5 are shown in light blue and dark blue respectively. Data for A18C, A90C and A140C are shown in top, middle and bottom panels respectively. The points were binned by a contour length of 5 nm. B) Schematics of hypothesised dimeric interactions for each immobilisation regime (different rows) and at both neutral (light blue box, left column) and acidic conditions (dark blue box, right column). The interaction regions are shown in the schematic at red sections. A gain in structure upon lowering of the pH is proposed which results in a lowering of the L_C for A140C, but not A18C or A90C.

5.4.3.3 SMFS homologues at acidic pH

SMFS experiments were also carried out on the human homologues of αSyn at acidic pH (Figure 5-9) in the same manner as those carried out at neutral pH in the previous chapter with proteins immobilised at their C-termini. As discussed previously, β- and γSyn are highly homologous proteins to αSyn, but, are not associated with PD and have a lower aggregation propensity and may not aggregate at all[274,275]. The data for β- and γSyn, as with WT is the sum of triplicate experiments. Each individual experiment for β- and γSyn consisted of at least 4

force maps of 500 approach-retract cycles. The overall number of approach-retract cycles for the experiment in triplicate was 9500 and 9500 for β- and γSyn, respectively.

The results show that there is a similar pH induced reduction in the L_C for both β- and γSyn. This shows that the pH induced structure is not limited to αSyn but also includes its human homologues and so is possibly due to the similarity in the gross properties of these proteins. It has been shown previously that a reduction in pH leads to a similar increase in structure and compaction in the synuclein homologues as observed for αSyn in studies on the monomeric proteins[275]. This is reflected in the dimeric interactions presented here.

The three proteins have similar gross properties to each other as they contain three domains as highlighted in Section 1.5.2 (Figure 1-14). The main differences between the proteins are an absence of a 12 residue stretch in the central NAC region in βSyn that is present in both α- and γSyn (as discussed in Chapter 1) and divergence at the acidic C-termini of the protein homologues. Importantly, the homologues dimerise at both physiological and acidic pH despite the absence of the central NAC region in βSyn. The fact that the dimerisation can still be observed and is similar suggests that this region doesn't play a role in the association of the proteins observed in these experiments. The divergence in the C-termini of the homologues may explain some of the variations in the L_C distributions between these proteins. A reduction in the L_C of γSyn at acidic pH can be observed, however, this homologue maintains a longer L_C in acidic conditions (27.9 nm) than that of either α- or βSyn (23.5 and 22.2 nm respectively). This may be due to the shorter C-terminus of γSyn compared to either α- and βSyn. It has been hypothesised in this chapter that the pH dependent reduction in the L_C of the αSyn dimer is due to the collapse of the acidic C-terminus. Given that γSyn has a shorter C-terminus but, maintains a longer L_C than the other synucleins upon acidification supports this finding that the C-terminal forms some pH dependent structure.

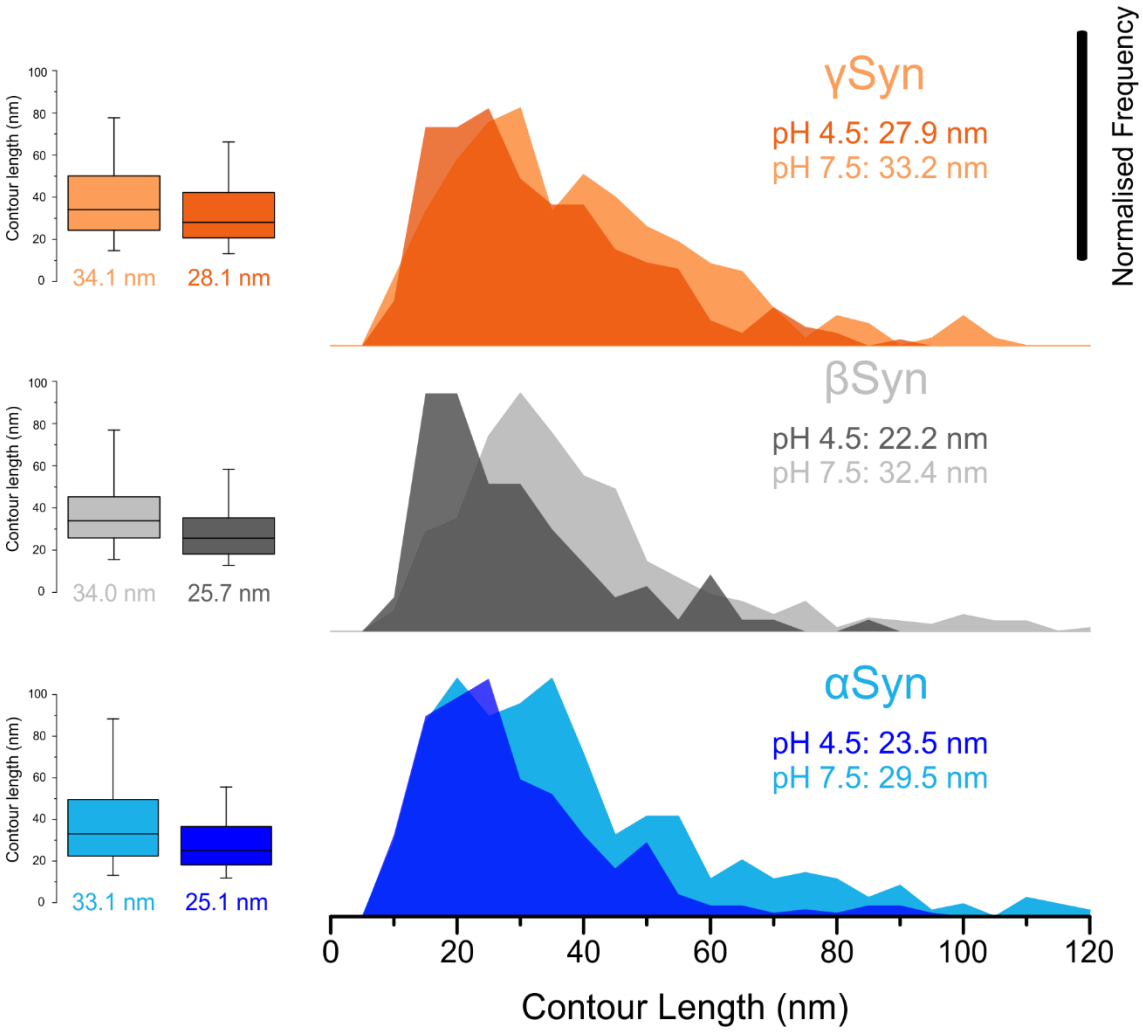


Figure 5-9. L_C distributions from SMFS data from C-terminal immobilised synuclein homologues at both physiological and neutral pH. Histograms of L_C distributions are shown as area plots (bin width: 5 nm). Blue, grey and orange histograms show α-, β- and γSyn L_C distributions in both neutral (light coloured histograms) and acidic (dark coloured histograms) conditions. Modal L_C values from Gaussian fits are denoted inset. The collapse of the L_C at acidic pH is clear in all three homologues. Box plots of L_C distributions are shown (left-hand side). Whiskers represent the 5th and 95th percentile. Median values are shown below the plots. Experiments were carried out at either 20mM Tris, 200 mM NaCl pH 7.5 or 20 mM acetate, 200 mM NaCl, pH 4.5.

5.4.3.4 SMFS E46K at acidic pH

The effect of pH on the dimerisation of E46K was also investigated in SMFS studies (Figure 5-10) in a similar manner to αSyn as discussed previously in this chapter. Each individual experiment for E46K consisted of at least 4 force maps of 500 approach-retract cycles. The data for E46K, as with WT is the sum of triplicate experiments. The overall number of approach-retract cycles for the experiment in triplicate was 8000.

The overall effect of decreasing the pH in SMFS studies of E46K is similar to that of WT αSyn (Figure 5-7). There is an obvious reduction in the L_C of the dimeric interaction. The distribution of L_C at pH 4.5 is very similar between WT and E46K with modal values of 23.5 and 20.9 nm respectively. This is despite the fact that there are obvious differences in the L_C distribution between WT and E46K at neutral pH. SMFS results suggest the dimerisation of E46K is different at acidic pH relative to physiological pH, this is despite the similar structural properties (Figure 5-5) and similar aggregation propensities (lag times: 13.2 and 16.3 hours in acidic and neutral conditions respectively) of the monomeric protein. The SAXS and EOM data (Figure 5-3 Figure 5-4) do however suggest a more collapsed conformation of E46K at lower pH in a similar manner to WT (SAXS data in neutral conditions is presented in Chapter 4). The F_R of αSyn dimerisation is overall similar (Figure 5-10 B and C). However, at pH 4.5 there is a higher proportion of events with higher F_R values (Figure 5-10 C) which may indicate an additional population of more avid dimers.

The collapse of the L_C upon acidification of E46K is consistent with WT and the synuclein homologues; this fact suggests that the interaction interface and the proposed structure in the dimer in these conditions are similar in E46K. In acidic conditions, the K46 residue will be protonated, the E46 will be ≈ 50% protonated (pK_A of the E sidechain is 4.4) and so, it is logical that the dimerisation of these two variants will be similar in these conditions, but, show differences at neutral pH when E46 is protonated but the mutated K46 is not. It has been shown throughout this thesis that charge contributions play a significant role in the structure, conformation, aggregation and single molecule association of αSyn, these data therefore supports this observation.

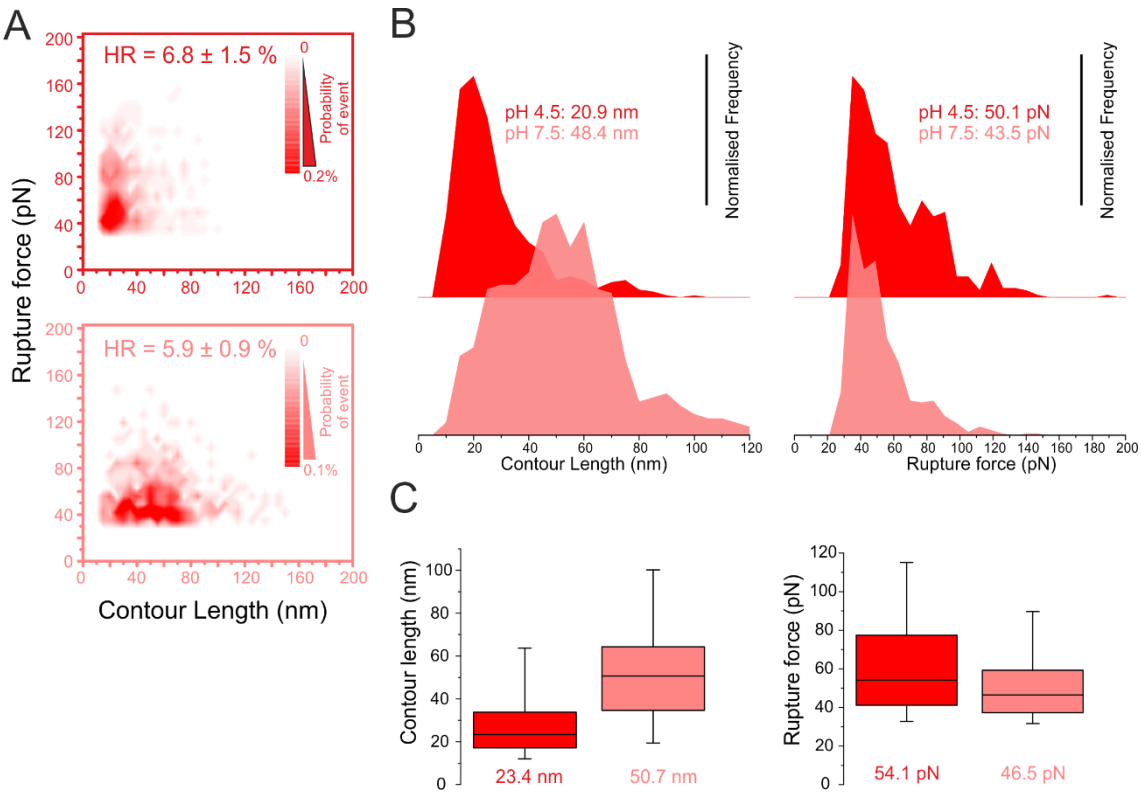


Figure 5-10. SMFS experiments of αSyn E46K A140C at both physiological and acidic pH. A) Scatter plots of SMFS from αSyn E46K at pH 4.5 (dark red) and pH 7.5 (light red). The points were binned by a contour length of 5 nm and a rupture force of 7 pN. The data was normalised to the total number of approach-retract cycles in the experiment (total pH 7.5 = 9,000 and pH 4.5 = 8000). Deeper colours denote higher frequency of events. A 'hotspot' is clearly present in the data from both conditions suggesting a specific single molecule interaction. The hit rate (HR) is shown inset for the combined triplicate experiments. B) L_C distribution histograms with the same colour coding as in A) (bin size 5 nm) displayed as area plots with modal values from Gaussian fittings denoted inset. The data shows that the modal L_C in acidic conditions (20.9 nm) is smaller than that at neutral conditions (48.4 nm). C) Box plots of L_C and F_R distributions Whiskers represent the 5th and 95th percentile. Median values are shown below the plots. Experiments were carried out at either 20mM Tris, 200mM NaCl pH 7.5 or 20mM Acetate, 200mM NaCl, pH 4.5.

5.5 Discussion

The results presented in this chapter have demonstrated several important findings about the dimerisation interaction of αSyn. Firstly, the novel use of L_C simulations has been used in parallel with experimental data to reveal the presence of force resistant structure in the dimeric protein at physiological pH. Different immobilisation regimes were used in order to gain more information about the dimeric interaction. When analysed alongside L_C simulations, the data reveal discrepancies in the proposed interaction region between different immobilisation regimes. This finding showed that the dimeric interaction of αSyn cannot be modelled by the simple interaction of two disordered monomers. There must be some

structure in the αSyn dimer and, moreover, this structure has to be force-resistant as there is an absence of additional unfolding events in the data.

Furthermore, SMFS experiments in acidic conditions showed that the dimerisation interaction is different to that in physiological conditions. These data show a pH dependent collapse of the L_C in acidic conditions. However, when the same experiments were carried out using different immobilisation points (A18C and A90C), the same collapse was absent. This suggested that the interaction region of αSyn at both physiological and acidic pH is the same but there is an increased level of collapse or structure in the dimer. Moreover, the data enabled some degree of localisation of the proposed structured region. The fact that the A140C interaction showed a pH dependent collapse whereas A18C and A90C did not allowed the proposal of a structured C-terminus in the dimeric species of αSyn in acidic conditions. The SMFS experiments indicated a modal L_C value of 23.5 nm for A90C at pH 7.5, the same as A140C at pH 4.5. This suggests then that at least this region of the protein: residues 90-140 collapse in acidic conditions and don't contribute to the L_C value.

The pH dependent formation of structure in the acidic C-terminus is supported by various studies of monomeric αSyn and biophysical data presented in this thesis. Intramolecular NMR PRE studies have shown that at acidic pH, the C-terminus is significantly more collapsed[220,221]. These studies also showed that this compaction extends into the NAC region. Here, we shown that pH induced structural formation doesn't just effect the intramolecular conformation of monomeric αSyn but also causes changes in the structure of the dimeric protein. The data presented here is the first evidence of this.

An interaction interface for the dimerisation of αSyn was proposed based on the SMFS for A18C. This immobilisation regime is distil to the proposed structure in the C-terminus of the protein. The experimentally observed L_C distributions were therefore used in parallel with simulated values to suggest an interaction region from at residues 40-70. Given that the L_C value doesn't change between neutral and acidic pH for this immobilisation regime, it was proposed that this is the interaction region in both conditions. This is a novel region of proposed dimeric interaction located mostly in the N-terminus and interestingly, mostly not in the highly hydrophobic NAC region of the protein (residues 65-90). It is also interesting to note that this region contains all but one of the six familial PD mutations. The data from SMFS experiments on E46K provide further evidence that this is the site of the interaction interface as at pH 7.5, the dimerisation interaction of E46K is vastly different to WT suggesting this mutation disrupts the interaction interface.

6 Postulating a protective dimeric interaction

6.1 Abstract

From the data presented in previous chapters, an interaction interface in the αSyn dimer was proposed to be present in the N-terminal region of the protein. The nature of the dimerisation interaction of αSyn is further explored in this chapter. SMFS and ThT aggregation assays were carried out in different ionic strengths and different salts. The results reveal that the dimeric interaction observed by SMFS is hydrophobically driven but the contributions of charge are critically important. Under conditions where an interaction is absent in SMFS studies, the rate of aggregation increases which offers more evidence that the interaction proposed here is protective to aggregation. Bioinformatics analysis of the αSyn sequence revealed aggregation-prone stretches in the N-terminus which likely form part of the interaction interface and corroborate SMFS experiments in which this region is proposed to interact in the dimeric species of αSyn. The same analysis on synuclein homologues was carried out in which the same aggregation-prone regions of the N-terminus are revealed, further suggesting that the dimeric interaction, proposed to be driven by these regions, is protective to aggregation and may be a physiological interaction.

6.2 Introduction

As discussed previously in this thesis, a variety of information can be extracted from the L_C in SMFS data. It was demonstrated in the previous chapter that using this information alongside simulated L_C data, the postulation of novel dimeric conformations are possible. In turn, the presence of force-resistant structure in the dimeric species increases the difficulty of predicting the exact location of the interaction interface as both the force-resistant dimeric structure and the location of the dimeric interface contribute towards the experimentally observed L_C data. By utilising different immobilisation regimes and analysing these in parallel with simulated L_C data, an interaction region was proposed, which is located in the amphipathic N-terminus of the protein.

Importantly, it was observed that dimerisation of synuclein homologues was also observed. As discussed previously the homologues are less aggregation-prone and not linked to PD[274,275]. There is literature precedent for this as all the synucleins have been shown to dimerise, however, this is not proposed to report on their respective aggregation propensity[276]. Given that the dimerisation interaction of the homologues and WT αSyn is similar in SMFS experiments, the interaction may not be one in which aggregation would be promoted, and

may therefore be proposed as a physiological interaction. The presence of physiological multimers of αSyn have been studied previously in cross-linking studies[46,255-259] in which they have been showed to be protective to cytotoxicity. These physiological multimeric species were proposed to be mediated by the N-terminal region of the protein, specifically the KTKEGV imperfect repeat motives. Altering these repeats with PD familial[258] or synthetic mutations[46] was shown to inhibit multimer formation and increase toxicity. Indeed, the localisation of all six known familial PD mutations in this region of the protein indicates its importance in the balance between health and disease.

The importance of charge on the dimerisation interaction and aggregation of αSyn has been demonstrated in this thesis, pH, ionic strength and charge mutations have all been shown to play an important role. In this chapter, the effect of different ionic strengths and also a different salt in the Hoffmeister series, on the dimerisation and aggregation of αSyn, was investigated.

6.3 Salt dependence of αSyn dimerisation

6.3.1 SMFS experiments reveal different populations of dimeric interaction

The effect of different ionic strength on the dimerisation of αSyn was investigated in SMFS experiments as described previously (Section 2.2.4). All SMFS experiments on full length αSyn in previous chapters were carried out in 200 mM NaCl. SMFS studies were also carried out at ten-times lower ionic strength at neutral pH (Figure 6-1). This was compared to the effects of a different salt, $(NH_3)_2SO_4$, which is lower in the Hoffmeister series of salts and acts to 'salt out' proteins by increasing the contribution of hydrophobic interactions[366]. The data show that at physiological pH, altering the ionic strength and salt does not cause dramatic differences in the dimerisation of αSyn. However, there are some subtle differences in the data: in the presence of low ionic strength $(NH_3)_2SO_4$, the distributions of both L_C and F_R become narrower than in other conditions. The width of the data were analysed by the full width at half maximum values (FWHM) of Gaussian fitted data. This may suggest that at other conditions, the data may contain contributions from different dimer populations. This goes some way to explain the large spread of the data observed at this pH throughout this thesis. In particular, the F_R data in the presence of low ionic strength $(NH_3)_2SO_4$ is much narrower than at other conditions with a FWHM of 15.9 pN compared to that of 43 pN for SMFS experiments carried out in high ionic strength NaCl. The modal Gaussian value is lower at 30.4 pN in low ionic strength

$(NH_3)_2SO_4$ than the other conditions. The data suggests that there are contributions of dimeric interactions that occur at higher F_R that are present in all salt conditions at pH 7.5, but, which are absent in low ionic strength $(NH_3)_2SO_4$. This suggests, therefore, that in these conditions αSyn dimerises by a more homogenous interaction, or that the interaction surface is the same, but, with a narrower range of dimeric conformations, this is shown in the more distinctive hotspots in the SMFS data (Figure 6-1 A far right panel).

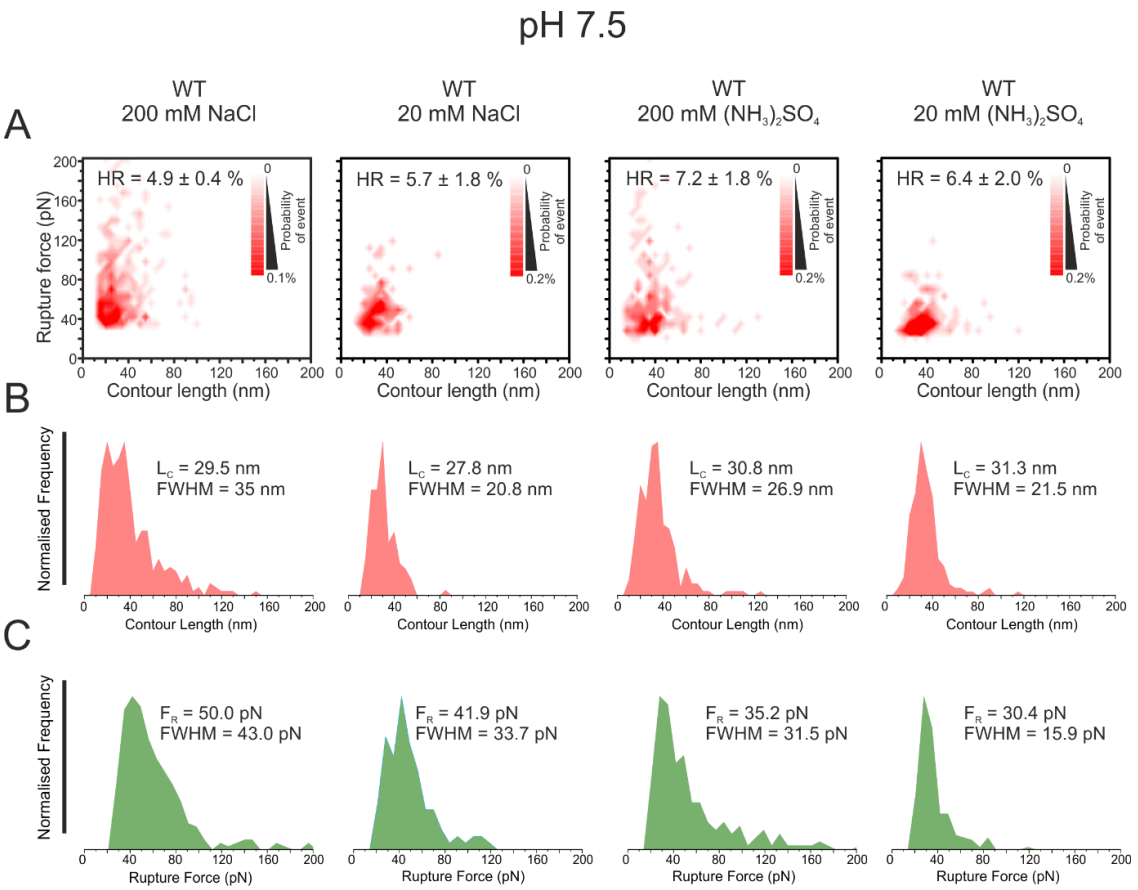


Figure 6-1. SMFS data of αSyn A140C under different salt conditions at neutral pH. A) Scatter plots of SMFS data carried out in different conditions denoted at the top of each panel. Concentrations denoted at the top of each panel are ionic strengths of the salt in the solution (experiments were carried out in 200 or 20 mM molar concentration of NaCl and 67 or 6.7 mM molar concentration of $(NH_3)_2SO_4$). The points were binned by a contour length of 5 nm and a rupture force of 7 pN. The data were normalised to the total number of approach-retract cycles in the experiment. Deeper colours denote higher frequency of events. The hit rate (HR) is denoted inset. B) Histograms of L_C distributions for each condition. The modal L_C value from Gaussian fits are shown inset. The width of the data is shown via the FWHM value of the Gaussian fits and shown inset. C) F_R distributions displayed in the same manner as L_C data in B.

The same series of experiments were carried out in acidic conditions to investigate further the importance of charge on the dimerisation of αSyn (Figure 6-2). SMFS experiments carried out in the previous chapter showed that at acidic pH, a characteristic pH dependent collapse of the

L_C occurred. This is consistent with the data presented in Figure 6-2. Strikingly, at low ionic strength NaCl, the hit rate of dissociation drops and a hotspot in the scatter plot disappears suggesting that the dimerisation interaction is no longer formed in these conditions. However, when experiments carried out in the same ionic strength of $(NH_3)_2SO_4$, a dimerisation interaction returns. These data suggest that the interaction is hydrophobically driven in that it is absent in a low ionic strength of NaCl, but present in the same ionic strength of $(NH_3)_2SO_4$, a salt lower in the Hoffmeister series that strengthens the hydrophobic effect of proteins.

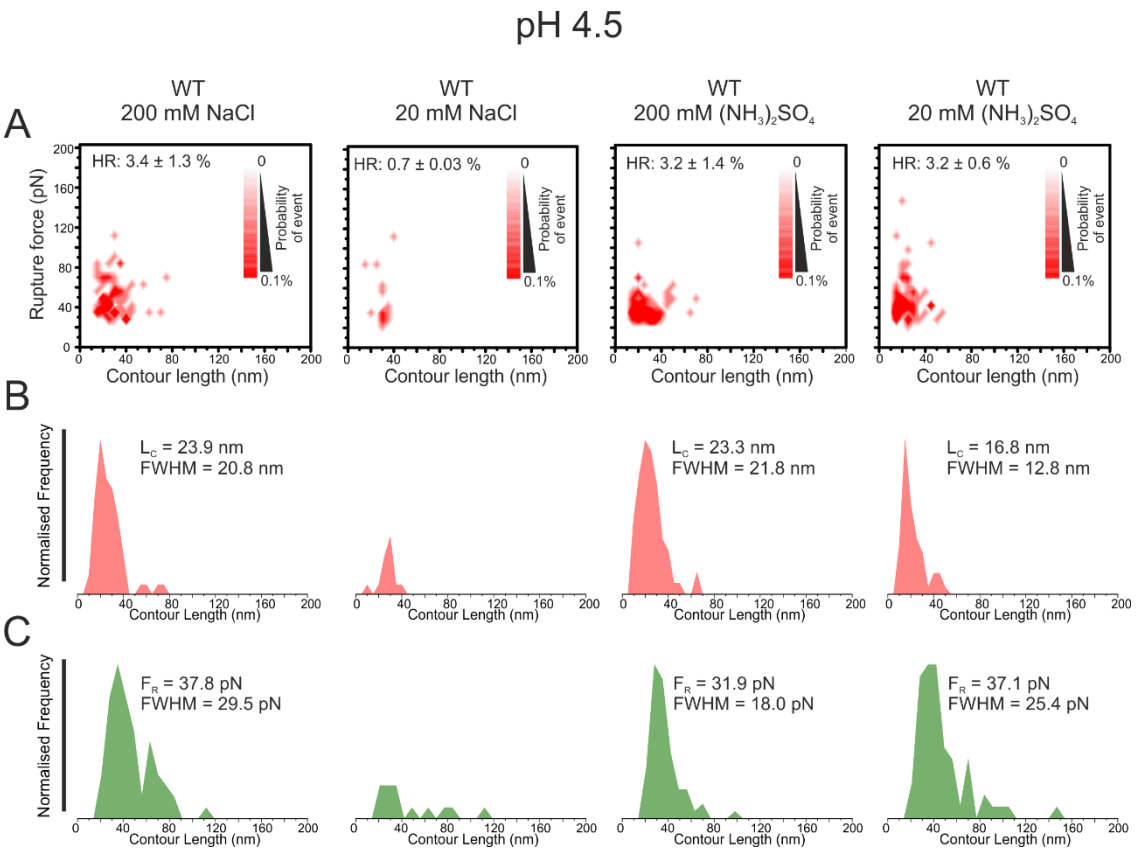


Figure 6-2. SMFS data of αSyn A140C under different salt conditions at acidic pH. A) Scatter plots of SMFS data carried out in different conditions denoted at the top of each panel. Concentrations denoted at the top of each panel are ionic strengths of the salt in the solution (experiments were carried out in 200 or 20 mM molar concentration of NaCl and 67 or 6.7 mM molar concentration of $(NH_3)_2SO_4$). The points were binned by a contour length of 5 nm and a rupture force of 7 pN. The data was normalised to the total number of approach-retract cycles in the experiment. Deeper colours denote higher frequency of events. The hit rate (HR) is denoted inset. B) L_C distributions for each condition. The modal L_C value from Gaussian fits are shown inset. The width of each distribution is shown via the FWHM value of the Gaussian fits and shown inset. C) F_R distributions displayed in the same manner as L_C data in B. L_C and F_R distributions in B and C for αSyn in low ionic strength NaCl (second column) are normalised to the distributions for those in high ionic strength NaCl (first column) in order to graphically show the absence of events.

At pH 4.5, αSyn is almost neutral (pI 4.7). It may be hypothesised that in these conditions, salt mediates a hydrophobically driven dimerisation interaction by shielding charges such as positive K residues. This would indeed have to be the case, if as proposed, the N-terminal region of αSyn was the location of a hydrophobically driven dimerisation interaction. This explains why an interaction would be observed in high NaCl but not in low NaCl but also why the interaction can also be observed at low ionic strength $(NH_3)_2SO_4$ where the hydrophobic effect is more pronounced.

The absence of interaction was validated via carrying out SMFS experiments on the same functionalised AFM tip and surface in the same experimental setup and simply changing the buffer conditions. This experiment is shown in Figure 6-3 and demonstrates that there is indeed an absence of dimerisation at low ionic strength NaCl but when the same functionalised tip and surface is incubated in higher ionic strength conditions, a distinctive hotspot appears in the data denoting the presence of a dimeric interaction.

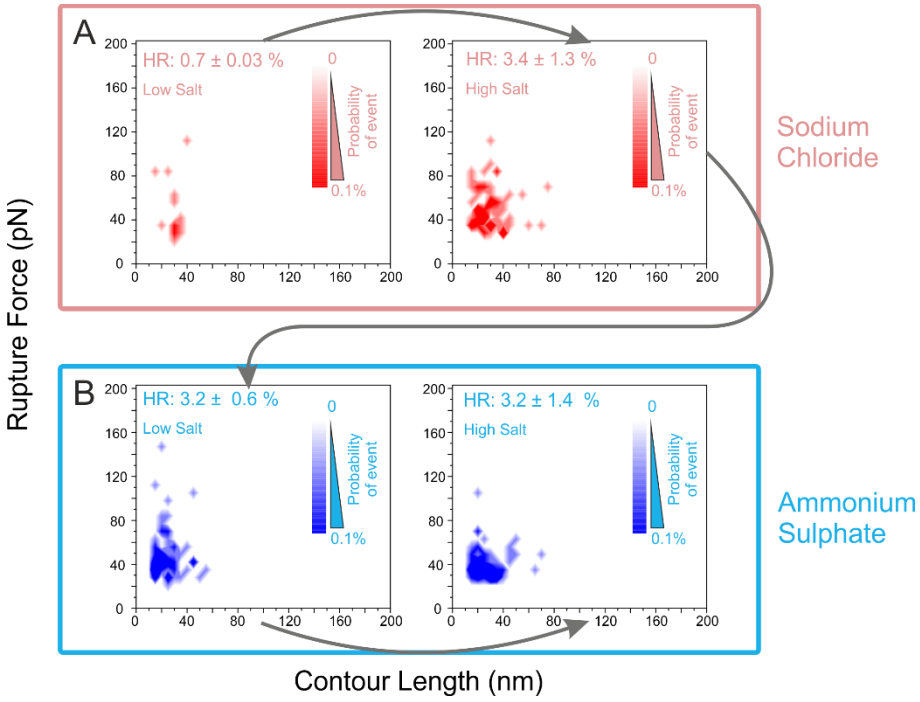


Figure 6-3. Scatter plots of SMFS experiments for αSyn A140C at different ionic strengths of both NaCl and $(NH)_3SO_4$ in acidic conditions. A) Scatter plots for SMFS experiments carried out in low ionic strength (20 mM) NaCl (first panel) and those carried out in high ionic strength (200 mM) NaCl (second panel). B) Scatter plots carried out in $(NH_3)_2SO_4$ presented in the same manner as in A with the same ionic strength conditions used as in A. All scatter plots presented here were carried out in the same experimental set up with the same functionalised cantilever tip and silicon nitride surface in order to confirm the absence of events in low ionic strength NaCl. Arrows show the sequence in which the experiment was carried out.

6.3.2 Aggregation of αSyn reveal promotion of a protective interaction

Aggregation assays were performed on αSyn in the various ionic strength and pH conditions used in the SMFS experiments as described in the previous section. Firstly, the aggregation kinetics at high and low ionic strength NaCl, conditions in which dimerisation events were absent and then reappeared respectively in SMFS experiments, were analysed (Figure 6-4). Interestingly, in conditions in which a dimerisation interaction is not observed in the SMFS experiments, the rate of fibril formation was more than twice as fast when compared to the same conditions with high ionic strength NaCl (2.2 ± 1.7 compared to 7.3 ± 0.5 hours). Both the aggregate yields and morphology were almost indistinguishable between the two conditions. Taken together with the SMFS data, this suggests that the interaction that is observed in SMFS studies is a protective to aggregation and offers more evidence that the interaction may be physiological. When this interaction can not be observed in SMFS studies, the aggregation rate speeds up suggesting that there is less competition from the proposed protective interaction and αSyn can go on to form interactions that are on pathway to aggregation.

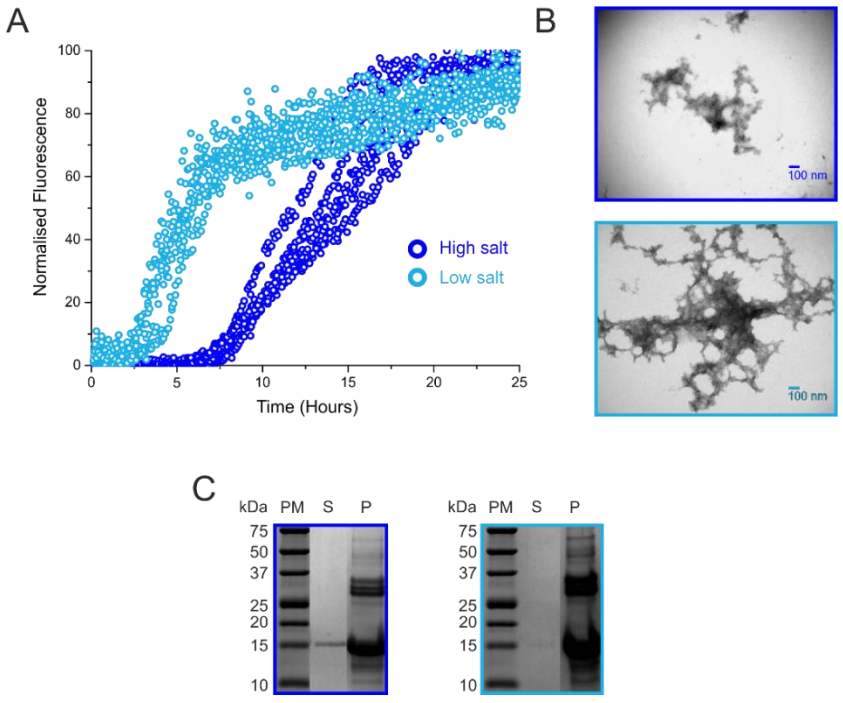


Figure 6-4. ThT aggregation assay of αSyn at different ionic strength NaCl in acidic conditions. A) ThT fibril growth assays of 100 µM αSyn in 20 mM acetate buffer pH 4.5 with either low or high ionic strength NaCl (20 or 200 mM in light and dark blue respectively). αSyn incubated in low ionic strength conditions has a faster lag time (2.2 ± 0.7 hours) relative to high ionic strength conditions (7.3 ± 0.5). B) Negative stain TEM images of aggregated structures at the end-point of ThT incubations colour coding is consistent throughout. Amorphous aggregates are observed in both conditions. C) Fibril yields as assessed by SDS-PAGE. The protein marker (PM) soluble (S) and pelleted (P) samples are shown. Fibril yields for both high and low salt conditions are identical at 76%.

The lag times determined from ThT aggregation assays for a variety of conditions are shown in Figure 6-5 and Table 6-1. Incubations at both acidic and physiological pH, in high and low ionic strength of both NaCl (Figure 6-5 A) and $(NH_3)_2SO_4$ (Figure 6-5 B) were carried out on both WT αSyn and αSyn E46K. The lag time analyses reveal differences between the speeds of fibril formation in different conditions of the different αSyn variants. As discussed in previous chapters, reducing the pH has the general effect of increasing the aggregation rate (incubations carried out at neutral and acidic pH are represented by green and red data points respectively). At pH 7.5 in NaCl, the aggregation of WT αSyn is generally unaffected by a change in ionic strength. E46K in the same conditions, however, is significantly affected by changing ionic strength. The lag time for E46K is more than 3 times faster in high salt compared to low salt. This effect may occur in E46K as it is a charge mutation in an important N-terminal region of the protein which has been proposed in this thesis to be protective from aggregation. It can be proposed that an increase in salt concentration acts to generically increase aggregation by inducing a Hoffmeister salting out effect at high salt concentrations[366], which, with a disrupted region than can no longer form a protective interaction, leads to an increased aggregation rate.

As discussed previously, the rate of aggregation of WT αSyn is increased with decreasing ionic strength, under these conditions an interaction is also absent in SMFS studies. This indicates that the proposed protective, hydrophobic interaction requires a high ionic strength in order to shield charges and facilitate the formation of a protective dimer. The same trend occurs for E46K but at a much lower extent as there is a lower average lag time for this variant in low ionic strength NaCl, but, this is not significantly different from high ionic strength conditions. This may be due to the fact that at pH 4.5, the E46K mutant retains an extra positive charge in respect to WT at the same pH, the association of the protein may thus be less susceptible to a change in ionic strength. The aggregation rate for E46K is also slightly slower than that of WT under acidic conditions. It could be postulated, therefore, that the protective hydrophobic interaction in which charges have been proposed to play a key role, still forms in E46K slowing down the aggregation of the protein. Indeed at pH 4.5 in high salt, both E46K and WT show a similar dimerisation interaction (see previous chapter).

The lag time analyses for synuclein variants in the presence of $(NH_3)_2SO_4$ under neutral conditions retains a similar trend with similar lag times to those in the presence of NaCl (only WT at high ionic strength is significantly different between the salt types). This indicates that

the enhanced hydrophobic effect in $(NH_3)_2SO_4$ does not play a significant role in the rates of aggregation. This may be caused by the protein possessing a highly negative charge at this pH.

Conversely, in acidic conditions in the presence of $(NH_3)_2SO_4$, the trends in the lag times are reversed in respect to the lag times recorded in the presence of NaCl. This indicates that enhancing the hydrophobic effect in these conditions plays a key role in the aggregation of αSyn variants. This may be because at pH 4.5, the protein is almost neutrally charged (pI 4.7) and so hydrophobic interactions are more important. The aggregation of the WT protein in these conditions at low ionic strength is significantly slower than in high ionic strength conditions. This could be because the protective hydrophobic interaction is promoted by $(NH_3)_2SO_4$ in which the hydrophobic effect is strengthened. The interaction was proposed to be modulated by charged residues in the N-terminal region, these however, are minimised at pH 4.5. The fact that this interaction in low ionic strength $(NH_3)_2SO_4$, slows down aggregation, and causes the reappearance of a dimerisation event in SMFS experiments supports the hypothesis that a hydrophobically driven protective interaction is being observed in SMFS studies. The same trend can be observed for E46K in the same conditions, but, again to a lesser extent.

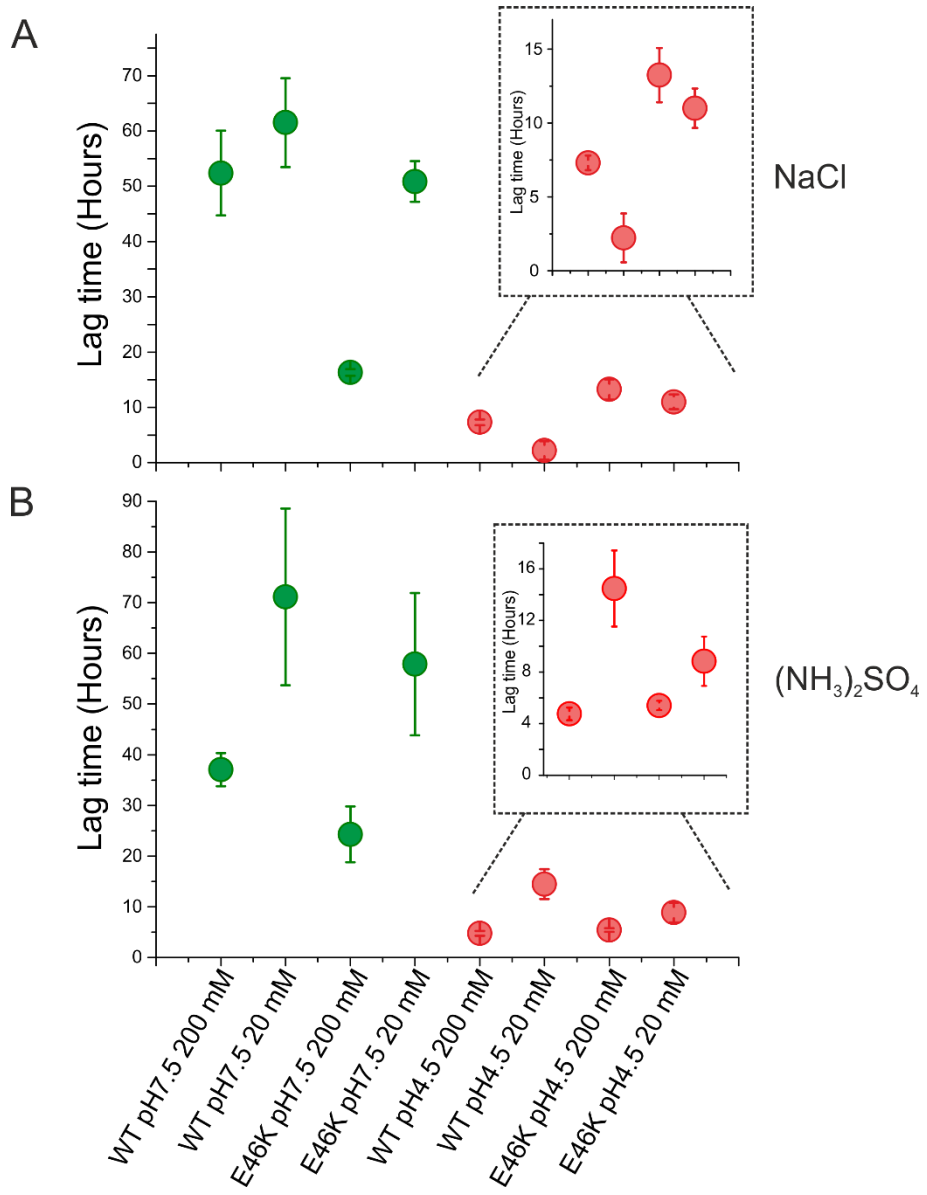


Figure 6-5. Lag time analysis from ThT aggregation assays of WT and E46K αSyn at 20 and 200 mM ionic strength, NaCl or $(NH_3)_2SO_4$ and pH 4.5 or 7.5. A) Incubations in the presence of NaCl, neutral (pH 7.5) and acidic (pH 4.5) conditions are shown in green and red respectively. Expansions of the lag times in acidic conditions are shown inset. Each incubation is as titled at the bottom of the figure labelled on the x-axis. Ionic strengths of NaCl are denoted. The centres of filled circles represent the average of at least 4 ThT incubations. Error bars represent SD. B) presented the same as in A, but showing lag times for incubations in the presence of $(NH_3)_2SO_4$. Each incubation is as titled at the bottom of the figure labelled on the x-axis. Ionic strengths of $(NH_3)_2SO_4$ are denoted.

pH	Protein variant	Ionic strength (mM)	Lag time (hours)	pH	Protein variant	Ionic strength (mM)	Lag time ± SD (hours)
		+ NaCl				*+ $(NH_3)_2SO_4$*	
7.5	WT	200	52.4 ± 7.7	7.5	WT	200	37.1 ± 3.3
		20	61.5 ± 8.0			20	71.1 ± 17.4
	E46K	200	16.3 ± 0.6		E46K	200	24.3 ± 5.5
		20	50.9 ± 3.7			20	57.9 ± 14.0
4.5	WT	200	7.3 ± 0.5	4.5	WT	200	4.8 ± 0.5
		20	2.2 ± 0.7			20	14.5 ± 3.0
	E46K	200	13.2 ± 1.8		E46K	200	5.4 ± 0.4
		20	11.0 ± 1.3			20	8.8 ± 1.9

Table 6-1. Lag times determined from ThT assays in various conditions for both WT and E46K.

6.4 Proposing a novel interaction region

The bioinformatic tool Zyggregator was used to analyse the aggregation propensity of αSyn in different conditions. Zyggregator is an algorithm derived by Vendruscolo and colleagues determining the intrinsic amyloid aggregation propensities of proteins[337-339]. The algorithm takes into account the intrinsic physio-chemical properties of a protein or peptide sequence and their contributions to its amyloid propensity. It takes into account the hydrophobicity, charge and secondary structural propensity of a polypeptide. It also takes into account the existence of patterned regions of residues with alternating hydrophobic and hydrophilic residues that have been shown to influence strongly the aggregation process[367]. Regions of the polypeptide sequence that have a high aggregation propensity also influence the likeliness of amyloid formation[338], therefore the algorithm takes into account intrinsic aggregation propensity in a seven residue sequence of the protein centred on a specific residue and this is carried out throughout the protein sequence. Residues that are considered "gatekeepers" are also taken into account. This term considers that if a hydrophobic sequence is flanked by charged residues its contribution to the aggregation propensity is much reduced by electrostatic repulsions[339]. Local structural properties are also taken into account in the algorithm as hydrophobic regions that are sufficiently unstructured increase the aggregation propensity[368].

The extrinsic effects of pH on the amyloid propensity are taken into account in the software and have been utilised in this section. The Zyggregator scores for both WT and E46K in neutral and acidic conditions are shown in Figure 6-6. A score of 0 in the sequence profiles indicates that the aggregation propensity is equal to that of a random sequence; a score of 1 indicates

regions which are one SD more aggregation-prone. There are three main aggregation-prone peaks in the αSyn and the PD mutant E46K sequence highlighted by shaded regions in Figure 6-6. The highly hydrophobic NAC region is highlighted, as are two additional regions in the amphipathic N-terminal region labelled 1 and 2 in Figure 6-6. These peaks are centred on residues 39 and 51. This data offers further evidence that αSyn may be likely to form intermolecular interactions via the N-terminal region, so far not greatly implicated in the aggregation of the protein.

This novel interaction has previously been proposed in this thesis. The data reveals that the aggregation-prone peaks in the N-terminal region are among the most pH dependent regions in the protein sequence. Neutral and acidic Zyggregator profiles (light and dark coloured respectively) are shown for both WT and E46K (Figure 6-6). At pH 4, the second peak in the profiles drop significantly. Taken with the data from SMFS studies in which the dimerisation interaction disappears in acidic conditions at low ionic strength (Figure 6-2 and Figure 6-3), this suggests that this region is crucial in the formation of a dimeric interaction observed in SMFS experiments. Data presented in previous chapters and in Figure 6-5 here, demonstrate that the rates of aggregation of αSyn at acidic pH are faster than those in neutral conditions. This may suggest that the balance of hydrophobicity in this region compared to the highly aggregation-prone NAC region plays an important role in aggregation. In neutral conditions the Zyggregator scores of this region is higher and so more likely to form an intermolecular interface that is protective to the aggregation of the protein. It has to be noted however that the conformation of the monomeric protein and the dimeric species changes with pH and so may also play an important role in the aggregation propensity of αSyn.

The pH dependency of region two in the Zyggregator plots also helps to rationalise SMFS carried out at different ionic strengths. At acidic pH, the fact that the dimerisation interaction requires high ionic strength NaCl conditions in order to form and be observed indicates the importance of charge in modulating this interaction (Figure 6-2 and Figure 6-3). This dependence on ionic strength is not present in SMFS experiments at neutral pH (Figure 6-1), as region 2 has a higher hydrophobicity in these conditions, and so is able to more readily form an interaction. The N-terminal region including peaks 1 and 2 denoted in Figure 6-6 is highly charged. The aggregation-prone peaks include and are surrounded by K residues that remain positively charged at acidic pH. High ionic strength NaCl would therefore aid in shielding these charges and mediate the formation of a hydrophobic interface. A lack of ionic strength dependence on the observation of the dimeric interaction also occurs in SMFS experiments

carried out in $(NH_3)_2SO_4$ (Figure 6-2 and Figure 6-3). In these conditions the hydrophobic effect is strengthened and so N-terminal interactions are more likely to form even without charge shielding interactions taking place.

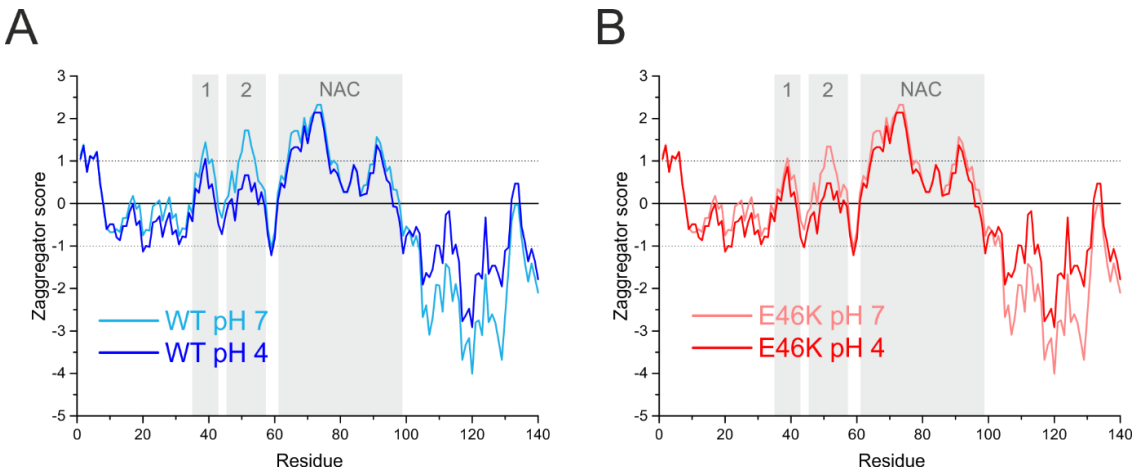


Figure 6-6. Zyggregator plots of the WT (A) and E46K (B) αSyn sequences in neutral and acidic conditions. A) Zyggregator profiles for WT αSyn at pH 7 (light blue) and pH 4.5 (dark blue). B) Zyggregator profiles for E46K αSyn at pH 7 (light orange) and pH 4.5 (dark orange). Regions with high amyloid propensity are shaded with grey boxes and labelled 1, 2 and NAC. A score of 0 in the sequence profiles indicates that the aggregation propensity is equal to that of a random sequence; a score of 1 indicates regions which are one SD more aggregation-prone.

The Zyggregator scores of WT and E46K αSyn are compared in Figure 6-7. The main difference in the aggregation propensity in the two proteins is in the 2 N-terminal peak regions as highlighted previously. This is expected as the mutation is close to both these regions at residue 46 and the algorithm takes into account local regions of the protein such as the hydrophobicity of 7 residue segments as well as the contribution of charged gatekeeper residues are taken into account over longer regions of 21 residues. E46K causes a drop in aggregation propensity in peaks 1 and 2 at both acidic and neutral conditions. The effects that the mutation has on the profiles are subtle, but may contribute to the significant difference observed in the dimerisation of the proteins and also in their aggregation. The subtle drop of Zyaggregator score of these regions in E46K may lead to a reduction in the propensity of a proposed protective interaction in this region and may offer and explanation as to the different dimerisation interaction relative to WT observed in SMFS studies (Chapter 4).

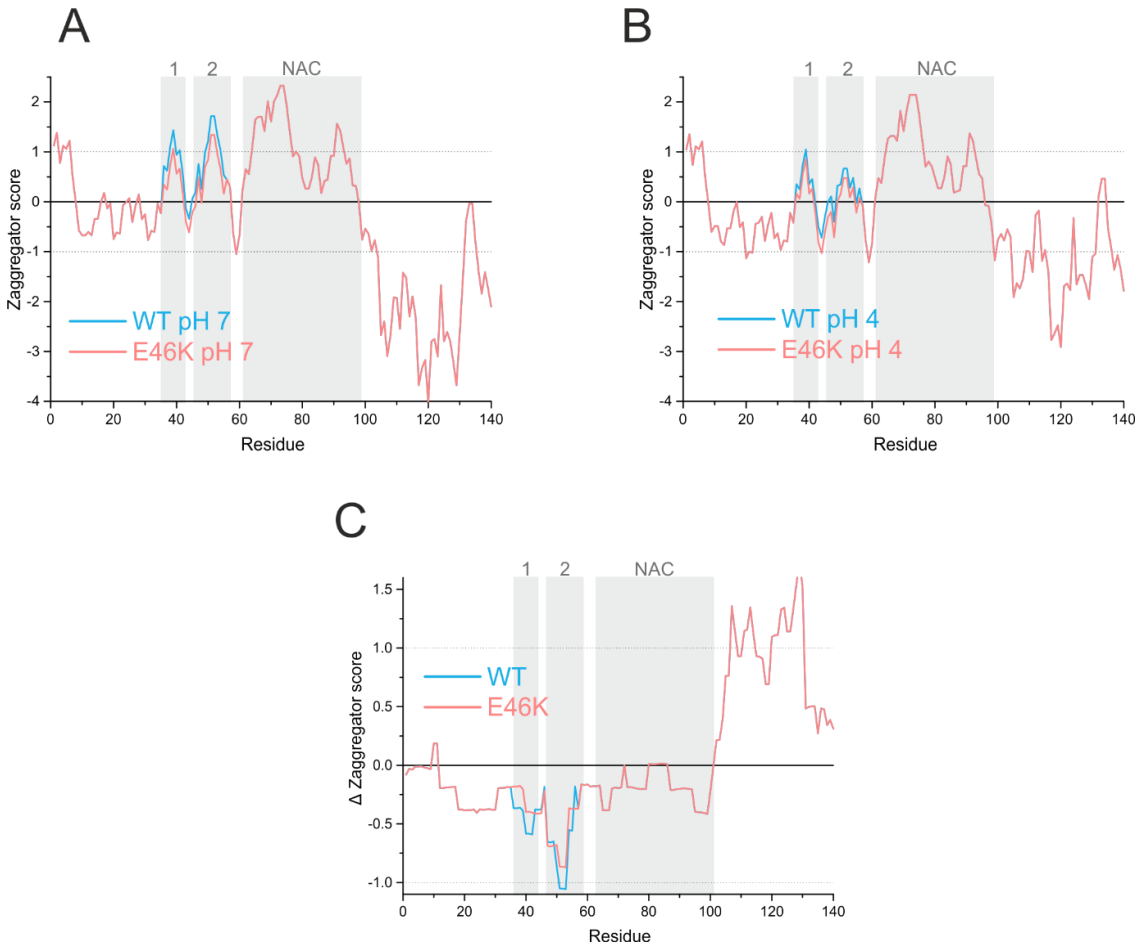


Figure 6-7. Zyggregator plots of the WT and E46K αSyn sequences at pH 7 and 4. A) Comparison of Zyggregator profiles of WT (blue) and E46K (red) at pH 7. The plots are presented in the same way as in Figure 6-6 with shaded regions indicating the main sections with high aggregation propensity. B) Difference of Zyggregator profiles of WT (blue) and E46K (red) at pH 4. C) The ΔZyggregator scores for both WT (blue) and E46K (red) calculated by subtracting the Zyggregator scores at pH 7 from those at pH 4. In this way, a negative value shows a decrease in amyloid propensity at pH 4 relative to pH 7. The ΔZyggregator profiles indicate that E46K is less pH dependent than WT.

The same bioinformatics approach was carried out on the synuclein homologues β- and γSyn (Figure 6-8). Very similar profiles were observed for the non-PD associated homologues as were previously observed for αSyn. The main features in the profiles are retained such as the presence of 2 N-terminal regions of high aggregation propensity as well as hydrophobic NAC regions in the centre of the proteins. The highlighted regions in Figure 6-8 are identical to those regions in αSyn (Figure 6-6 and Figure 6-7) for reference. Both the synuclein homologues (β- and γSyn) contain hydrophobic NAC regions which are generally less hydrophobic and shorter in length than αSyn, illustrating their lower aggregation propensity. The pH dependent reduction in the aggregation propensity of the N-terminal regions is also consistent in the homologues. The presence of N-terminal peaks in the homologues may suggest that the

interaction of these proteins are also mediated by the N-terminal region. Taken together with the fact that the dimerisation interaction observed by SMFS studies is similar between α-, β- and γSyn provides evidence that the interaction in αSyn is protective to aggregation.

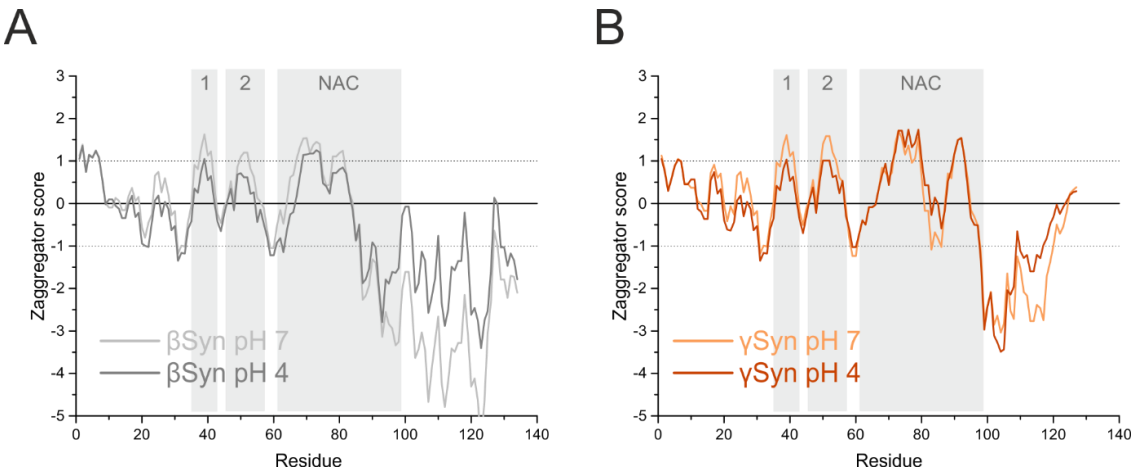


Figure 6-8. Zyggregator plots of synuclein homologues: β- (A) and γSyn (B). A) Zyggregator profiles for βSyn at both pH 7 (light grey) and pH 4 (dark grey). B) Zyggregator profiles for γSyn at both pH 7 (light grey) and pH 4 (dark grey). The same shaded areas showing high aggregation propensity of αSyn as shown in Figure 6-6 and Figure 6-7 are shown in the plots. These highlight the same regions of high aggregation propensity present in the synuclein homologues.

The intrinsic hydrophobicity of αSyn and the dependence on pH was investigated by carrying out a different bioinformatics approach (Figure 6-9). The bioinformatics tool CamSol, an algorithm that has also been derived by Vendruscolo and colleagues[369]. The CamSol method yields a solubility profile (one score per residue in the protein sequence) where regions with scores larger than 1 denote highly soluble regions, while scores smaller than -1 denote poorly soluble ones. CamSol profiles indicate that same peaks that were shown to have a high Zyggregator score are also have low hydrophobicity (negative peak in CamSol profile, Figure 6-9). The same peaks are present in the synuclein homologues and E46K. CamSol profiles do not have a high dependence on pH unlike the Zyggregator scores shown previously (Figure 6-6 and Figure 6-8).

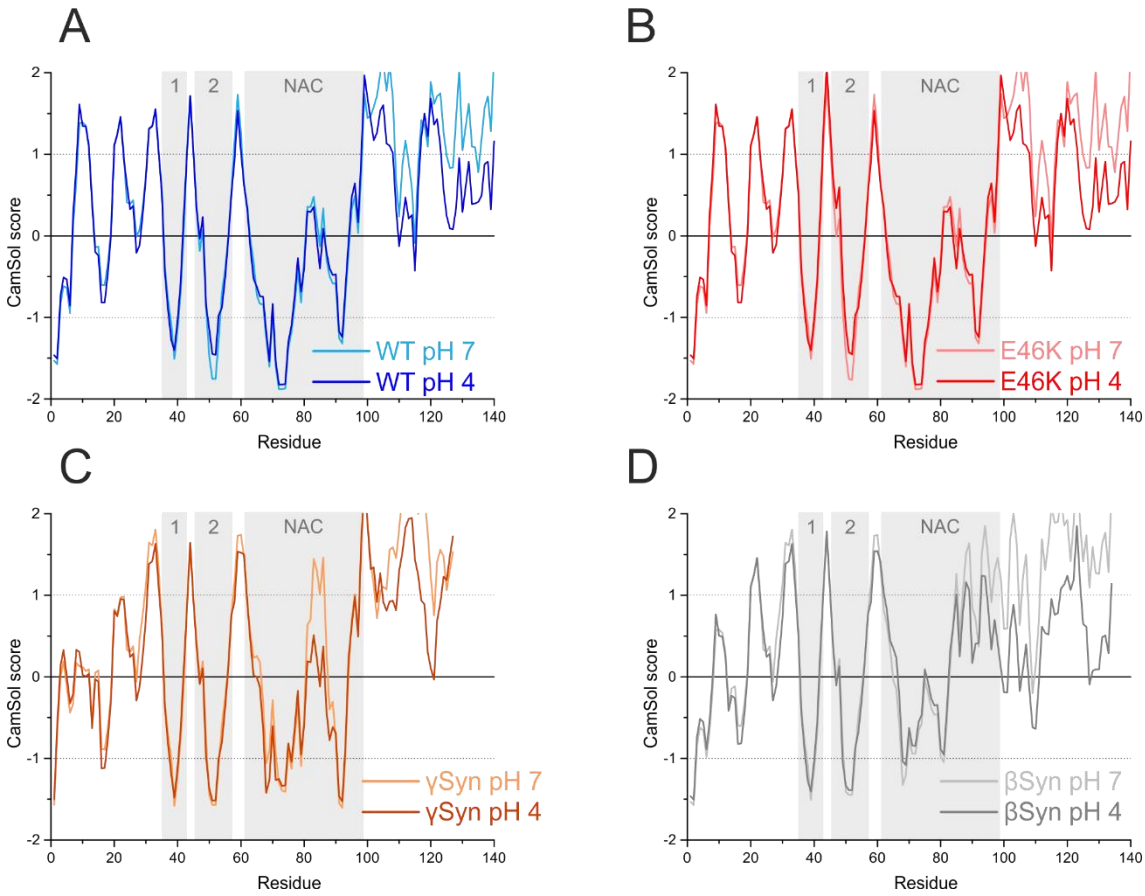


Figure 6-9. CamSol profiles from synuclein variants at pH 7.5 and 4.5. WT αSyn, E46K αSyn, γSyn and and βSyn CamSol profiles are shown in A-D, respectively. Profiles for pH 7 and 4 are shown for all the variants in light and dark colours, respectively. Regions with scores larger than 1 denote highly soluble regions, while scores smaller than -1 poorly soluble ones. Regions of high hydrophobicity are highlighted in grey and are the same regions as for the Zyggregator profiles (Figure 6-6, Figure 6-7 and Figure 6-8).

6.5 Discussion

In this chapter, the nature of the dimerisation interface has been further explored. It has been shown that an interaction is dependent on the fine balance of hydrophobic and charged interactions. In previous chapters, SMFS studies of full length αSyn extended from different immobilisation points suggested that the interaction interface occurred in the N-terminal region of the protein and not in the central highly hydrophobic central NAC region. This was further supported by bioinformatics data presented in this chapter in which two predicted regions in the N-terminal region, close in proximity relative to each other, were identified as highly aggregation-prone. The data suggests that these aggregation-prone sections may indeed be able to interact, and taken together with observed and simulated L_C data from SMFS studies suggest that these regions are most likely to form the intermolecular interaction interface in WT αSyn.

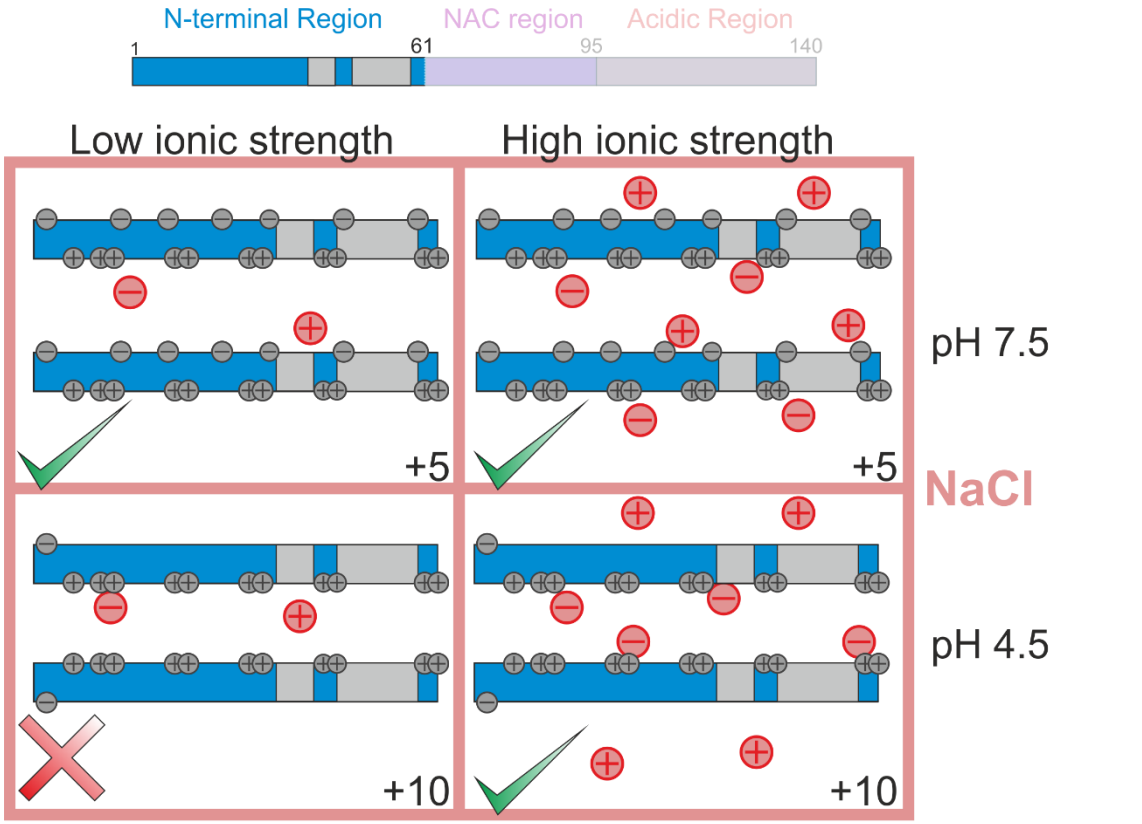
The contributions of charge and of different Hoffmeister salts were also explored in SMFS and aggregation experiments. The results shed light on the main properties that are driving the dimeric interaction that has been observed in SMFS studies. Under acidic conditions at low ionic strength NaCl, a dimeric interaction disappears relative to high ionic strength conditions (Figure 6-2 and Figure 6-3). This suggests that the interaction at this pH requires a high ionic strength in order to act as a charge screen. The Zyggregator profiles revealed that the interactions are most likely in regions of high amyloid-propensity in the N-terminal region. High ionic strength NaCl would therefore be required for this interaction to shield local charges in order to promote a hydrophobically driven dimerisation interface. Interestingly, aggregation assays reveal that in low ionic strength conditions where a dimerisation interaction is absent in SMFS studies, the rate of aggregation actually increases. This provides further support of the proposed hypothesis that the dimerisation interaction is in fact protective to aggregation. This hypothesis is corroborated by SMFS experiments carried out on the lower aggregation-prone, non-PD linked synuclein homologues presented in previous chapters. The Zyggregator data here show very similar profiles for the homologues in respect to αSyn. Importantly, Zyggregator peaks are also present in the N-terminal region of the synuclein homologues suggesting that their dimerisation is driven by the same regions of the homologues as for αSyn. The same peaks are confirmed in CamSol predictions revealing the high hydrophobicity of these regions (Figure 6-9).

Experiments carried out in $(NH_3)_2SO_4$ also suggest that the interaction is hydrophobically driven. At acidic pH with low ionic strength $(NH_3)_2SO_4$, a dimerisation interaction is once again observed. $(NH_3)_2SO_4$ is lower than NaCl on the Hoffmeister series and so strengthens the hydrophobic effect in proteins[366]. The reappearance of a dimerisation event highlights the fine balance between hydrophobicity and charges driving an intermolecular interaction of αSyn. The balance of charge and hydrophobicity and the contributions of each on dimerisation is summarised in Figure 6-10.

A subtle fall in aggregation propensity of the N-terminal region in E46K may provide some explanation as to why the dimerisation interactions revealed by SMFS for this protein are significantly different (Chapter 4). It reveals the fine balance between intermolecular interactions in these regions, in that inhibiting the interaction by subtly reducing hydrophobicity in the N-terminal region, may alter the balance between a protective dimerisation interaction and an interaction that may promote aggregation in this disease-

linked mutant. Modelled L_C data for E46K in fact, located a wide interaction interface that extends into the NAC region.

As discussed throughout this thesis, previous studies have suggested that physiological multimeric species of αSyn may exist and are protective to aggregation[46,255-259]. These studies also showed that by mutating important repeat motives in the N-terminal region, the area of αSyn that has been shown to form part of the dimerisation interface this thesis, multimeric interactions are perturbed and in cell models these mutants cause toxicity. Importantly, these studies support the data presented in this thesis that protective, possibly physiological intermolecular interactions of αSyn take place and they are driven by interactions in the N-terminal region.


N-terminal Region
NAC region
Acidic Region
1
61
95
140
Low ionic strength
High ionic strength
pH 7.5
+5
+5
NaCl
pH 4.5
+10
+10

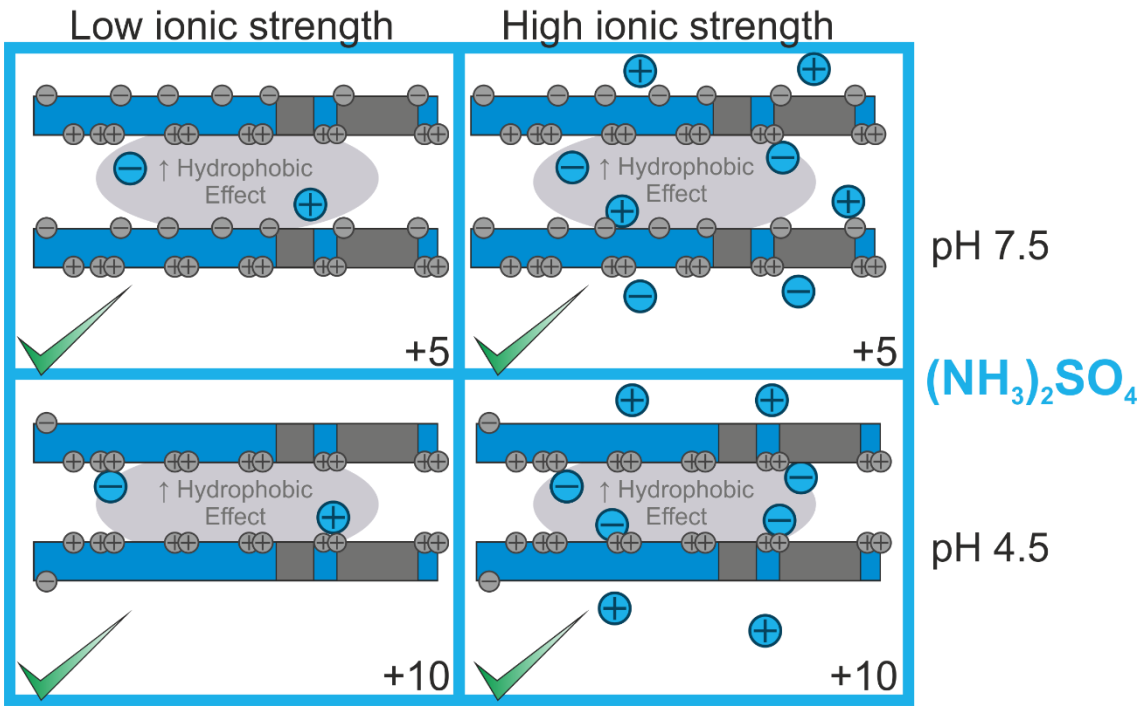

Low ionic strength
High ionic strength
↑ Hydrophobic Effect
↑ Hydrophobic Effect
pH 7.5
+5
+5
(NH3)2SO4
↑ Hydrophobic Effect
↑ Hydrophobic Effect
pH 4.5
+10
+10

Figure 6-10. Schematic representation of the dimerisation interface in the N-terminal region (residues 1-60) of αSyn. Schematic of full length αSyn (top) with the N-terminal region highlighted (blue). Peaks 1 and 2 from Zyggregator (Figure 6-6) analysis are shown in grey. The top panel (red box) shows the interaction in the presence of NaCl. Interactions at pH 7.5 and 4.5 are shown in the top and bottom boxes in this panel, respectively. Interactions at low (20 mM) and high (200 mM) ionic strength are shown in left and right boxes, respectively. Charged residues are represented by grey plus and negative signs. The negative values attributed to E are absent at pH 4.5. Net charge of the N-terminal region are shown at the bottom right in each box. The charge increases from +5 at pH 7.5 to +10 at pH 4.5 due to neutralisation of E residues. Ticks and crosses in each box represent the presence or absence of dimerisation interaction observed in SMFS experiments. An interaction is absent in low ionic strength (20 mM) NaCl, pH 4.5 as there is insufficient charge shielding for the highly positively charged protein region. The bottom panel (blue box) shows the interaction in the presence of $(NH_3)_2SO_4$. The Zyggregator peak regions are shown in darker grey to represent the increased hydrophobic effect on these regions. Interactions are observed in conditions in which an interaction was absent in NaCl due to strengthened hydrophobic interactions.

7 Conclusions and future directions

The conversion of normally soluble functioning proteins into intractable aggregates or amyloid has been the subject of intense research over recent decades as amyloid disorders represent some of the most debilitating diseases in the world at present. The aggregation of αSyn in the CNS causes several prevalent human diseases including PD, DLB and ALS amongst several others. Although the association of αSyn with PD was made 20 years ago[181], the aggregation process is still not fully understood and, like other amyloid diseases, there are as yet no disease modifying therapies. It is therefore important to carry out studies in order to gain more fundamental knowledge regarding the properties of αSyn self-association.

In this thesis, the intermolecular self-association of αSyn was interrogated on a single molecule scale primarily using SMFS. These experiments were carried out in parallel with various biophysical analyses in order to better understand the process of self-association. Utilising SMFS in this manner allowed the interrogation of the first association step in the complex association pathway. As it is not known where the toxic species lie on this pathway, it is a logical step to attempt to identify possible therapeutic targets.

7.1 A display system for aggregation-prone peptides

A SMFS technique was developed in this thesis which involved a novel display system in order to identify accurately, interactions of small aggregation-prone peptides on a single molecule scale. This would be impossible to achieve using a conventional AFM SMFS setup as interactions of the small peptides used (12 residues) would be obscured by noise that occur in force-distance plots proximal to the surface. Importantly, biophysical analysis confirmed that the engineered display constructs, mechanically strong pL with displaying peptides of interest, maintained the same structure which may otherwise affect SMFS experiments.

This novel technique was utilised to study dimerisation on the single molecule scale of $\alpha Syn_{71\text{-}82}$, a region of αSyn shown to be both necessary and sufficient for aggregation[204]. DFS showed that the dimerisation interaction had a lifetime in the range of seconds (a k_{off}^{0F} of 0.18 s^{-1} corresponding to a lifetime of 5.79 s). Comparable lifetimes, also on a seconds time scale, have been reported for the dimerisation of full length αSyn (4.00 s at pH 2.7 and 1.35 s at pH 3.7)[316]. The finding that the lifetimes of full length αSyn and $\alpha Syn_{71\text{-}82}$ dimers are of comparable magnitudes suggests that residues 71-82 play a key role in the stability of dimeric species formed from the intact protein. However, it was shown in subsequent chapters in this thesis that this region so not part of the dimerisation interface of full length αSyn. It is possible

therefore that the avid interaction of the central NAC region revealed in Chapter 3 is shielded by other regions of the protein in the context of full length αSyn.

Additionally, a novel heterodimeric interaction was identified between αSyn_{71-82} and the same region from the non-PD associated protein γSyn_{71-82} and validated by native MS techniques. The studies revealed that γSyn_{71-82} had an inhibitory role on the aggregation of αSyn_{71-82} suggesting that this sequence plays a key role in the inhibition of aggregation observed in the full length proteins[275]. Moreover, SMFS and MS experiments suggested a similar dimerisation affinity of the αSyn_{71-82}/ γSyn_{71-82} heterodimer to the αSyn_{71-82} homodimer. This allowed the proposal of a kinetic competition mechanism of inhibition similar to that recently reported for the inhibition of full length αSyn by βSyn.

7.2 SMFS reveals dimerisation interaction of αSyn

SMFS studies were also carried out on full length αSyn. The data revealed that dimerisation interactions could indeed be observed and analysed on the single molecule scale. SMFS hits were also observed under conditions in which dimerisation interactions have been previously reported to be absent at neutral pH[316,318,319]. This is most likely due to the difference in protein concentrations immobilised on AFM tips and surfaces (5 μM in this study compared to 19 nM in previous studies). The hit rate for αSyn dimerisation at pH 7.0 was reported to be 0.8 % by this group[318] and not further analysed. It is likely, therefore, that interactions of αSyn were missed and not analysed in these studies at neutral pH. We have therefore presented the first SMFS of αSyn dimerisation under physiological-like conditions in this thesis.

The dimerisation of an early onset familial PD mutant: E46K was also analysed throughout this thesis, the L_C distributions showed significantly differences to WT at neutral pH. Biophysical analysis however, revealed very similar structural and conformational properties of WT αSyn and E46K with some subtle differences. Modelled L_C data localised a wide interaction interface for E46K dimerisation that includes the NAC region (residues 60-130). The dimerisation interaction of E46K therefore, may report more clearly on an aggregation promoting pathway.

7.3 Proposed novel dimeric structure of αSyn and its environmental dependence

SMFS experiments were carried out in parallel with a L_C simulation method. By comparing the experimentally observed data with simulated values, important information on the interaction interface and the conformational structure in the dimer was revealed. Different immobilisation

regimes were then used to yield more information about the dimer interaction. When L_C data of αSyn under different immobilisation regimes were analysed in parallel with simulated values, there appeared to be a divergence of reported interaction regions. This offered evidence that that the dimerisation interaction cannot be modelled as two polypeptide chains lacking structure interacting and must therefore contain some structure. Importantly, in order for this structure to contribute to the L_C values, it must be force resistant and therefore, it can be hypothesised that this dimeric structure is mechanically strong β-sheet in nature.

Experiments carried out in acidic conditions revealed a pH dependence in the dimerisation of the protein. SMFS experiments showed that the L_C became more collapsed (in A140C immobilisation experiments). Different immobilisation regimes in parallel with simulated L_C values indicated that the location of the pH induced collapse was at the C-terminus. There is precedent from previous studies by other groups that αSyn does indeed undergo a pH induced collapse of the C-terminus as discussed previously[220,221]. These studies have analysed the monomeric species of the protein. It should be noted that the collapse proposed here is novel in that it has been proposed to occur within the dimeric species of αSyn. Acidic pH also cause the rates of aggregation to increase, the more compact structure of the dimeric species may therefore play an important role in increased aggregation propensity.

The conformations of the αSyn dimer is of great relevance in physiological and pathological settings. Different pH environments are present in vivo to which αSyn is exposed. Specifically, the acidic environment of the lysosome (pH 4.5) which plays an important role in the cellular trafficking of αSyn in the endocytic pathway in the constitutive process of proteostasis[331,332]. The importance of lysosomal impairment in PD is supported by the identification of genetic associations between PD and Gaucher's disease, the most common lysosomal storage disorder, which is caused by mutations in the gene encoding the lysosomal hydrolytic enzyme glucocerebrosidase[365]. Studies have shown that 5-10% of PD patients possess glucocerebrosidase mutations[365]. The self-association of αSyn at lysosomal pH is therefore highly relevant and important to study.

7.4 Proposed novel interaction interface

SMFS experiments on different immobilisation positions allowed the proposal of a novel dimeric interface in the amphipathic N-terminal region of the protein and not in the central, highly hydrophobic NAC region. It is also interesting to note that this region contains all but one of the six familial PD mutations. The data from SMFS experiments on E46K provide further evidence that this is the site of the interaction interface as at pH 7.5, the dimerisation

interaction of E46K is vastly different to WT which may suggest that this mutation disrupts the interaction interface.

Bioinformatic analysis also supported the hypothesis of an interaction in the N-terminal region. The Zyggregator algorithm highlighted two highly aggregation-prone areas, other than the NAC region in the sequence of αSyn and both were localised in close proximity in the N-terminal region. The presence of the same regions in the synuclein homologues, which in SMFS studies have very similar dimerisation profiles, further suggest the interactions observed are driven by these regions. CamSol predictions showed regions of high hydrophobicity correlated with the high Zyggregator scores.

7.5 Postulating a protective physiological interaction

Various data in this thesis indicate that the interactions observed in SMFS experiments are not ones which promote aggregation and may therefore be protective. Firstly, the dimeric interactions of non-aggregation-prone, non PD-linked synuclein homologues detailed by SMFS are very similar to those of αSyn. This idea is supported by previous studies in which all synucleins were shown to dimerise, but this interaction was not predictive towards aggregation[276]. Crosslinking studies have also showed that all three synucleins can exist as physiological multimers[46,257], with very similar distributions of multimeric species. These studies indicate that the dimeric interaction being interrogated in SMFS studies may be in fact a physiological one and not necessarily one that is on-pathway or promotes pathological aggregation. This is further supported by the fact that the dimeric interaction for the PD familial mutation E46K revealed by SMFS showed significantly different L_C distributions to that of WT αSyn. This mutant was also shown to disrupt protective multimeric interactions in cross linking studies and increased toxicity[258]. A recent *in vivo* study provided evidence to the physiological importance of regions in the N-terminal region[370]. This study showed that alleles of the major histocompatibility complex that are associated with PD recognise residues 31-46 of αSyn. This indicates that when this "protective" region of αSyn is targeted by variants of the major histocompatibility complex, the protein can aggregate and this leads to PD.

The presence of aggregation-prone stretches in the N-terminus of αSyn and its homologues, revealed by Zyggregator analysis as discussed above, led to the hypothesis that the interactions were driven by hydrophobic interactions in these regions. Interestingly, at acidic, low ionic strength conditions (20 mM NaCl), dimerisation events as viewed by SMFS are absent. Aggregation assays showed that in these conditions, the rate of aggregation of αSyn was actually increased which suggested that the dimeric interactions may be protective to

aggregation. Moreover, when the NaCl was substituted for a salt lower in the Hoffmeister series, $(NH_3)_2SO_4$, the dimerisation interaction was once again observed. $(NH_3)_2SO_4$ acts as a kosmotrope in that it increases the stability and structure of water-water interactions, it therefore, effectively strengthens the hydrophobic effect in the protein. It shows that the dimeric interaction is indeed hydrophobically driven. The aggregation of αSyn in these conditions is much slower which also supports the fact that this interaction is protective to aggregation. The fact that ionic strength also plays an important role in the dimerisation indicates that the interaction is more complex than simply driven by hydrophobic effect, but mediated by local electrostatics. This would therefore go some way to rationalising the differences in aggregation due to ionic strength, pH and charged mutations.

A schematic representation of the possible dimeric interaction is shown in Figure 7-1.

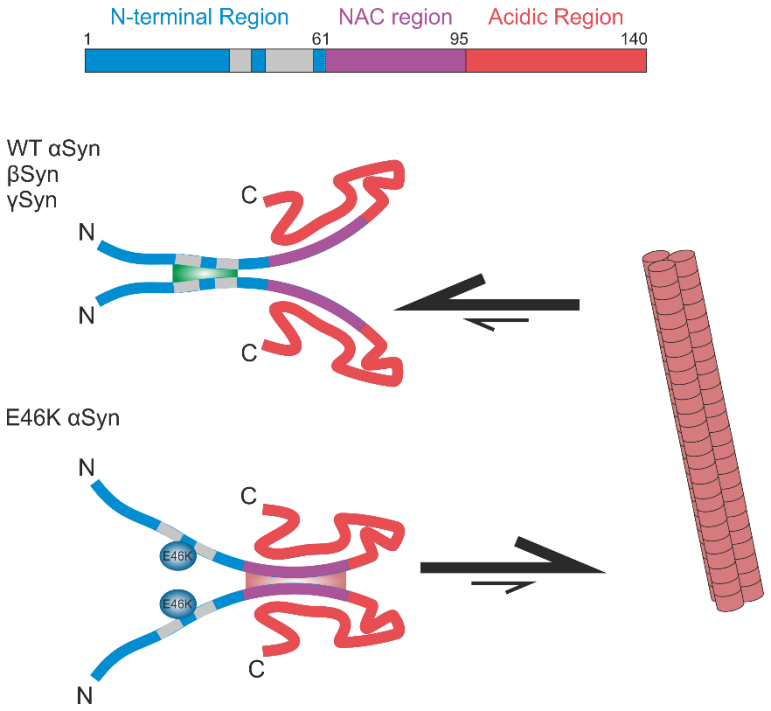


Figure 7-1. Schematic representation of one possible hypothesised mechanism of synuclein interactions. Schematic of the primary sequence of αSyn is shown at the top of the figure. The N-terminal region, NAC and acidic C-terminal are shown in blue, pink and red, respectively. Grey regions (residues 36-42 and 45-57) are regions with high Zyggregator scores. Colours are the same throughout. The top interaction schematic shows the possible physiological interaction of αSyn which is similar to the interactions of non-aggregation-prone synuclein homologues. The interaction is driven by hydrophobic stretches of the N-terminal region. E46K disrupts these interactions, increasing aggregation promoting interactions in the NAC region. This subsequently leads to increased fibril formation. The C-terminal region of the αSyn monomers in the dimeric species are shown as being structures.

7.6 Future directions

The novel dimeric structure revealed in this thesis requires more study. The data presented here presents a coarse model as to the nature of a structured dimer and its dependence on environmental conditions. Further structural analysis of the dimeric species would be difficult given the probable transient nature of any dimeric interaction in this pathway, however it would be an important area to study given that the conformation of the dimer may play a key role in the shifting the protein population to a less or more aggregation-prone state. In order to carry out such studies, these interactions would have to be trapped, possibly by crosslinking techniques similar to the techniques used in the literature proposing multimeric interactions of αSyn[46,255-259]. This could also be an approach for structural studies to confirm the region of intermolecular interaction proposed here to be in the N-terminus of the protein. Importantly, in this thesis we have shed light on the nature of a proposed protective interaction and therefore have a more coherent and rational starting point to derive experimental conditions in which the protective interaction is promoted.

Proposed here is some evidence of an interaction which may be protective to aggregation. However this has only been analysed in *in vitro* aggregation experiments. It would be important to study whether these interactions relate to reduced cytotoxicity and aggregation *in vivo*.

7.7 Final remarks

The data presented in this thesis has provided various novel insights into the initial intermolecular self-association events of αSyn. The data has suggested that the interaction observed is one of a protective, possibly physiological interaction that has varying degrees of force-resistant structure depending on the environmental conditions. If these conclusions and hypothesis are correct, there is a huge potential for targeting these novel events with therapeutics. Indeed, the only regulatory-approved drug that can slow the progression of a human amyloid disease (familial amyloid polyneuropathy, FAP) is Tafamidis[60]. In FAP, tetrameric species of TTR dissociate into misfolded monomers which leads to the formation of non-fibrillar and amyloid aggregates in the plasma[371]. Tetramer dissociation is the rate-limiting step and Tafamidis acts to bind and kinetically stabilise this species[60,371-373]. Exciting parallels can be drawn with this, the only example we have of a drug which slows amyloid progression, and with αSyn. If the interactions that have been proposed in this thesis are indeed protective and exist physiologically, the disease process could be targeted by promoting these interactions.

8 Appendix

8.1 Protein and peptide sequences

pL αSyn$_{71-82}$

MAMEEVTIKANLIFACGSTQTAEFKGTFEKATSEAYAYADTLKKDNGEWTVDVADKGGGARGSV
TGVTAVAQKTVGSARGGGYTLNFKFAGSGHHHHHH

αSyn$_{71-82}$

pL γSyn$_{71-82}$

MAMEEVTIKANLIFACGSTQTAEFKGTFEKATSEAYAYADTLKKDNGEWTVDVADKGGGARGSV
SSVNTVATKTVGSARGGGYTLNFKFAGSGHHHHHH

γSyn$_{71-82}$

pL GS

MAMEEVTIKANLIFACGSTQTAEFKGTFEKATSEAYAYADTLKKDNGEWTVDVADKGGGARGSG
SGSGSGSGSGSGSARGGGYTLNFKFAGSGHHHHHH

GS extension

αSyn$_{71-82}$ peptide

VTGVTAVAQKTV

pL γSyn$_{71-82}$ peptide

VSSVNTVATKTV

αSyn A140C

MDVFMKGLSKAKEGVVAAAEKTKQGVAEAAGKTKEGVLYVGSKTKEGVVHGVATVAEKTKEQVT
NVGGAVVTGVTAVAQKTVEGAGSIAAATGFVKKDQLGKNEEGAPQEGILEDMPVDPDNEAYEMP
SEEGYQDYEPEC

A140C

αSyn E46K A140C

MDVFMKGLSKAKEGVVAAAEKTKQGVAEAAGKTKEGVLYVGSKTKKGVVHGVATVAEKTKEQVT
NVGGAVVTGVTAVAQKTVEGAGSIAAATGFVKKDQLGKNEEGAPQEGILEDMPVDPDNEAYEMP
SEEGYQDYEPEC

E46K
A140C

βSyn A134C

MDVFMKGLSMAKEGVVAAAEKTKQGVTEAAEKTKEGVLYVGSKTREGVVQGVASVAEKTKEQAS
HLGGAVFSGAGNIAAATGLVKREEFPTDLKPEEVAQEAAEEPLIEPLMEPEGESYEDPPQEEYQ
EYEPEC

A134C

γSyn D127C

MDVFKKGFSIAKEGVVGAVEKTKQGVTEAAEKTKEGVMYVGAKTKENVVQSVTSVAEKTKEQAN
AVSEAVVSSVNTVATKTVEEAENIAVTSGVVRKEDLRPSAPQQEGEASKEKEEVAEEAQSGGC

D127C

References

1. Dobson, C.M. Principles of protein folding, misfolding and aggregation. *Semin. Cell Dev. Biol.* **15**, 3-16 (2004).
2. Anfinsen, C.B., Haber, E., Sela, M. & White, F.H. Kinetics of formation of native ribonuclease during oxidation of reduced polypeptide chain. *Proc. Natl. Acad. Sci. USA* **47**, 1309-& (1961).
3. Levinthal, C. Are there pathways for protein folding. *J. Chim. Phys. Phys.-Chim. Biol.* **65**, 44-+ (1968).
4. Zwanzig, R., Szabo, A. & Bagchi, B. Levinthals paradox. *Proc. Natl. Acad. Sci. USA* **89**, 20-22 (1992).
5. Englander, S.W. & Mayne, L. The case for defined protein folding pathways. *Proc. Natl. Acad. Sci. USA* **114**, 8253-8258 (2017).
6. Wetlaufe.Db. Nucleation, rapid folding, and globular intrachain regions in proteins. *Proc. Natl. Acad. Sci. USA* **70**, 697-701 (1973).
7. Tsong, T.Y., Baldwin, R.L. & Elson, E.L. Properties of the refolding and unfolding reactions of ribonuclease a. *Proc Natl Acad Sci U S A* **69**, 1809-12 (1972).
8. Fersht, A.R. Nucleation mechanisms in protein folding. *Curr. Opin. Struct. Biol.* **7**, 3-9 (1997).
9. Kim, P.S. & Baldwin, R.L. Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. *Annu. Rev. Biochem.* **51**, 459-489 (1982).
10. Baldwin, R.L. The nature of protein-folding pathways - the classical versus the new view. *J. Biomol. NMR* **5**, 103-109 (1995).
11. Karplus, M. & Weaver, D.L. Diffusion-collision model for protein folding. *Biopolymers* **18**, 1421-1437 (1979).
12. Karplus, M. & Weaver, D.L. Protein-folding dynamics - the diffusion-collision model and experimental-data. *Protein Sci.* **3**, 650-668 (1994).
13. Tanford, C. Contribution of hydrophobic interactions to stability of globular conformation of proteins. *J. Am. Chem. Soc.* **84**, 4240-& (1962).
14. Baldwin, R.L. How does protein folding get started. *Trends Biochem. Sci.* **14**, 291-294 (1989).
15. Jackson, S.E. & Fersht, A.R. Folding of chymotrypsin inhibitor-2 .1. Evidence for a 2-state transition. *Biochem.* **30**, 10428-10435 (1991).
16. Itzhaki, L.S., Otzen, D.E. & Fersht, A.R. The structure of the transition-state for folding of chymotrypsin inhibitor-2 analyzed by protein engineering methods - evidence for a nucleation-condensation mechanism for protein-folding. *J. Mol. Biol.* **254**, 260-288 (1995).
17. Viguera, A.R., Martinez, J.C., Filimonov, V.V., Mateo, P.L. & Serrano, L. Thermodynamic and kinetic-analysis of the sh3 domain of spectrin shows a 2-state folding transition. *Biochem.* **33**, 2142-2150 (1994).
18. Schindler, T., Herrler, M., Marahiel, M.A. & Schmid, F.X. Extremely rapid protein-folding in the absence of intermediates. *Nat. Struct. Biol.* **2**, 663-673 (1995).
19. Fersht, A.R. Optimization of rates of protein-folding - the nucleation-condensation mechanism and its implications. *Proc. Natl. Acad. Sci. USA* **92**, 10869-10873 (1995).
20. Daggett, V. & Fersht, A.R. Is there a unifying mechanism for protein folding? *Trends Biochem. Sci.* **28**, 18-25 (2003).
21. Gianni, S. et al. Unifying features in protein-folding mechanisms. *Proc. Natl. Acad. Sci. USA* **100**, 13286-13291 (2003).

22. Morris, E.R. & Searle, M.S. Overview of protein folding mechanisms: Experimental and theoretical approaches to probing energy landscapes. in *Current protocols in protein science* (John Wiley & Sons, Inc., 2001).
23. Dill, K.A. & Chan, H.S. From levinthal to pathways to funnels. *Nat. Struct. Biol.* **4**, 10-19 (1997).
24. Jahn, T.R. & Radford, S.E. Folding versus aggregation: Polypeptide conformations on competing pathways. *Archives of Biochemistry and Biophysics* **469**, 100-117 (2008).
25. Taverna, D.M. & Goldstein, R.A. Why are proteins marginally stable? *Prot. Struct. Func. Genet.* **46**, 105-109 (2002).
26. Henzler-Wildman, K. & Kern, D. Dynamic personalities of proteins. *Nature* **450**, 964-972 (2007).
27. Palmer, A.G. Nmr probes of molecular dynamics: Overview and comparison with other techniques. *Annu. Rev. Biophys. Biomol. Struct.* **30**, 129-155 (2001).
28. Eisenmesser, E.Z., Bosco, D.A., Akke, M. & Kern, D. Enzyme dynamics during catalysis. *Science* **295**, 1520-1523 (2002).
29. Gelis, I. et al. Structural basis for signal-sequence recognition by the translocase motor seca as determined by nmr. *Cell* **131**, 756-769 (2007).
30. Tokuriki, N. & Tawfik, D.S. Protein dynamism and evolvability. *Science* **324**, 203-207 (2009).
31. Williamson, M.P. & Potts, J.R. Intrinsically disordered proteins: Administration not executive. *Biochem. Soc. Trans.* **40**, 945-949 (2012).
32. Dunker, A.K., Silman, I., Uversky, V.N. & Sussman, J.L. Function and structure of inherently disordered proteins. *Curr. Opin. Struct. Biol.* **18**, 756-764 (2008).
33. Ramachandran, G.N., Ramakrishnan, C. & Sasisekharan, V. Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* **7**, 95-& (1963).
34. Dyson, H.J. & Wright, P.E. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* **6**, 197-208 (2005).
35. Uversky, V.N. Intrinsically disordered proteins from a to z. *Int. J. Biochem. Cell Biol.* **43**, 1090-1103 (2011).
36. Mittag, T. & Forman-Kay, J.D. Atomic-level characterization of disordered protein ensembles. *Curr. Opin. Struct. Biol.* **17**, 3-14 (2007).
37. Tompa, P. & Fuxreiter, M. Fuzzy complexes: Polymorphism and structural disorder in protein-protein interactions. *Trends Biochem. Sci.* **33**, 2-8 (2008).
38. Monod, J., Wyman, J. & Changeux, J.-P. On the nature of allosteric transitions: A plausible model. *J. Mol. Biol.* **12**, 88-118 (1965).
39. Dyson, H.J. & Wright, P.E. Coupling of folding and binding for unstructured proteins. *Curr. Opin. Struct. Biol.* **12**, 54-60 (2002).
40. Kuriyan, J. & Eisenberg, D. The origin of protein interactions and allostery in colocalization. *Nature* **450**, 983-990 (2007).
41. Neduva, V. & Russell, R.B. Linear motifs: Evolutionary interaction switches. *FEBS Lett.* **579**, 3342-3345 (2005).
42. Davidson, W.S., Jonas, A., Clayton, D.F. & George, J.M. Stabilization of alpha-synuclein secondary structure upon binding to synthetic membranes. *J. Biol. Chem.* **273**, 9443-9449 (1998).
43. Jo, E.J., McLaurin, J., Yip, C.M., St George-Hyslop, P. & Fraser, P.E. Alpha-synuclein membrane interactions and lipid specificity. *J. Biol. Chem.* **275**, 34328-34334 (2000).
44. Lee, H.J., Choi, C. & Lee, S.J. Membrane-bound alpha-synuclein has a high aggregation propensity and the ability to seed the aggregation of the cytosolic form. *J. Biol. Chem.* **277**, 671-678 (2002).
45. Tokuriki, N., Oldfield, C.J., Uversky, V.N., Berezovsky, I.N. & Tawfik, D.S. Do viral proteins possess unique biophysical features? *Trends Biochem. Sci.* **34**, 53-59 (2009).

46. Dettmer, U., Newman, A.J., von Saucken, V.E., Bartels, T. & Selkoe, D. Ktkegv repeat motifs are key mediators of normal alpha-synuclein tetramerization: Their mutation causes excess monomers and neurotoxicity. *Proc. Natl. Acad. Sci. USA* **112**, 9596-9601 (2015).
47. Kay, B.K., Williamson, M.P. & Sudol, P. The importance of being proline: The interaction of proline-rich motifs in signaling proteins with their cognate domains. *FASEB J.* **14**, 231-241 (2000).
48. Breidenbach, M.A. & Brunger, A.T. Substrate recognition strategy for botulinum neurotoxin serotype a. *Nature* **432**, 925-929 (2004).
49. Yoon, M.K., Mitrea, D.M., Ou, L. & Kriwacki, R.W. Cell cycle regulation by the intrinsically disordered proteins p21 and p27. *Biochem. Soc. Trans.* **40**, 981-988 (2012).
50. Chiti, F. & Dobson, C.M. Protein misfolding, functional amyloid, and human disease. in *Annual review of biochemistry*, Vol. 75 333-366 (2006).
51. Knowles, T.P.J., Vendruscolo, M. & Dobson, C.M. The amyloid state and its association with protein misfolding diseases. *Nat. Rev. Mol. Cell Biol.* **15**, 384-396 (2014).
52. Baldwin, A.J. et al. Metastability of native proteins and the phenomenon of amyloid formation. *J. Am. Chem. Soc.* **133**, 14160-14163 (2011).
53. Gazit, E. The "correctly folded" state of proteins: Is it a metastable state. *Angew. Chem.-Int. Edit.* **41**, 257-+ (2002).
54. Karamanos, T.K. et al. A population shift between sparsely populated folding intermediates determines amyloidogenicity. *J. Am. Chem. Soc.* **138**, 6271-6280 (2016).
55. Hartl, F.U. & Hayer-Hartl, M. Converging concepts of protein folding in vitro and in vivo. *Nat. Struct. Mol. Biol.* **16**, 574-581 (2009).
56. Dobson, C.M. Protein folding and misfolding. *Nature* **426**, 884-890 (2003).
57. Eisenberg, D. & Jucker, M. The amyloid state of proteins in human diseases. *Cell* **148**, 1188-1203 (2012).
58. Monsellier, E. & Chiti, F. Prevention of amyloid-like aggregation as a driving force of protein evolution. *Embo Reports* **8**, 737-742 (2007).
59. Uversky, V.N. Natively unfolded proteins: A point where biology waits for physics. *Protein Sci.* **11**, 739-756 (2002).
60. Johnson, S.M., Connelly, S., Fearns, C., Powers, E.T. & Kelly, J.W. The transthyretin amyloidoses: From delineating the molecular mechanism of aggregation linked to pathology to a regulatory-agency-approved drug. *J. Mol. Biol.* **421**, 185-203 (2012).
61. Hammarstrom, P., Jiang, X., Hurshman, A.R., Powers, E.T. & Kelly, J.W. Sequence-dependent denaturation energetics: A major determinant in amyloid disease diversity. *Proc. Natl. Acad. Sci. USA* **99**, 16427-16432 (2002).
62. Scarpioni, R. et al. Dialysis-related amyloidosis: Challenges and solutions. *Int. J. Nephrol. Renov. Dis.* **9**, 319-328 (2016).
63. Jahn, T.R., Parker, M.J., Homans, S.W. & Radford, S.E. Amyloid formation under physiological conditions proceeds via a native-like folding intermediate. *Nat. Struct. Mol. Biol.* **13**, 195-201 (2006).
64. Balch, W.E., Morimoto, R.I., Dillin, A. & Kelly, J.W. Adapting proteostasis for disease intervention. *Science* **319**, 916-919 (2008).
65. Hipp, M.S., Park, S.H. & Hartl, F.U. Proteostasis impairment in protein-misfolding and -aggregation diseases. *Trends Cell Biol.* **24**, 506-514 (2014).
66. Alzheimer's disease international. *World Alzheimer's report* (2016).
67. Dobson, C.M. Protein misfolding, evolution and disease. *Trends Biochem. Sci.* **24**, 329-332 (1999).
68. Greenwald, J. & Riek, R. On the possible amyloid origin of protein folds. *J. Mol. Biol.* **421**, 417-426 (2012).

69. Carny, O. & Gazit, E. A model for the role of short self-assembled peptides in the very early stages of the origin of life. *FASEB J.* **19**, 1051-1055 (2005).
70. DePas, W.H. & Chapman, M.R. Microbial manipulation of the amyloid fold. *Res. Microbiol.* **163**, 592-606 (2012).
71. Fowler, D.M., Koulov, A.V., Balch, W.E. & Kelly, J.W. Functional amyloid - from bacteria to humans. *Trends Biochem. Sci.* **32**, 217-224 (2007).
72. Fandrich, M. & Dobson, C.M. The behaviour of polyamino acids reveals an inverse side chain effect in amyloid structure formation. *EMBO J.* **21**, 5682-5690 (2002).
73. Virchow, R. Weitere mittheilungen über das vorkommen der pflanzlichen cellulose beim menschen. *Archiv. F. Pathol. Anat* **6**, 268-271 (1854).
74. Sipe, J.D. & Cohen, A.S. Review: History of the amyloid fibril. *J. Struct. Biol.* **130**, 88-98 (2000).
75. Friedreich, N. & Kekulé, A. Zur amyloidfrage. *Archiv. F. Pathol. Anat* **16**, 50-65 (1859).
76. Bennhold, H. Eine specifische amyloid-farbung mit kongorot. *Munch. Med. Wochenschr* **69**, 1537-1538 (1922).
77. Divry, P. & Florkin, M. The optic properties of amyloid. *Comptes Rendus Des Seances De La Societe De Biologie Et De Ses Filiales* **97**, 1808-1810 (1927).
78. Cohen, A.S. & Calkins, E. Electron microscopic observations on a fibrous component in amyloid of diverse origins. *Nature* **183**, 1202-1203 (1959).
79. Fruhling, L., Kempf, J. & Porte, A. Structure et formation de la substance amyloide dans lamylose experimentale de la souris - etude au microscope electronique. *Comptes Rendus Hebdomadaires Des Seances De L Academie Des Sciences* **250**, 1385-& (1960).
80. Caesar, R. Die feinstruktur von milz und leber bei experimenteller amyloidose. *Zeitschrift Fur Zellforschung Und Mikroskopische Anatomie* **52**, 653-673 (1960).
81. Cohen, A.S., Frensdor.A, Calkins, E. & Lamprecht, S. A study of fine structure of amyloid associated with familial mediterranean fever. *Am. J. Pathol.* **41**, 567-& (1962).
82. Geddes, A.J., Parker, K.D., Atkins, E.D.T. & Beighton, E. Cross-beta conformation in proteins. *J. Mol. Biol.* **32**, 343-& (1968).
83. Eanes, E.D. & Glenner, G.G. X-ray diffraction studies on amyloid filaments. *J. Histochem. Cytochem.* **16**, 673-& (1968).
84. Bonar, L., Cohen, A.S. & Skinner, M.M. Characterization of amyloid fibril as a cross-beta protein. *Proc. Soc. Exp. Biol. Med.* **131**, 1373-& (1969).
85. Glenner, G.G., Terry, W., Harada, M., Isersky, C. & Page, D. Amyloid fibril proteins - proof of homology with immunoglobulin light chains by sequence analyses. *Science* **172**, 1150-& (1971).
86. Benditt, E.P. & Eriksen, N. Chemical classes of amyloid substance. *Am. J. Pathol.* **65**, 231-& (1971).
87. Sletten, K., Westermark, P. & Natvig, J.B. Characterization of amyloid fibril proteins from medullary carcinoma of thyroid. *J. Exp. Med.* **143**, 993-998 (1976).
88. Benditt, E.P. & Eriksen, N. Amyloid. 3. A protein related to the subunit structure of human amyloid fibrils. *Proc Natl Acad Sci U S A* **55**, 308-16 (1966).
89. Glenner, G.G. & Wong, C.W. Alzheimer's disease: Initial report of the purification and characterization of a novel cerebrovascular amyloid protein. *Biochem. Biophys. Res. Commun.* **120**, 885-90 (1984).
90. Sipe, J.D. et al. Amyloid fibril proteins and amyloidosis: Chemical identification and clinical classification international society of amyloidosis 2016 nomenclature guidelines. *Amyloid-J. Protein Fold. Disord.* **23**, 209-213 (2016).
91. Sunde, M. et al. Common core structure of amyloid fibrils by synchrotron x-ray diffraction. *J. Mol. Biol.* **273**, 729-739 (1997).
92. Sawaya, M.R. et al. Atomic structures of amyloid cross-beta spines reveal varied steric zippers. *Nature* **447**, 453-457 (2007).

93. Fitzpatrick, A.W.P. et al. Atomic structure and hierarchical assembly of a cross-beta amyloid fibril. *Proc. Natl. Acad. Sci. USA* **110**, 5468-5473 (2013).
94. Makin, O.S. & Serpell, L.C. Structures for amyloid fibrils. *FEBS J.* **272**, 5950-5961 (2005).
95. Chamberlain, A.K. et al. Ultrastructural organization of amyloid fibrils by atomic force microscopy. *Biophys. J.* **79**, 3282-3293 (2000).
96. Petkova, A.T. et al. A structural model for alzheimer's beta-amyloid fibrils based on experimental constraints from solid state nmr. *Proc. Natl. Acad. Sci. USA* **99**, 16742-16747 (2002).
97. Sachse, C., Fandrich, M. & Grigorieff, N. Paired beta-sheet structure of an a beta(1-40) amyloid fibril revealed by electron microscopy. *Proc. Natl. Acad. Sci. USA* **105**, 7462-7466 (2008).
98. Wasmer, C. et al. Amyloid fibrils of the het-s(218-289) prion form a beta solenoid with a triangular hydrophobic core. *Science* **319**, 1523-1526 (2008).
99. Luhrs, T. et al. 3d structure of alzheimer's amyloid-beta(1-42) fibrils. *Proc. Natl. Acad. Sci. USA* **102**, 17342-17347 (2005).
100. Tuttle, M.D. et al. Solid-state nmr structure of a pathogenic fibril of full-length human alpha-synuclein. *Nat. Struct. Mol. Biol.* **23**, 409-415 (2016).
101. Serpell, L. Amyloid structure. in *Amyloids in health and disease*, Vol. 56 (ed. Perrett, S.) 1-10 (Portland Press Ltd, London, 2014).
102. Fitzpatrick, A.W.P. et al. Cryo-em structures of tau filaments from alzheimer's disease. *Nature* **547**, 185-+ (2017).
103. Walti, M.A. et al. Atomic-resolution structure of a disease-relevant a beta(1-42) amyloid fibril. *Proc. Natl. Acad. Sci. USA* **113**, E4976-E4984 (2016).
104. Cremades, N. et al. Direct observation of the interconversion of normal and toxic forms of alpha-synuclein. *Cell* **149**, 1048-1059 (2012).
105. Nettleton, E.J. et al. Characterization of the oligomeric states of insulin in self-assembly and amyloid fibril formation by mass spectrometry. *Biophys. J.* **79**, 1053-1065 (2000).
106. Smith, D.P., Radford, S.E. & Ashcroft, A.E. Elongated oligomers in beta(2)-microglobulin amyloid assembly revealed by ion mobility spectrometry-mass spectrometry. *Proc. Natl. Acad. Sci. USA* **107**, 6794-6798 (2010).
107. Bernstein, S.L. et al. Amyloid-beta protein oligomerization and the importance of tetramers and dodecamers in the aetiology of alzheimer's disease. *Nat. Chem.* **1**, 326-331 (2009).
108. Narayan, P. et al. The extracellular chaperone clusterin sequesters oligomeric forms of the amyloid-beta(1-40) peptide. *Nat. Struct. Mol. Biol.* **19**, 79-U97 (2012).
109. Stefani, M. Protein misfolding and aggregation: New examples in medicine and biology of the dark side of the protein world. *Biochim. Biophys. Acta Mol. Basis Dis.* **1739**, 5-25 (2004).
110. Kopito, R.R. Aggresomes, inclusion bodies and protein aggregation. *Trends Cell Biol.* **10**, 524-530 (2000).
111. Fandrich, M. Absolute correlation between lag time and growth rate in the spontaneous formation of several amyloid-like aggregates and fibrils. *J. Mol. Biol.* **365**, 1266-1270 (2007).
112. Cohen, S.I.A., Vendruscolo, M., Dobson, C.M. & Knowles, T.P.J. From macroscopic measurements to microscopic mechanisms of protein aggregation. *J. Mol. Biol.* **421**, 160-171 (2012).
113. Knowles, T.P.J. et al. An analytical solution to the kinetics of breakable filament assembly. *Science* **326**, 1533-1537 (2009).
114. Cohen, S.I.A. et al. Proliferation of amyloid-beta 42 aggregates occurs through a secondary nucleation mechanism. *Proc. Natl. Acad. Sci. USA* **110**, 9758-9763 (2013).

115. Eichner, T. & Radford, S.E. A diversity of assembly mechanisms of a generic amyloid fold. *Mol. Cell* **43**, 8-18 (2011).
116. Chia, S. et al. Monomeric and fibrillar alpha-synuclein exert opposite effects on the catalytic cycle that promotes the proliferation of a beta 42 aggregates. *Proc. Natl. Acad. Sci. USA* **114**, 8005-8010 (2017).
117. Gaspar, R. et al. Secondary nucleation of monomers on fibril surface dominates alpha-synuclein aggregation and provides autocatalytic amyloid amplification. *Q. Rev. Biophys.* **50**, 12 (2017).
118. Flagmeier, P. et al. Mutations associated with familial parkinson's disease alter the initiation and amplification steps of alpha-synuclein aggregation. *Proc. Natl. Acad. Sci. USA* **113**, 10328-10333 (2016).
119. Meisl, G. et al. Molecular mechanisms of protein aggregation from global fitting of kinetic models. *Nat. Protoc.* **11**, 252-272 (2016).
120. Bolognesi, B. et al. Single point mutations induce a switch in the molecular mechanism of the aggregation of the alzheimer's disease associated a beta(42) peptide. *ACS Chem. Biol.* **9**, 378-382 (2014).
121. Selkoe, D.J. & Hardy, J. The amyloid hypothesis of alzheimer's disease at 25years. *Embo Molecular Medicine* **8**, 595-608 (2016).
122. Haass, C. & Selkoe, D.J. Soluble protein oligomers in neurodegeneration: Lessons from the alzheimer's amyloid beta-peptide. *Nat. Rev. Mol. Cell Biol.* **8**, 101-112 (2007).
123. Karran, E., Mercken, M. & De Strooper, B. The amyloid cascade hypothesis for alzheimer's disease: An appraisal for the development of therapeutics. *Nat. Rev. Drug Discov.* **10**, 698-U1600 (2011).
124. Lue, L.F. et al. Soluble amyloid beta peptide concentration as a predictor of synaptic change in alzheimer's disease. *Am. J. Pathol.* **155**, 853-862 (1999).
125. Soto, C. Unfolding the role of protein misfolding in neurodegenerative diseases. *Nature. Rev. Neurosci.* **4**, 49-60 (2003).
126. Stefani, M. & Dobson, C.M. Protein aggregation and aggregate toxicity: New insights into protein folding, misfolding diseases and biological evolution. *J. Mol. Med.* **81**, 678-699 (2003).
127. Baglioni, S. et al. Prefibrillar amyloid aggregates could be generic toxins in higher organisms. *J. Neurosci.* **26**, 8160-8167 (2006).
128. Conway, K.A. et al. Acceleration of oligomerization, not fibrillization, is a shared property of both alpha-synuclein mutations linked to early-onset parkinson's disease: Implications for pathogenesis and therapy. *Proc. Natl. Acad. Sci. USA* **97**, 571-576 (2000).
129. Zhu, Y.J., Lin, H. & Lal, R. Fresh and nonfibrillar amyloid beta protein(1-40) induces rapid cellular degeneration in aged human fibroblasts: Evidence for a beta p-channel-mediated cellular toxicity. *FASEB J.* **14**, 1244-1254 (2000).
130. Nilsberth, C. et al. The 'arctic' app mutation (e693g) causes alzheimer's disease by enhanced a beta protofibril formation. *Nat. Neurosci.* **4**, 887-893 (2001).
131. Sousa, M.M., Cardoso, I., Fernandes, R., Guimaraes, A. & Saraiva, M.J. Deposition of transthyretin in early stages of familial amyloidotic polyneuropathy - evidence for toxicity of nonfibrillar aggregates. *Am. J. Pathol.* **159**, 1993-2000 (2001).
132. Verma, M., Vats, A. & Taneja, V. Toxic species in amyloid disorders: Oligomers or mature fibrils. *Ann. Indian Acad. Neurol.* **18**, 138-145 (2015).
133. Walsh, D.M. et al. Naturally secreted oligomers of amyloid beta protein potently inhibit hippocampal long-term potentiation in vivo. *Nature* **416**, 535-539 (2002).
134. Lesne, S. et al. A specific amyloid-beta protein assembly in the brain impairs memory. *Nature* **440**, 352-357 (2006).

135. Campioni, S. et al. A causative link between the structure of aberrant protein oligomers and their toxicity. *Nat. Chem. Biol.* **6**, 140-147 (2010).
136. Cheon, M. et al. Structural reorganisation and potential toxicity of oligomeric species formed during the assembly of amyloid fibrils. *PLoS Comp. Biol.* **3**, 1727-1738 (2007).
137. Bolognesi, B. et al. Ans binding reveals common features of cytotoxic amyloid species. *ACS Chem. Biol.* **5**, 735-740 (2010).
138. Olzscha, H. et al. Amyloid-like aggregates sequester numerous metastable proteins with essential cellular functions. *Cell* **144**, 67-78 (2011).
139. Narayan, P. et al. Single molecule characterization of the interactions between amyloid-beta peptides and the membranes of hippocampal cells. *J. Am. Chem. Soc.* **135**, 1491-1498 (2013).
140. Lashuel, H.A. & Lansbury, P.T. Are amyloid diseases caused by protein aggregates that mimic bacterial pore-forming toxins? *Q. Rev. Biophys.* **39**, 167-201 (2006).
141. Lashuel, H.A. et al. Alpha-synuclein, especially the parkinson's disease-associated mutants, forms pore-like annular and tubular protofibrils. *J. Mol. Biol.* **322**, 1089-1102 (2002).
142. Wahlbom, M. et al. Fibrillogenic oligomers of human cystatin c are formed by propagated domain swapping. *J. Biol. Chem.* **282**, 18318-18326 (2007).
143. Kayed, R. et al. Annular protofibrils are a structurally and functionally distinct type of amyloid oligomer. *J. Biol. Chem.* **284**, 4230-4237 (2009).
144. Serio, T.R. et al. Nucleated conformational conversion and the replication of conformational information by a prion determinant. *Science* **289**, 1317-1321 (2000).
145. Bucciantini, M. et al. Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases. *Nature* **416**, 507-511 (2002).
146. Kirkitadze, M.D., Bitan, G. & Teplow, D.B. Paradigm shifts in alzheimer's disease and other neuro degenerative disorders: The emerging role of oligomeric assemblies. *J. Neurosci. Res.* **69**, 567-577 (2002).
147. Stroud, J.C., Liu, C., Teng, P.K. & Eisenberg, D. Toxic fibrillar oligomers of amyloid-beta have cross-beta structure. *Proc. Natl. Acad. Sci. USA* **109**, 7717-7722 (2012).
148. Orte, A. et al. Direct characterization of amyloidogenic oligomers by single-molecule fluorescence. *Proc. Natl. Acad. Sci. USA* **105**, 14424-14429 (2008).
149. Oliveira, C.L.P. et al. A saxs study of glucagon fibrillation. *J. Mol. Biol.* **387**, 147-161 (2009).
150. Ahmed, M. et al. Structural conversion of neurotoxic amyloid-beta(1-42) oligomers to fibrils. *Nat. Struct. Mol. Biol.* **17**, 561-U56 (2010).
151. Stefani, M. Biochemical and biophysical features of both oligomer/fibril and cell membrane in amyloid cytotoxicity. *FEBS J.* **277**, 4602-4613 (2010).
152. Tipping, K.W. et al. Ph-induced molecular shedding drives the formation of amyloid fibril-derived oligomers. *Proc. Natl. Acad. Sci. USA* **112**, 5691-5696 (2015).
153. Labbadia, J. & Morimoto, R.I. The biology of proteostasis in aging and disease. in *Annual review of biochemistry, vol 84*, Vol. 84 (ed. Kornberg, R.D.) 435-464 (Annual Reviews, Palo Alto, 2015).
154. Bertram, L. & Tanzi, R.E. The genetic epidemidogy of neurodegenerative disease. *J. Clin. Invest.* **115**, 1449-1457 (2005).
155. Citron, M. Alzheimer's disease: Strategies for disease modification. *Nat. Rev. Drug Discov.* **9**, 387-398 (2010).
156. Westermark, P., Andersson, A. & Westermark, G.T. Islet amyloid polypeptide, islet amyloid, and diabetes mellitus. *Physiol. Rev.* **91**, 795-826 (2011).
157. Yan, S.D. et al. Rage and amyloid-beta peptide neurotoxicity in alzheimer's disease. *Nature* **382**, 685-691 (1996).

158. Yan, S.D. et al. Receptor-dependent cell stress and amyloid accumulation in systemic amyloidosis. *Nat. Med.* **6**, 643-651 (2000).
159. Engel, M.F.M. Membrane permeabilization by islet amyloid polypeptide. *Chem. Phys. Lipids* **160**, 1-10 (2009).
160. Engel, M.F.M. et al. Membrane damage by human islet amyloid polypeptide through fibril growth at the membrane. *Proc. Natl. Acad. Sci. USA* **105**, 6033-6038 (2008).
161. Pieri, L., Madiona, K., Bousset, L. & Melki, R. Fibrillar alpha-synuclein and huntingtin exon 1 assemblies are toxic to the cells. *Biophys. J.* **102**, 2894-2905 (2012).
162. Lin, H., Bhatia, R. & Lal, R. Amyloid beta protein forms ion channels: Implications for alzheimer's disease pathophysiology. *FASEB J.* **15**, 2433-2444 (2001).
163. Lin, M.X., Mirzabekov, T. & Kagan, B.L. Channel formation by a neurotoxic prion protein fragment. *J. Biol. Chem.* **272**, 44-47 (1997).
164. Cummings, C.J. et al. Chaperone suppression of aggregation and altered subcellular proteasome localization imply protein misfolding in sca1. *Nat. Genet.* **19**, 148-154 (1998).
165. Cook, C. & Petrucelli, L. A critical evaluation of the ubiquitin-proteasome system in parkinson's disease. *Biochim. Biophys. Acta Mol. Basis Dis.* **1792**, 664-675 (2009).
166. Douglas, P.M. & Dillin, A. Protein homeostasis and aging in neurodegeneration. *J. Cell Biol.* **190**, 719-729 (2010).
167. Sherman, M.Y. & Goldberg, A.L. Cellular defenses against unfolded proteins: A cell biologist thinks about neurodegenerative diseases. *Neuron* **29**, 15-32 (2001).
168. Morimoto, R.I. Proteotoxic stress and inducible chaperone networks in neurodegenerative disease and aging. *Genes Dev.* **22**, 1427-1438 (2008).
169. Cattaneo, E. et al. Loss of normal huntingtin function: New developments in huntington's disease research. *Trends Neurosci.* **24**, 182-188 (2001).
170. Reaume, A.G. et al. Motor neurons in cu/zn superoxide dismutase-deficient mice develop normally but exhibit enhanced cell death after axonal injury. *Nat. Genet.* **13**, 43-47 (1996).
171. Friedrich, R.P. et al. Mechanism of amyloid plaque formation suggests an intracellular basis of a beta pathogenicity. *Proc. Natl. Acad. Sci. USA* **107**, 1942-1947 (2010).
172. Milanesi, L. et al. Direct three-dimensional visualization of membrane disruption by amyloid fibrils. *Proc. Natl. Acad. Sci. USA* **109**, 20455-20460 (2012).
173. Lashuel, H.A., Hartley, D., Petre, B.M., Walz, T. & Lansbury, P.T. Neurodegenerative disease - amyloid pores from pathogenic mutations. *Nature* **418**, 291-291 (2002).
174. Olanow, C.W. & Brundin, P. Parkinson's disease and alpha synuclein: Is parkinson's disease a prion-like disorder? *Mov. Disord.* **28**, 31-40 (2013).
175. Weinreb, P.H., Zhen, W.G., Poon, A.W., Conway, K.A. & Lansbury, P.T. Nacp, a protein implicated in alzheimer's disease and learning, is natively unfolded. *Biochem.* **35**, 13709-13715 (1996).
176. Iwai, A. et al. The precursor protein of non-a-beta component of alzheimers-disease amyloid is a presynaptic protein of the central-nervous-system. *Neuron* **14**, 467-475 (1995).
177. Jakes, R., Spillantini, M.G. & Goedert, M. Identification of 2 distinct synucleins from human brain. *FEBS Lett.* **345**, 27-32 (1994).
178. Withers, G.S., George, J.M., Banker, G.A. & Clayton, D.F. Delayed localization of synelfin (synuclein, nacp) to presynaptic terminals in cultured rat hippocampal neurons. *Dev. Brain Res.* **99**, 87-94 (1997).
179. Wong, Y.C. & Krainc, D. Alpha-synuclein toxicity in neurodegeneration: Mechanism and therapeutic strategies. *Nat. Med.* **23**, 151-163 (2017).

180. Maroteaux, L., Campanelli, J.T. & Scheller, R.H. Synuclein - a neuron-specific protein localized to the nucleus and presynaptic nerve-terminal. *J. Neurosci.* **8**, 2804-2815 (1988).
181. Polymeropoulos, M.H. et al. Mutation in the alpha-synuclein gene identified in families with parkinson's disease. *Science* **276**, 2045-2047 (1997).
182. Spillantini, M.G. et al. Alpha-synuclein in lewy bodies. *Nature* **388**, 839-840 (1997).
183. Villar-Pique, A., da Fonseca, T.L. & Outeiro, T.F. Structure, function and toxicity of alpha-synuclein: The bermuda triangle in synucleinopathies. *J. Neurochem.* **139**, 240-255 (2016).
184. Lashuel, H.A., Overk, C.R., Oueslati, A. & Masliah, E. The many faces of alpha-synuclein: From structure and toxicity to therapeutic target. *Nature. Rev. Neurosci.* **14**, 38-48 (2013).
185. Abeliovich, A. et al. Mice lacking alpha-synuclein display functional deficits in the nigrostriatal dopamine system. *Neuron* **25**, 239-252 (2000).
186. Cabin, D.E. et al. Synaptic vesicle depletion correlates with attenuated synaptic responses to prolonged repetitive stimulation in mice lacking alpha-synuclein. *J. Neurosci.* **22**, 8797-8807 (2002).
187. Jin, J.H. et al. Identification of novel proteins associated with both alpha-synuclein and dj-1. *Mol. Cell. Preoteomics* **6**, 845-859 (2007).
188. McLean, P.J., Kawamata, H., Ribich, S. & Hyman, B.T. Membrane association and protein conformation of alpha-synuclein in intact neurons - effect of parkinson's disease-linked mutations. *J. Biol. Chem.* **275**, 8812-8816 (2000).
189. Kahle, P.J. et al. Subcellular localization of wild-type and parkinson's disease-associated mutant alpha-synuclein in human and transgenic mouse brain. *J. Neurosci.* **20**, 6365-6373 (2000).
190. Zhang, L. et al. Semi-quantitative analysis of alpha-synuclein in subcellular pools of rat brain neurons: An immunogold electron microscopic study using a c-terminal specific monoclonal antibody. *Brain Res.* **1244**, 40-52 (2008).
191. Lee, S.J., Jeon, H. & Kandror, K.V. Alpha-synuclein is localized in a subpopulation of rat brain synaptic vesicles. *Acta Neurobiol. Exp.* **68**, 509-515 (2008).
192. Murphy, D.D., Rueter, S.M., Trojanowski, J.Q. & Lee, V.M.Y. Synucleins are developmentally expressed, and alpha-synuclein regulates the size of the presynaptic vesicular pool in primary hippocampal neurons. *J. Neurosci.* **20**, 3214-3220 (2000).
193. Yavich, L., Tanila, H., Vepsalainen, S. & Jakala, P. Role of alpha-synuclein in presynaptic dopamine recruitment. *J. Neurosci.* **24**, 11165-11170 (2004).
194. Scott, D.A. et al. A pathologic cascade leading to synaptic dysfunction in alpha-synuclein-induced neurodegeneration. *J. Neurosci.* **30**, 8083-8095 (2010).
195. Nemani, V.M. et al. Increased expression of alpha-synuclein reduces neurotransmitter release by inhibiting synaptic vesicle reclustering after endocytosis. *Neuron* **65**, 66-79 (2010).
196. Gaugler, M.N. et al. Nigrostriatal overabundance of alpha-synuclein leads to decreased vesicle density and deficits in dopamine release that correlate with reduced motor activity. *Acta Neuropathol.* **123**, 653-669 (2012).
197. Scott, D. & Roy, S. Alpha-synuclein inhibits intersynaptic vesicle mobility and maintains recycling-pool homeostasis. *J. Neurosci.* **32**, 10129-10135 (2012).
198. Sudhof, T.C. The synaptic vesicle cycle. *Annu. Rev. Neurosci.* **27**, 509-547 (2004).
199. Burré, J. et al. A-synuclein promotes snare-complex assembly in vivo and in vitro. *Science* **329**, 1663-1667 (2010).
200. Fusco, G. et al. Structural basis of synaptic vesicle assembly promoted by alpha-synuclein. *Nat. Commun.* **7**(2016).

201. Chandra, S. et al. Double-knockout mice for alpha- and beta-synucleins: Effect on synaptic functions. *Proc. Natl. Acad. Sci. USA* **101**, 14966-14971 (2004).
202. Chandra, S., Gallardo, G., Fernandez-Chacon, R., Schluter, O.M. & Sudhof, T.C. Alpha-synuclein cooperates with csp alpha in preventing neurodegeneration. *Cell* **123**, 383-396 (2005).
203. Ueda, K. et al. Molecular-cloning of cdna-encoding an unrecognized component of amyloid in alzheimer-disease. *Proc. Natl. Acad. Sci. USA* **90**, 11282-11286 (1993).
204. Giasson, B.I., Murray, I.V.J., Trojanowski, J.Q. & Lee, V.M.Y. A hydrophobic stretch of 12 amino acid residues in the middle of alpha-synuclein is essential for filament assembly. *J. Biol. Chem.* **276**, 2380-2386 (2001).
205. Bussell, R., Jr. & Eliezer, D. A structural and functional role for 11-mer repeats in alpha-synuclein and other exchangeable lipid binding proteins. *J. Mol. Biol.* **329**, 763-78 (2003).
206. Ulmer, T.S., Bax, A., Cole, N.B. & Nussbaum, R.L. Structure and dynamics of micelle-bound human alpha-synuclein. *J. Biol. Chem.* **280**, 9595-9603 (2005).
207. Jensen, P.H., Nielsen, M.S., Jakes, R., Dotti, G. & Goedert, M. Binding of alpha-synuclein to brain vesicles is abolished by familial parkinson's disease mutation. *J. Biol. Chem.* **273**, 26292-26294 (1998).
208. Hoyer, W., Cherny, D., Subramaniam, V. & Jovin, T.M. Impact of the acidic c-terminal region comprising amino acids 109-140 on alpha-synuclein aggregation in vitro. *Biochem.* **43**, 16233-16242 (2004).
209. Breydo, L., Wu, J.W. & Uversky, V.N. Alpha-synuclein misfolding and parkinson's disease. *Biochim. Biophys. Acta Mol. Basis Dis.* **1822**, 261-285 (2012).
210. Uversky, V.N., Li, J. & Fink, A.L. Evidence for a partially folded intermediate in alpha-synuclein fibril formation. *J. Biol. Chem.* **276**, 10737-10744 (2001).
211. Eliezer, D., Kutluay, E., Bussell, R. & Browne, G. Conformational properties of alpha-synuclein in its free and lipid-associated states. *J. Mol. Biol.* **307**, 1061-1073 (2001).
212. Cho, M.K. et al. Amino acid bulkiness defines the local conformations and dynamics of natively unfolded alpha-synuclein and tau. *J. Am. Chem. Soc.* **129**, 3032-+ (2007).
213. Allison, J.R., Varnai, P., Dobson, C.M. & Vendruscolo, M. Determination of the free energy landscape of alpha-synuclein using spin label nuclear magnetic resonance measurements. *J. Am. Chem. Soc.* **131**, 18314-18326 (2009).
214. Salmon, L. et al. Nmr characterization of long-range order in intrinsically disordered proteins. *J. Am. Chem. Soc.* **132**, 8407-8418 (2010).
215. Bertoncini, C.W. et al. Release of long-range tertiary interactions potentiates aggregation of natively unstructured alpha-synuclein. *Proc. Natl. Acad. Sci. USA* **102**, 1430-1435 (2005).
216. Bertoncini, C.W., Fernandez, C.O., Griesinger, C., Jovin, T.M. & Zweckstetter, M. Familial mutants of alpha-synuclein with increased neurotoxicity have a destabilized conformation. *J. Biol. Chem.* **280**, 30649-30652 (2005).
217. Rao, J.N., Jao, C.C., Hegde, B.G., Langen, R. & Ulmer, T.S. A combinatorial nmr and epr approach for evaluating the structural ensemble of partially folded proteins. *J. Am. Chem. Soc.* **132**, 8657-8668 (2010).
218. Lee, J.C., Lai, B.T., Kozak, J.J., Gray, H.B. & Winkler, J.R. Alpha-synuclein tertiary contact dynamics. *J. Phys. Chem. B.* **111**, 2107-2112 (2007).
219. Dedmon, M.M., Lindorff-Larsen, K., Christodoulou, J., Vendruscolo, M. & Dobson, C.M. Mapping long-range interactions in alpha-synuclein using spin-label nmr and ensemble molecular dynamics simulations. *J. Am. Chem. Soc.* **127**, 476-477 (2005).
220. Cho, M.K. et al. Structural characterization of alpha-synuclein in an aggregation prone state. *Protein Sci.* **18**, 1840-1846 (2009).

221. Wu, K.P., Weinstock, D.S., Narayanan, C., Levy, R.M. & Baum, J. Structural reorganization of alpha-synuclein at low ph observed by nmr and remd simulations. *J. Mol. Biol.* **391**, 784-796 (2009).
222. Kruger, R. et al. Ala30pro mutation in the gene encoding alpha-synuclein in parkinson's disease. *Nat. Genet.* **18**, 106-108 (1998).
223. Zarranz, J.J. et al. The new mutation, e46k, of alpha-synuclein causes parkinson and lewy body dementia. *Ann. Neurol.* **55**, 164-173 (2004).
224. Appel-Cresswell, S. et al. Alpha-synuclein p.H50q, a novel pathogenic mutation for parkinson's disease. *Mov. Disord.* **28**, 811-813 (2013).
225. Proukakis, C. et al. A novel alpha-synuclein missense mutation in parkinson disease. *Neurology* **80**, 1062-1064 (2013).
226. Lesage, S. et al. G51d alpha-synuclein mutation causes a novel parkinsonian-pyramidal syndrome. *Ann. Neurol.* **73**, 459-471 (2013).
227. Pasanen, P. et al. A novel alpha-synuclein mutation a53e associated with atypical multiple system atrophy and parkinson's disease-type pathology. *Neurobiol. Aging* **35**, 5 (2014).
228. Singleton, A.B. et al. Alpha-synuclein locus triplication causes parkinson's disease. *Science* **302**, 841-841 (2003).
229. Chartier-Harlin, M.C. et al. Alpha-synuclein locus duplication as a cause of familial parkinson's disease. *Lancet* **364**, 1167-1169 (2004).
230. Simon-Sanchez, J. et al. Genome-wide association study reveals genetic risk underlying parkinson's disease. *Nat. Genet.* **41**, 1308-U68 (2009).
231. Nalls, M.A. et al. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for parkinson's disease. *Nat. Genet.* **46**, 989-+ (2014).
232. Soldner, F. et al. Parkinson-associated risk variant in distal enhancer of alpha-synuclein modulates target gene expression. *Nature* **533**, 95-+ (2016).
233. Kalia, L.V. & Lang, A.E. Parkinson's disease. *Lancet* **386**, 896-912 (2015).
234. Spillantini, M.G., Crowther, R.A., Jakes, R., Hasegawa, M. & Goedert, M. Alpha-synuclein in filamentous inclusions of lewy bodies from parkinson's disease and dementia with lewy bodies. *Proc. Natl. Acad. Sci. USA* **95**, 6469-6473 (1998).
235. Spillantini, M.G. et al. Filamentous alpha-synuclein inclusions link multiple system atrophy with parkinson's disease and dementia with lewy bodies. *Neurosci. Lett.* **251**, 205-208 (1998).
236. Shachar, T. et al. Lysosomal storage disorders and parkinson's disease: Gaucher disease and beyond. *Mov. Disord.* **26**, 1593-1604 (2011).
237. Bachhuber, T. et al. Inhibition of amyloid-beta plaque formation by alpha-synuclein. *Nat. Med.* **21**, 802-+ (2015).
238. Guo, J.L. et al. Distinct alpha-synuclein strains differentially promote tau inclusions in neurons. *Cell* **154**, 103-117 (2013).
239. Yoshimoto, M. et al. Nacp, the precursor protein of the nonamyloid beta/a4 protein (a-beta) component of alzheimer-disease amyloid, binds a-beta and stimulates a-beta aggregation. *Proc. Natl. Acad. Sci. USA* **92**, 9141-9145 (1995).
240. Masliah, E. et al. Beta-amyloid peptides enhance alpha-synuclein accumulation and neuronal deficits in a transgenic mouse model linking alzheimer's disease and parkinson's disease. *Proc. Natl. Acad. Sci. USA* **98**, 12245-12250 (2001).
241. Giasson, B.I. et al. Initiation and synergistic fibrillization of tau and alpha-synuclein. *Science* **300**, 636-640 (2003).
242. Vilar, M. et al. The fold of alpha-synuclein fibrils. *Proc. Natl. Acad. Sci. USA* **105**, 8637-8642 (2008).
243. Rodriguez, J.A. et al. Structure of the toxic core of alpha-synuclein from invisible crystals. *Nature* **525**, 486-+ (2015).

244. Peelaerts, W. et al. Alpha-synuclein strains cause distinct synucleinopathies after local and systemic administration. *Nature* **522**, 340-+ (2015).
245. Luk, K.C. et al. Pathological alpha-synuclein transmission initiates parkinson-like neurodegeneration in nontransgenic mice. *Science* **338**, 949-953 (2012).
246. Masuda-Suzukake, M. et al. Prion-like spreading of pathological alpha-synuclein in brain. *Brain* **136**, 1128-1138 (2013).
247. Recasens, A. et al. Lewy body extracts from parkinson disease brains trigger alpha-synuclein pathology and neurodegeneration in mice and monkeys. *Ann. Neurol.* **75**, 351-362 (2014).
248. Bousset, L. et al. Structural and functional characterization of two alpha-synuclein strains. *Nat. Commun.* **4**(2013).
249. Apetri, M.M., Maiti, N.C., Zagorski, M.G., Carey, P.R. & Anderson, V.E. Secondary structure of alpha-synuclein oligomers: Characterization by raman and atomic force microscopy. *J. Mol. Biol.* **355**, 63-71 (2006).
250. Ghosh, D. et al. Structure based aggregation studies reveal the presence of helix-rich intermediate during alpha-synuclein aggregation. *Sci. Rep.* **5**(2015).
251. Roberts, H.L. & Brown, D.R. Seeking a mechanism for the toxicity of oligomeric alpha-synuclein. *Biomolecules* **5**, 282-305 (2015).
252. Karpinar, D.P. et al. Pre-fibrillar alpha-synuclein variants with impaired beta-structure increase neurotoxicity in parkinson's disease models. *EMBO J.* **28**, 3256-3268 (2009).
253. Winner, B. et al. In vivo demonstration that alpha-synuclein oligomers are toxic. *Proc. Natl. Acad. Sci. USA* **108**, 4194-4199 (2011).
254. Rockenstein, E. et al. Accumulation of oligomer-prone alpha-synuclein exacerbates synaptic and neuronal degeneration in vivo. *Brain* **137**, 1496-1513 (2014).
255. Bartels, T., Choi, J.G. & Selkoe, D.J. Alpha-synuclein occurs physiologically as a helically folded tetramer that resists aggregation. *Nature* **477**, 107-U123 (2011).
256. Wang, W. et al. A soluble alpha-synuclein construct forms a dynamic tetramer. *Proc. Natl. Acad. Sci. USA* **108**, 17797-17802 (2011).
257. Dettmer, U., Newman, A.J., Luth, E.S., Bartels, T. & Selkoe, D. In vivo cross-linking reveals principally oligomeric forms of alpha-synuclein and beta-synuclein in neurons and non-neural cells. *J. Biol. Chem.* **288**, 6371-6385 (2013).
258. Dettmer, U. et al. Parkinson-causing alpha-synuclein missense mutations shift native tetramers to monomers as a mechanism for disease initiation. *Nat. Commun.* **6**(2015).
259. Luth, E.S., Bartels, T., Dettmer, U., Kim, N.C. & Selkoe, D.J. Purification of alpha-synuclein from human brain reveals an instability of endogenous multimers as the protein approaches purity. *Biochem.* **54**, 279-292 (2015).
260. Li, J., Uversky, V.N. & Fink, A.L. Conformational behavior of human alpha-synuclein is modulated by familial parkinson's disease point mutations a30p and a53t. *Neurotoxicology* **23**, 553-567 (2002).
261. Li, J., Uversky, V.N. & Fink, A.L. Effect of familial parkinson's disease point mutations a30p and a53t on the structural properties, aggregation, and fibrillation of human alpha-synuclein. *Biochem.* **40**, 11604-11613 (2001).
262. Greenbaum, E.A. et al. The e46k mutation in alpha-synuclein increases amyloid fibril formation. *J. Biol. Chem.* **280**, 7800-7807 (2005).
263. Tsigelny, I.F. et al. Role of alpha-synuclein penetration into the membrane in the mechanisms of oligomer pore formation. *FEBS J.* **279**, 1000-1013 (2012).
264. Jo, E.J., Fuller, N., Rand, R.P., St George-Hyslop, P. & Fraser, P.E. Defective membrane interactions of familial parkinson's disease mutant a30p alpha-synuclein. *J. Mol. Biol.* **315**, 799-807 (2002).

265. Fares, M.B. et al. The novel parkinson's disease linked mutation g51d attenuates in vitro aggregation and membrane binding of alpha-synuclein, and enhances its secretion and nuclear localization in cells. *Hum. Mol. Genet.* **23**, 4491-4509 (2014).
266. Ghosh, D. et al. The newly discovered parkinson's disease associated finnish mutation (a53e) attenuates alpha-synuclein aggregation and membrane binding. *Biochem.* **53**, 6419-6421 (2014).
267. Sahay, S. et al. Familial parkinson disease-associated mutations alter the site-specific microenvironment and dynamics of alpha-synuclein. *J. Biol. Chem.* **290**, 7804-7822 (2015).
268. Conway, K.A., Harper, J.D. & Lansbury, P.T. Accelerated in vitro fibril formation by a mutant alpha-synuclein linked to early-onset parkinson disease. *Nat. Med.* **4**, 1318-1320 (1998).
269. Burre, J., Sharma, M. & Sudhof, T.C. Systematic mutagenesis of alpha-synuclein reveals distinct sequence requirements for physiological and pathological activities. *J. Neurosci.* **32**, 15227-15242 (2012).
270. Lazaro, D.F. et al. Systematic comparison of the effects of alpha-synuclein mutations on its oligomerization and aggregation. *PLoS Genet.* **10**(2014).
271. Rutherford, N.J., Moore, B.D., Golde, T.E. & Giasson, B.I. Divergent effects of the h5oq and g51d snca mutations on the aggregation of alpha-synuclein. *J. Neurochem.* **131**, 859-867 (2014).
272. van Raaij, M.E., Segers-Nolten, I.M.J. & Subramaniam, V. Quantitative morphological analysis reveals ultrastructural diversity of amyloid fibrils from alpha-synuclein mutants. *Biophys. J.* **91**, L96-L98 (2006).
273. George, J.M. The synucleins. *Genome Biol.* **3**(2002).
274. Biere, A.L. et al. Parkinson's disease-associated alpha-sylnuclein is more fibrillogenic than beta- and gamma-synuclein and cannot cross-seed its homologs. *J. Biol. Chem.* **275**, 34574-34579 (2000).
275. Uversky, V.N. et al. Biophysical properties of the synucleins and their propensities to fibrillate - inhibition of alpha-synuclein assembly by beta- and gamma-synucleins. *J. Biol. Chem.* **277**, 11970-11978 (2002).
276. Eckermann, K., Kugler, S. & Bahr, M. Dimerization propensities of synucleins are not predictive for synuclein aggregation. *Biochim. Biophys. Acta Mol. Basis Dis.* **1852**, 1658-1664 (2015).
277. Hashimoto, M., Rockenstein, E., Mante, M., Mallory, M. & Masliah, E. Beta-synuclein inhibits alpha-synuclein aggregation: A possible role as an anti-parkinsonian factor. *Neuron* **32**, 213-223 (2001).
278. Hashimoto, M. et al. An antiaggregation gene therapy strategy for lewy body disease utilizing beta-synuclein lentivirus in a transgenic model. *Gene Ther.* **11**, 1713-1723 (2004).
279. Park, J.Y. & Lansbury, P.T. Beta-synuclein inhibits formation of alpha-synuclein protofibrils: A possible therapeutic strategy against parkinson's disease. *Biochem.* **42**, 3696-3700 (2003).
280. Janowska, M.K., Wu, K.P. & Baum, J. Unveiling transient protein-protein interactions that modulate inhibition of alpha-synuclein aggregation by beta-synuclein, a pre-synaptic protein that co-localizes with alpha-synuclein. *Sci. Rep.* **5**(2015).
281. Israeli, E. & Sharon, R. Beta-synuclein occurs in vivo in lipid-associated oligomers and forms hetero-oligomers with alpha-synuclein. *J. Neurochem.* **108**, 465-474 (2009).
282. Lee, C.K., Wang, Y.M., Huang, L.S. & Lin, S.M. Atomic force microscopy: Determination of unbinding force, off rate and energy barrier for protein-ligand interaction. *Micron* **38**, 446-461 (2007).

283. Hoffmann, T. & Dougan, L. Single molecule force spectroscopy using polyproteins. *Chem. Soc. Rev.* **41**, 4781-4796 (2012).
284. Hughes, M.L. & Dougan, L. The physics of pulling polyproteins: A review of single molecule force spectroscopy using the afm to study protein unfolding. *Rep. Prog. Phys.* **79**(2016).
285. Rief, M., Gautel, M., Oesterhelt, F., Fernandez, J.M. & Gaub, H.E. Reversible unfolding of individual titin immunoglobulin domains by afm. *Science* **276**, 1109-1112 (1997).
286. Chen, Y., Radford, S.E. & Brockwell, D.J. Force-induced remodelling of proteins and their complexes. *Curr. Opin. Struct. Biol.* **30**, 89-99 (2015).
287. Crampton, N. & Brockwell, D.J. Unravelling the design principles for single protein mechanical strength. *Curr. Opin. Struct. Biol.* **20**, 508-517 (2010).
288. Hugel, T. & Seitz, M. The study of molecular interactions by afm force spectroscopy. *Macromol. Rapid Commun.* **22**, 989-1016 (2001).
289. Lee, G.U., Kidwell, D.A. & Colton, R.J. Sensing discrete streptavidin biotin interactions with atomic-force microscopy. *Langmuir* **10**, 354-357 (1994).
290. Dammer, U. et al. Binding strength between cell-adhesion proteoglycans measured by atomic-force microscopy. *Science* **267**, 1173-1175 (1995).
291. Churnside, A.B. & Perkins, T.T. Ultrastable atomic force microscopy: Improved force and positional stability. *FEBS Lett.* **588**, 3621-3630 (2014).
292. Bull, M.S., Sullan, R.M.A., Li, H.B. & Perkins, T.T. Improved single molecule force spectroscopy using micromachined cantilevers. *Acs Nano* **8**, 4984-4995 (2014).
293. Stuart, J.K. & Hlady, V. Effects of discrete protein surface interactions in scanning force microscopy adhesion force measurements. *Langmuir* **11**, 1368-1374 (1995).
294. Hinterdorfer, P., Baumgartner, W., Gruber, H.J., Schilcher, K. & Schindler, H. Detection and localization of individual antibody-antigen recognition events by atomic force microscopy. *Proc. Natl. Acad. Sci. USA* **93**, 3477-3481 (1996).
295. Riener, C.K. et al. Simple test system for single molecule recognition force microscopy. *Anal. Chim. Acta* **479**, 59-75 (2003).
296. Lo, Y.S. et al. Specific interactions between biotin and avidin studied by atomic force microscopy using the poisson statistical analysis method. *Langmuir* **15**, 1373-1382 (1999).
297. Nakajima, H. et al. Scanning force microscopy of the interaction events between a single molecule of heavy meromyosin and actin. *Biochem. Biophys. Res. Commun.* **234**, 178-182 (1997).
298. Kienberger, F. et al. Recognition force spectroscopy studies of the nta-his6 bond. *Single Mol.* **1**, 59-65 (2000).
299. Willemsen, O.H. et al. Simultaneous height and adhesion imaging of antibody-antigen interactions by atomic force microscopy. *Biophys. J.* **75**, 2220-2228 (1998).
300. Cai, X.E. & Yang, J. The binding potential between the cholera toxin b-oligomer and its receptor. *Biochem.* **42**, 4028-4034 (2003).
301. Marko, J.F. & Siggia, E.D. Stretching DNA. *Macromolecules* **28**, 8759-8770 (1995).
302. Bustamante, C., Marko, J.F., Siggia, E.D. & Smith, S. Entropic elasticity of lambda-phage DNA. *Science* **265**, 1599-1600 (1994).
303. Carrion-Vazquez, M. et al. Mechanical and chemical unfolding of a single protein: A comparison. *Proc. Natl. Acad. Sci. USA* **96**, 3694-3699 (1999).
304. Marszalek, P.E. et al. Mechanical unfolding intermediates in titin modules. *Nature* **402**, 100-103 (1999).
305. Farrance, O.E. et al. A force-activated trip switch triggers rapid dissociation of a colicin from its immunity protein. *PLoS Biol.* **11**(2013).

306. Farrance, O.E., Paci, E., Radford, S.E. & Brockwell, D.J. Extraction of accurate biomolecular parameters from single-molecule force spectroscopy experiments. *Acs Nano* **9**, 1315-1324 (2015).
307. Hickman, S.J., Cooper, R.E.M., Bellucci, L., Paci, E. & Brockwell, D.J. Gating of tonb-dependent transporters by substrate-specific forced remodelling. *Nat. Commun.* **8**(2017).
308. Evans, E. Probing the relation between force - lifetime - and chemistry in single molecular bonds. *Annu. Rev. Biophys. Biomol. Struct.* **30**, 105-128 (2001).
309. Evans, E. & Ritchie, K. Dynamic strength of molecular adhesion bonds. *Biophys. J.* **72**, 1541-1555 (1997).
310. Bell, G.I. Models for the specific adhesion of cells to cells. *Science* **200**, 618-27 (1978).
311. Hervas, R. et al. Common features at the start of the neurodegeneration cascade. *PLoS Biol.* **10**(2012).
312. Sandal, M. et al. Conformational equilibria in monomeric alpha-synuclein at the single-molecule level. *PLoS Biol.* **6**, 99-108 (2008).
313. Brucale, M. et al. Pathogenic mutations shift the equilibria of alpha-synuclein single molecules towards structured conformers. *ChemBioChem* **10**, 176-183 (2009).
314. Dougan, L., Li, J., Badilla, C.L., Berne, B.J. & Fernandez, J.M. Single homopolypeptide chains collapse into mechanically rigid conformations. *Proc. Nat. Acad. Sci.* **106**, 12605-12610 (2009).
315. McAllister, C. et al. Protein interactions and misfolding analyzed by afm force spectroscopy. *J. Mol. Biol.* **354**, 1028-1042 (2005).
316. Yu, J.P., Malkova, S. & Lyubchenko, Y.L. Alpha-synuclein misfolding: Single molecule afm force spectroscopy study. *J. Mol. Biol.* **384**, 992-1001 (2008).
317. Yu, J.P. & Lyubchenko, Y.L. Early stages for parkinson's development: Alpha-synuclein misfolding and aggregation. *J. Neuroimmune Pharm.* **4**, 10-16 (2009).
318. Yu, J.P., Warnke, J. & Lyubchenko, Y.L. Nanoprobing of alpha-synuclein misfolding and aggregation with atomic force microscopy. *Nanomed. Nanotech. Biol. Med.* **7**, 146-152 (2011).
319. Krasnoslobodtsev, A.V. et al. Effect of spermidine on misfolding and interactions of alpha-synuclein. *Plos One* **7**(2012).
320. Krasnoslobodtsev, A.V. et al. Alpha-synuclein misfolding assessed with single molecule afm force spectroscopy: Effect of pathogenic mutations. *Biochem.* **52**, 7377-7386 (2013).
321. Kransnoslobodtsev, A.V. et al. Nanomedicine and protein misfolding diseases. *Nanomed. Nanotech. Biol. Med.* **1**, 300-305 (2005).
322. Kim, B.H. et al. Single-molecule atomic force microscopy force spectroscopy study of a beta-40 interactions. *Biochem.* **50**, 5154-5162 (2011).
323. Portillo, A.M., Krasnoslobodtsev, A.V. & Lyubchenko, Y.L. Effect of electrostatics on aggregation of prion protein sup35 peptide. *J. Phys. Condens. Matter* **24**(2012).
324. Wurth, C., Guimard, N.K. & Hecht, M.H. Mutations that reduce aggregation of the alzheimer's a beta 42 peptide: An unbiased search for the sequence determinants of a beta amyloidogenesis. *J. Mol. Biol.* **319**, 1279-1290 (2002).
325. Aubin-Tam, M.E., Olivares, A.O., Sauer, R.T., Baker, T.A. & Lang, M.J. Single-molecule protein unfolding and translocation by an atp-fueled proteolytic machine. *Cell* **145**, 257-267 (2011).
326. Maillard, R.A. et al. Clpx(p) generates mechanical force to unfold and translocate its protein substrates. *Cell* **145**, 459-469 (2011).
327. Lv, Z.J., Roychaudhuri, R., Condron, M.M., Teplow, D.B. & Lyubchenko, Y.L. Mechanism of amyloid beta-protein dimerization determined using single-molecule afm force spectroscopy. *Sci. Rep.* **3**(2013).

328. Roychaudhuri, R. et al. C-terminal turn stability determines assembly differences between a beta 40 and a beta 42. *J. Mol. Biol.* **425**, 292-308 (2013).
329. Tsigelny, I.F. et al. Dynamics of alpha-synuclein aggregation and inhibition of pore-like oligomer development by beta-synuclein. *FEBS J.* **274**, 1862-1877 (2007).
330. Hoyer, W. et al. Dependence of alpha-synuclein aggregate morphology on solution conditions. *J. Mol. Biol.* **322**, 383-393 (2002).
331. Jackson, M.P. & Hewitt, E.W. Cellular proteostasis: Degradation of misfolded proteins by lysosomes. in *Proteostasis*, Vol. 60 (ed. VanOostenHawle, P.) 173-180 (Portland Press Ltd, London, 2016).
332. Cuervo, A.M. Chaperone-mediated autophagy: Selectivity pays off. *Trends Endocrinol. Metab.* **21**, 142-150 (2010).
333. Tria, G., Mertens, H.D.T., Kachala, M. & Svergun, D.I. Advanced ensemble modelling of flexible macromolecules using x-ray solution scattering. *Iucrj* **2**, 207-217 (2015).
334. Hutter, J.L. & Bechhoefer, J. Calibration of atomic-force microscope tips. *Rev. Sci. Instrum.* **64**, 1868-1873 (1993).
335. deHoffmann, E. Tandem mass spectrometry: A primer. *J. Mass Spectrom.* **31**, 129-137 (1996).
336. Pringle, S.D. et al. An investigation of the mobility separation of some peptide and protein ions using a new hybrid quadrupole/travelling wave ims/oa-tof instrument. *Int. J. Mass Spec.* **261**, 1-12 (2007).
337. Dubay, K.F. et al. Prediction of the absolute aggregation rates of amyloidogenic polypeptide chains. *J. Mol. Biol.* **341**, 1317-1326 (2004).
338. Pawar, A.P. et al. Prediction of "aggregation-prone" and "aggregation-susceptible" regions in proteins associated with neurodegenerative diseases. *J. Mol. Biol.* **350**, 379-392 (2005).
339. Tartaglia, G.G. et al. Prediction of aggregation-prone regions in structured proteins. *J. Mol. Biol.* **380**, 425-436 (2008).
340. Brockwell, D.J. et al. Mechanically unfolding the small, topologically simple protein l. *Biophys. J.* **89**, 506-519 (2005).
341. Sadler, D.P. et al. Identification of a mechanical rheostat in the hydrophobic core of protein l. *J. Mol. Biol.* **393**, 237-248 (2009).
342. West, D.K., Brockwell, D.J., Olmsted, P.D., Radford, S.E. & Paci, E. Mechanical resistance of proteins explained using simple molecular models. *Biophys. J.* **90**, 287-297 (2006).
343. Curtain, C.C. et al. Alpha-synuclein oligomers and fibrils originate in two distinct conformer pools: A small angle x-ray scattering and ensemble optimisation modelling study. *Mol. Biosyst.* **11**, 190-196 (2015).
344. Karamanos, T.K., Kalverda, A.P., Thompson, G.S. & Radford, S.E. Visualization of transient protein-protein interactions that promote or inhibit amyloid assembly. *Mol. Cell* **55**, 214-226 (2014).
345. Groenning, M. Binding mode of thioflavin t and other molecular probes in the context of amyloid fibrils-current status. *Journal of chemical biology* **3**, 1-18 (2010).
346. Lv, Z.J. et al. Direct detection of alpha-synuclein dimerization dynamics: Single-molecule fluorescence analysis. *Biophys. J.* **108**, 2038-2047 (2015).
347. Tiede, C. et al. Adhiron: A stable and versatile peptide display scaffold for molecular recognition applications. *Protein Eng. Des. Sel.* **27**, 145-155 (2014).
348. Tiede, C. et al. Affimers proteins are versatile and renewable affinity reagents. *eLife* **6**(2017).
349. O'Neill, J.W., Kim, D.E., Baker, D. & Zhang, K.Y.J. Structures of the b1 domain of protein l from peptostreptococcus magnus with a tyrosine to tryptophan substitution. *Acta Cryst. D* **57**, 480-487 (2001).

350. Young, L.M., Cao, P., Raleigh, D.P., Ashcroft, A.E. & Radford, S.E. Ion mobility spectrometry-mass spectrometry defines the oligomeric intermediates in amylin amyloid formation and the mode of action of inhibitors. *J. Am. Chem. Soc.* **136**, 660-670 (2014).
351. Young, L.M. et al. Insights into the consequences of co-polymerisation in the early stages of iapp and a beta peptide assembly from mass spectrometry. *Analyst* **14**, 6990-6999 (2015).
352. Young, L.M. et al. Esi-ims-ms: A method for rapid analysis of protein aggregation and its inhibition by small molecules. *Methods* **95**, 62-69 (2016).
353. Woods, L.A. et al. Ligand binding to distinct states diverts aggregation of an amyloid-forming protein. *Nat. Chem. Biol.* **7**, 730-739 (2011).
354. Smith, D.P., Giles, K., Bateman, R.H., Radford, S.E. & Ashcroft, A.E. Monitoring copopulated conformational states during protein folding events using electrospray ionization-ion mobility spectrometry-mass spectrometry. *J. Am. Soc. Mass Spectrom.* **18**, 2180-2190 (2007).
355. Hirel, P.H., Schmitter, J.M., Dessen, P., Fayat, G. & Blanquet, S. Extent of n-terminal methionine excision from escherichia-coli proteins is governed by the side-chain length of the penultimate amino-acid. *Proc. Natl. Acad. Sci. USA* **86**, 8247-8251 (1989).
356. Kelly, S.M., Jess, T.J. & Price, N.C. How to study proteins by circular dichroism. *BBA-Proteins Proteomics* **1751**, 119-139 (2005).
357. Erba, E.B. & Petosa, C. The emerging role of native mass spectrometry in characterizing the structure and dynamics of macromolecular complexes. *Protein Sci.* **24**, 1176-1192 (2015).
358. Frimpong, A.K., Abzatimov, R.R., Uversky, V.N. & Kaltashov, I.A. Characterization of intrinsically disordered proteins with electrospray ionization mass spectrometry: Conformational heterogeneity of alpha-synuclein. *Proteins-Structure Function and Bioinformatics* **78**, 714-722 (2010).
359. Bernstein, S.L. et al. Alpha-synuclein: Stable compact and extended monomeric structures and ph dependence of dimer formation. *J. Am. Soc. Mass Spectrom.* **15**, 1435-1443 (2004).
360. Ullman, O., Fisher, C.K. & Stultz, C.M. Explaining the structural plasticity of alpha-synuclein. *J. Am. Chem. Soc.* **133**, 19536-19546 (2011).
361. Groenning, M. Binding mode of thioflavin t and other molecular probes in the context of amyloid fibrils—current status. *J. Chem. Biol.* **3**, 1-18 (2010).
362. Hong, D.P., Han, S.B., Fink, A.L. & Uversky, V.N. Characterization of the non-fibrillar alpha-synuclein oligomers. *Protein Pep Lett* **18**, 230-240 (2011).
363. Buell, A.K. et al. Solution conditions determine the relative importance of nucleation and growth processes in α-synuclein aggregation. *Proc. Natl. Acad. Sci. USA* **111**, 7671-7676 (2014).
364. Bujalowski, P.J. & Oberhauser, A.F. Tracking unfolding and refolding reactions of single proteins using atomic force microscopy methods. *Methods* **60**, 151-160 (2013).
365. Schapira, A.H.V. Glucocerebrosidase and parkinson disease: Recent advances. *Mol. Cell. Neurosci.* **66**, 37-42 (2015).
366. Zhang, Y.J. & Cremer, P.S. Interactions between macromolecules and ions: The hofmeister series. *Curr. Opin. Chem. Biol.* **10**, 658-663 (2006).
367. Broome, B.M. & Hecht, M.H. Nature disfavors sequences of alternating polar and non-polar amino acids: Implications for amyloidogenesis. *J. Mol. Biol.* **296**, 961-968 (2000).
368. Tartaglia, G.G., Cavalli, A. & Vendruscolo, M. Prediction of local structural stabilities of proteins from their amino acid sequences. *Structure* **15**, 139-143 (2007).
369. Sormanni, P., Aprile, F.A. & Vendruscolo, M. The camsol method of rational design of protein mutants with enhanced solubility. *J. Mol. Biol.* **427**, 478-490 (2015).

370. Sulzer, D. et al. T cells from patients with parkinson's disease recognize alpha-synuclein peptides. *Nature* **546**, 656-+ (2017).
371. Cohen, F.E. & Kelly, J.W. Therapeutic approaches to protein-misfolding diseases. *Nature* **426**, 905-909 (2003).
372. Razavi, H. et al. Benzoxazoles as transthyretin amyloid fibril inhibitors: Synthesis, evaluation, and mechanism of action. *Angew. Chem.-Int. Edit.* **42**, 2758-2761 (2003).
373. Hammarstrom, P., Wiseman, R.L., Powers, E.T. & Kelly, J.W. Prevention of transthyretin amyloid disease by changing protein misfolding energetics. *Science* **299**, 713-716 (2003).