Early molecular events
in the pathology of

Machado-Joseph's disease
or
Spinocerebellar ataxia type 3

Thesis submitted for the degree of PhD.

University of Coimbra
Faculty of Sciences and Technology

by Ricardo Jorge Ladeiro Tomé

Abstract

I - Abstract

From the original description of Machado-Joseph's disease (MJD) or Spinocerebellar ataxia type 3 (SCA3), much knowledge and understanding of the pathology of the disease has been accumulated. The affected gene in this disease is transcribed into multiple splice variants, recently found to be heterogeneously distributed across human cell types. Interestingly, the expression pattern of these transcripts has been shown to favour at least one of these variants in brain cells, coding specifically for a full-length protein with three ubiquitin interacting motifs (UIMs), a catalytic Josephin domain and a polyglutamine stretch (polyQ), the latter expanded in disease states. The *in vivo* function of the affected protein, ataxin-3, is not yet fully understood, though recent work has provided evidence for a role in shuttling ER proteins to the proteasome for degradation. What is clear, and has become the histological hallmark for the disease, is that the affected protein in the pathology, with its expanded polyQ stretch, shows a strong tendency towards self-association, typically accumulating in the nucleus of cells directly involved in pathological cell death. The biochemical and biophysical parameters of this self-association are a topic of growing interest, not just in the study of MJD/SCA3 pathology, but also in the study of other disease-related proteins found to share this genetic alteration, a polyQ expansion in the protein sequence. These diseases are generally referred to as "polyglutaminopathies". Each of the proteins involved in these diseases is linked to protein self-association and leads to neurological disorders, typically of late onset. Somewhat surprisingly, there appears to be some cross-over of effects, as co-aggregation of pathological and non-pathological polyQ-containing proteins has been frequently observed, suggesting a common link. Nevertheless, the patterns of cell death (degeneration) in each polyglutaminopathy are specific, even if somewhat overlapping, indicating a direct link between homeostasis and the normal function of the affected protein, justifying the study of each protein in both pathological and non-pathological states.

Regardless of the mechanism through which cell death occurs in these pathologies, the characteristic protein aggregation is ruled by biochemical and biophysical parameters and is at the start of disease progression. A hope for a future therapeutic approach lies in the determination of these parameters and the design of molecules to prevent or reduce this self-association. Work by several groups has provided insights into the aggregation parameters of ataxin-3, particularly a full-length isoform with two UIM, recently found to be less stable in cells. Together with the preferential expression of the three UIM isoform *in vivo*, it becomes clear that the research focus on ataxin-3 aggregation needs to be re-evaluated, as an extensive characterisation of this isoform's aggregation is currently lacking in literature. Furthermore, the suggestion that, for ataxin-3 as well as other native proteins and engineered peptides the polyQ tract is not directly involved in the early steps of aggregation, strengthens the importance of determining the interplay between the protein's regions and domains in the assembly process.

The work presented in this thesis aims at characterizing the aggregation details of the 3UIM isoform of ataxin-3, establishing a basis for comparison between the full-length ataxin-3 isoforms mentioned. Furthermore, the relative influence of domains and regions of the protein in self-association, most of which are shared by the different isoforms found, is addressed. Specifically, methods for expression and purification of recombinant full-length and C-terminally truncated ataxin-3 protein are described, and high throughput methods for reproducible aggregation kinetics, under physiologically-relevant conditions, developed. With these experimental tools the mechanics and thermodynamical parameters of ataxin-3 aggregation were determined and the role of each region/domain assessed, allowing a model for the early aggregation events of ataxin-3 to be postulated.

Sumário

I – Sumário

Desde a descrição inicial da doença de Machado-Joseph (MJD) ou Ataxia espinocerebelosa tipo 3 (SCA3) muitos foram os avanços, tanto no acumular de conhecimento como na compreensão dos mecanismos inerentes à doença. O gene afectado no estado patológico é transcrito em múltiplas variantes de ARN mensageiro, diferindo entre si no *splicing* alternativo de exões, recentemente descritas como tendo uma distribuição heterogénea nos vários tipos celulares humanos. O padrão de expressão destes ARN foi recentemente descrito como favorecendo a expressão de pelo menos uma das variantes em células do sistema nervoso central, levando à produção de uma isoforma (considerada completa) de ataxina-3, possuindo três motivos de interacção com ubiquitina (UIMs), um domínio catalítico (Josephin) e uma região rica em poliglutaminas (polyQ), esta expandida no estado patologico. A função *in vivo* da ataxina-3 é ainda pouco conhecida, apesar de dados recentes apontarem para um papel na transferência de proteínas do retículo endoplasmático para degradação pelo proteosoma. O que é claro no entanto, e se tornou chave para a identificação histológica da doença, é a tendência da proteína expandida para se associar em agregados intracelulares, tipicamente nucleares, directamente relacionados com a morte celular patológica. Os parâmetros bioquímicos e biofísicos desta agregação são um tópico de interesse na literatura actual, não só no estudo desta patologia mas também no estudo de outras patologias hereditárias que partilham a alteração genética responsável pela expansão da região polyQ. Estas doenças são geralmente designadas poliglutaminopatias, com sintomas neurológicos relacionados com a auto-associação da proteína afectada. Curiosamente, tanto versões expandidas como não expandidas de proteínas contendo uma região polyQ co-localizam nos agregados intracelulares, sugerindo um ponto em comum no mecanismo molecular de agregação. No entanto, os padrões de neurodegeneração são específicos para cada poliglutaminopatia, sugerindo uma ligação entre a homeostase e o papel

fisiológico da proteína afectada e justificando o estudo de cada proteína individualmente.

Independentemente do mecanismo pelo qual a degeneração celular ocorre nestas patologias, a característica agregação de proteínas é regulada por parâmetros bioquímicos e biofísicos e tem lugar no início da progressão da doença. Uma esperança para uma futura aplicação terapêutica poderá estar na determinação destes parâmetros e no desenho de moléculas para prevenir ou reduzir a auto-associação. O trabalho desenvolvido por vários grupos tem fornecido vislumbres dos parâmetros de agregação da ataxina-3, particularmente de uma isoforma, dita completa, contendo dois UIM e recentemente descrita como menos estável em células. Juntamente com a expressão diferencial da isoforma com três UIM *in vivo*, torna-se claro que o foco de investigação na agregação da ataxina-3 deve ser reavaliado, estando uma caracterização extensa da isoforma até à data ausente da literatura. Para mais, a sugestão de que, para a ataxina-3 bem como para outras proteínas (nativas ou desenhadas), a região de polyQ não se encontra directamente envolvida nos passos iniciais de agregação, reforça a importância da determinação das interacções entre regiões e domínios da proteína em processos de agregação.

O trabalho apresentado nesta tese almeja a caracterização detalhada da agregação da isoforma de ataxina-3 com três UIM, estabelecendo uma base para a comparação entre as isoformas ditas completas mencionadas acima, bem como para variantes truncados no terminal carboxílico, permitindo a avaliação da importância relativa dos domínios e regiões da proteína. Métodos de expressão e purificação de ataxina-3 são descritos, simultaneamente para proteína completa e variantes truncadas, e métodos de alta eficiência desenvolvidos para o estudo da agregação da proteína em condições fisiologicamente relevantes. A utilização destas ferramentas experimentais permitiu a determinação de parâmetros mecânicos e termodinâmicos da agregação de ataxina-3, levando à discussão do papel de cada região ou domínio da proteína no processo, bem como a apresentação de um modelo teórico para os passos iniciais de agregação de ataxina-3 em condições fisiologicamente relevantes.

Acknowledgments

II - Acknowledgments

The work presented in this thesis could not have been carried out without many individuals who have made their mark on me. Some of them are mentioned below, others are not, though none is forgotten.

To the members of the Radford lab, a special thank you for making me feel part of the group. Countless discussions with current and past members greatly improved the science conducted.

To my supervisor Dr. Sandra Macedo-Ribeiro, I thank for giving me the opportunity and means to carry out my work, as well as keeping a critical eye on the science developed.

To Prof. Sheena Radford, I thank for infinite patience and guidance through the last four years, without which I know I wouldn't have gotten this far.

To my brother, niece and nephews, who, unknowingly, have shaped many of my choices.

To my aunt, uncle and cousin, for the unconditional love shown over the years, which I hope they know is mutual.

To my partner for the greater part of this project, Eva, for sharing and creating so many good memories.

Most of all, I thank my parents, for always believing in me and supporting my decisions. To you I owe all that I am, was or ever will be.

Thank you!

Table of contents

TABLE OF CONTENTS

List of abbreviations

IV – List of abbreviations

[NaCl] – concentration of NaCl

[protein] – concentration of protein

2UIM – isoform of ataxin-3 described in Kawaguchi, 1994

3UIM – isoform of ataxin-3 described in Goto, 1997

4Ub – tetra-ubiquitin

5TXB – concentrated stock of aggregation buffer

AA – amorphous aggregates

APP – amyloid precursor protein

APS – ammonium persulphate

AT3[2U/Jos] – truncation variant of AT3[U3/14Q], comprising the Josephin domain and the two tandem UIMs of ataxin-3

AT3[Jos] - truncation variant of AT3[U3/14Q], comprising the Josephin domain

AT3[U3/14Q] – full-length ataxin-3 of the 3UIM isoform and with a 14 residue-long polyQ region

Aβ – Amyloid-β peptide

BME – 2-mercaptoethanol

BSA – bovine serum albumin

C_0 – critical concentration

CD – circular dichroism

CR – Congo-red

C-terminal – carboxy-terminal

DMPC/DMPG – cholesterol lipids

DRPLA – dentatorubral-pallidoluysian atrophy

DUB – deubiquitinating enzyme

E.coli – Escherichia coli

EM – electron microscopy

ESI-MS – electrospray ionization mass spectrometry

FAP – Familial amyloid neuropathy

FPLC – fast protein liquid chromatography

FTIR – Fourier-transformed infrared spectroscopy

HMMO - High molecular mass oligomer

Htt – huntingtin protein

IAPP – Islet amyloid peptide

LB – Luria-Bertani media

LBON – LB media without NaCl

LMMO – Low molecular mass oligomer

LS – long-straight fibrillar structures

ME – mixed cellulose esters

MJD - Machado-Joseph's disease (see SCA3)

NI – nuclear inclusions

NMR – nuclear magnetic resonance

N-terminal – amino-terminal

OD – optical density

p.s.i. – pounds per square inch

PDB – protein database

PES – polyethersulfone

pFib – protofibrils

pFil – protofilament

polyQ – polyglutamine

polyUb – polyubiquitin

QBP1 – polyglutamine binding peptide

RL – rod-like fibrillar structures

RO –purified - Reverse osmosis-purified

SAXS – small-angle X-ray scattering

SBMA – spinal bulbar muscular atrophy

SCA – spinocerebellar ataxia (several numbers)

 SCA3 – spinocerebellar ataxia type 3 (see MJD)

SDS-PAGE – polyacrilamide gel electrophoresis with sodium dodecyl sulphate

t_{50} – midpoint of a reaction

T7 – T7 RNA polymerase

TBP – TATA-box binding protein

TEV – Tobacco-etch virus

ThioT – ThioflavinT

ThT – Thioflavin T

TTR – transthyretin

TXBa – buffer from the dilution of 5TXB

U – unfolded

Ub-AMC – ubiquitin-7-amido-4-methylcoumarin

UCH – ubiquitin C-terminal hydrolase

UIM – Ubiquitin interacting motif

UK – United Kingdom

USP – ubiquitin specific protease

UV – ultra-violet

WL – worm-like fibrillar structures

αSyn – alfa-synuclein protein

$β_2m$ – beta-2-microglobulin

Chapter 1

1 – Introduction

1.1 The amyloid structure

Nearly, but not all, proteins and peptides spontaneously self-associate *in vivo* into three-dimensional structures with varied function. While most folding reactions occur with high efficiency, especially when facilitated by molecular chaperones (Hebert, 2007), protein self-assembly can also lead to cellular dysfunction through loss of function or toxic gain of function. A clear example is the pathology of Familial Amyloid Polyneuropathy (FAP), where under non-pathological states the protein transthyretin (TTR) self-associates into tetramers (Fig. 1.1 - B) able to perform its function as a carrier of thyroxine and retinol. In pathological states, however, the functional tetramers dissociate into monomeric species (Fig. 1.1 – A), whereupon they self-associate into another quaternary structure, one coordinating many more monomer copies (Fig. 1.1 – C) (Damas and Saraiva, 2000; Correia, 2006). The resulting structures are non-functional and insoluble, unlike the tetrameric quaternary structure in non-pathological states, and display a set of characteristics that are common to insoluble aggregates of other disease-related proteins, such as β_2-microglobulin (β_2m), cystatin C and Islet amyloid polypeptide (IAPP) (reviewed in Sipe, 2010).

The term amyloid to describe these aggregates was initially coined by Virchow in 1854 (reviewed in Sipe, 2000), mistaking them for starch deposits (*amylum* in latin). Eventually amyloid deposits were found to be composed of aggregated protein (Friedrich and Kekule, 1859, reviewed in Sipe, 2000), and it took almost 70 years to demonstrate a birefringence for these aggregates when visualized under polarized light (Divry and Florkin, 1927, reviewed in Sipe, 2000), indicating a highly ordered arrangement of the polypeptide chain. A little over a century after Virchow's original observation, electron microscopy studies by Cohen and Calkins demonstrated the fibrillar morphology of these aggregates (reviewed in Sipe, 2000). Following on Astbury's and Pauling's work with naturally-occurring protein fibres (such as silk, keratin and collagen), Geddes and co-workers used X-ray beam diffraction to determine the intrinsic order of protein fibrils (reviewed in

Caspar, 2009). Two sets of X-ray diffraction lines were observed by the authors, one longitudinal and one transverse, forming a "cross" pattern, originating the term "cross-β" to describe these structures (Geddes, 1967).
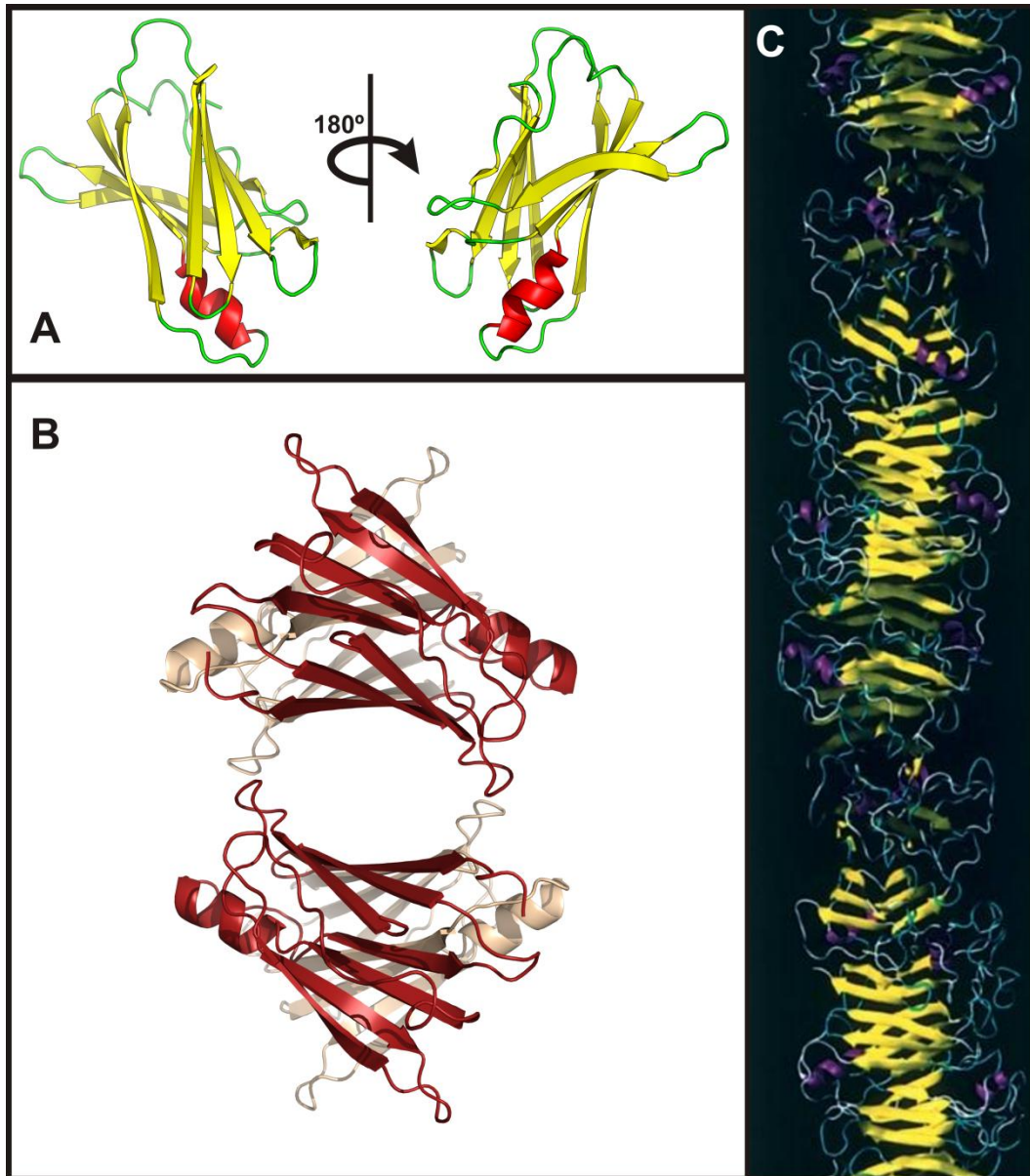
Fig. 1.1 – Cartoon representation of transthyretin protein (TTR) and its more frequent quaternary structures. A) and B) – Monomeric and tetrameric crystallographic structures, respectively (PDB access code for monomeric and oligomeric structures is 1F41 and 2ROX, respectively). C) – Modelled protofibrillar quaternary structure of TTR. Taken from Correia, 2006.

To this day, a definition for an amyloid aggregate is a matter of debate. A broad definition, based on biophysical characteristics, describes amyloid as ordered protein aggregates with fibrillar morphology (as viewed by electron microscopy

and, more recently, atomic force microscopy), that possess an internal structure in which β-strands are arranged perpendicular to the fibril axis, with backbone hydrogen bonds parallel to the fibril axis, the so-called cross-β structure (Fig. 1.1 – C). These aggregates display tinctorial properties upon the binding of planar dyes, leading to altered spectral properties of these molecules. Thioflavin T (ThioT) and Congo-red (CR) are common examples, displaying a shift in the fluorescence (LeVine III, 1999) or absorbance intensity (Klunk, 1999), respectively, or green birefringence under polarized light (in the case of CR), upon interaction with amyloid structures. The specificity of these dyes is questionable (reviewed recently in Reinke, 2011), as the understanding of their binding chemistry is limited, but they are thought to intercalate between beta-sheets (Mao, 2011). Indirect identification of amyloid structure, by judging the content of β conformation, can be achieved using circular dichroism (CD) or Fourier-transform Infrared (FTIR) spectroscopy, although more recently the latter has been shown to distinguish between native β-sheet and amyloid structures (Zandomeneghi, 2004). A direct identification, however, comes from X-ray diffraction studies, such as demonstrated by Geddes and co-workers (Geddes, 1967).

Recent guidelines from the Nomenclature Committee of the International Society of Amyloidosis, describe amyloid as "An insoluble protein fibril that is deposited, mainly, in the extracellular spaces […] that results in a condition known as amyloidosis." (taken from Sipe, 2010). The committee further defines an amyloid fibril as a 10nm wide rigid and non-branching filament, with Congo-red tinctorial properties (as discussed above) and a cross-β pattern upon X-ray diffraction. This set of characteristics is much more restrictive, with the authors suggesting the use of the term "amyloid-like" for fibrils created *in vitro*.

1.1.1 The amyloid fibril

Amyloid fibrils are straight and unbranched linear polymers (Fig. 1.2), consisting of protofilaments, 2-4nm wide (Serpell, 2000), arranged laterally into ribbons, up to 30nm wide (Saiki, 2005), or twisted together like a rope, 7-13nm wide (Serpell, 2000). The average length of fibrils has been shown to vary with aggregation conditions (Ward, 2000; Gosal, 2005; Jahn and Radford, 2006; Lee, 2007). Each protofilament is composed of several copies of the amyloid-forming protein, organized in a cross-β pattern (Fig. 1.3 – C). Typically, amyloid fibrils show X-ray diffraction signals at 4.7 and 10 Å, corresponding to the spacing between β-strands and the stacking distances of β-sheets, respectively (Fig. 1.3 – A, B and C). Within the fibrils, main-chain interactions account for the overall framework, with side-chain packing likely determining the structural variations and different fibril morphologies (Fig. 1.3 – D and Fig. 1.2) (reviewed in Chiti and Dobson, 2006). This array of interactions generally confers upon the fibrils a high stability (Makin, 2005).
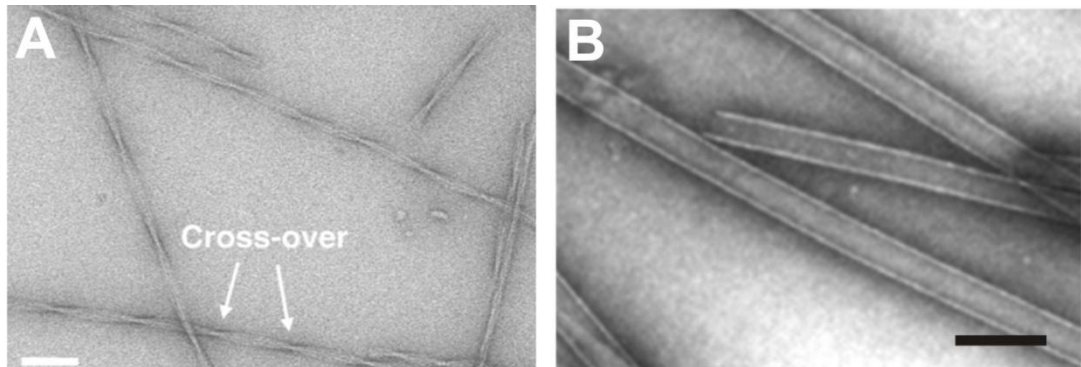

Fig. 1.2 – Electron micrographs of amyloid fibrils. A) – Negative stained twisted fibrils of β$_2$-microglobulin. Taken from White *et al.*, 2009. B) – Negative stained ribbon fibrils of amyloid-beta peptide (Aβ). Taken from Dong *et al.*, 2006. Bars represent 100nm (A) and 200nm (B).
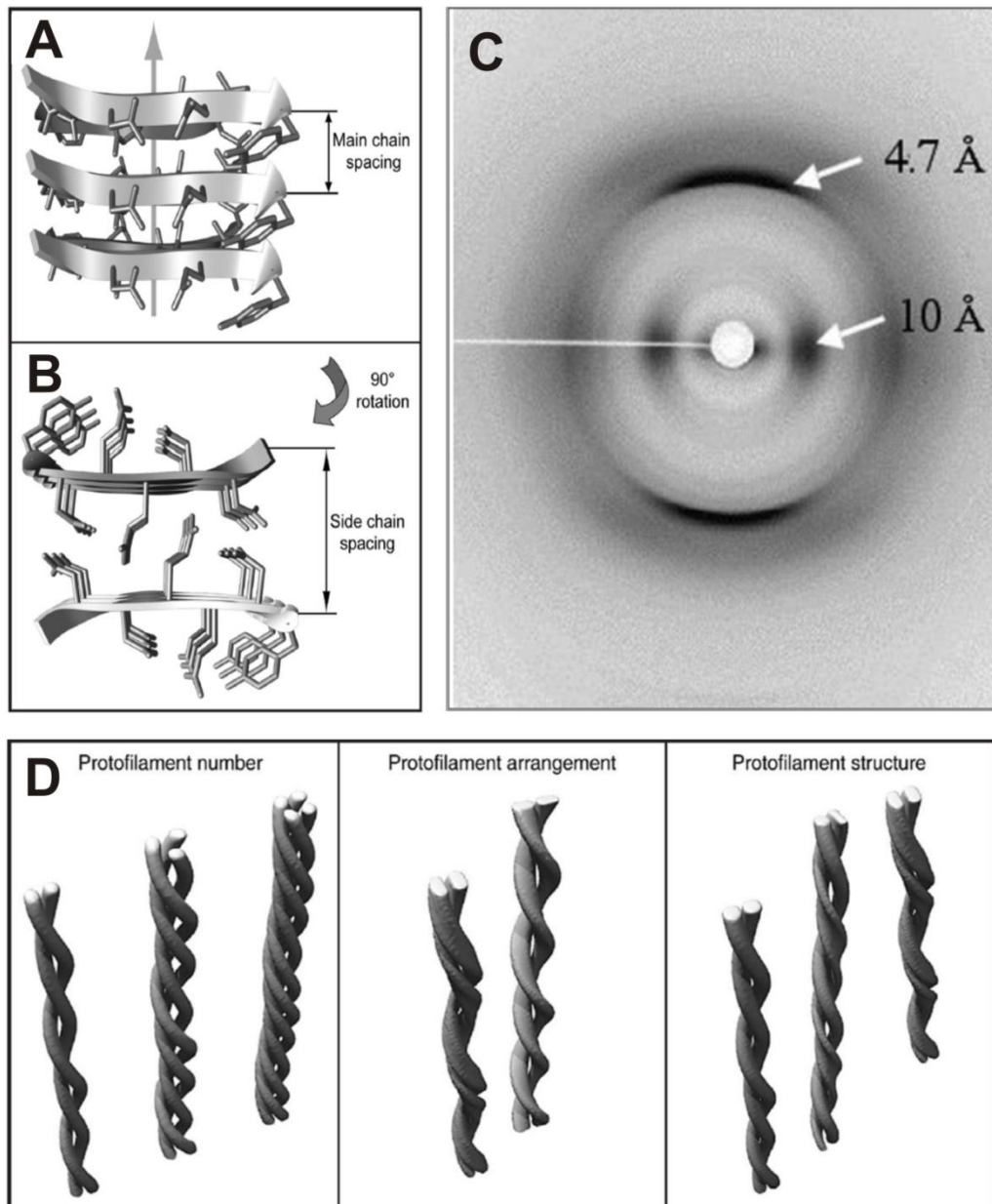
Fig.1.3 – Amyloid fibril architecture and morphology. A) and B) – crystallographic structure of aggregated peptide fragments of Sup35, taken from Fändrich *et al.*, 2009. Central arrow in A represents fibril axis. C) X-ray diffraction pattern of partially aligned Aβ fibrils showing the characteristic "cross" pattern, with reflections at 4.7 and 10Å, corresponding to the spacing marked in A) and B), respectively. Taken from Stromer and Serpell, 2005. D) Cartoon representation of fibril morphologies for rope-like fibrils differing in, from left to right, number, orientation or structure of the protofilaments. Taken from Fändrich, 2009.

1.2 Amyloid assembly

An aggregating protein or peptide can self-associate into several different structures, resulting in different morphologies of the fibrillar end-products. From monomer to fibril, intermediate, non-fibrillar, species are formed, globally referred to as "oligomers". With increasing size, the structures formed typically display increasing β-sheet conformation until the cross-β motif of fibrils is achieved (Gales, 2005; Milner-white, 2006; Tomaselli, 2006; Benseny-Cases, 2007). At this point the intermediate species are referred to as "protofibrils", short curvilinear species of 6-10nm width (Ross and Poirier, 2004), that grow in length by the successive association of aggregation-competent species, until eventually they can be described as protofilaments and associate to form amyloid fibrils (Fig. 1.6). To fully characterize an aggregation reaction requires the knowledge of all species formed, as well as rates involved.

1.2.1 Aggregation models

To describe the assembly process, two models are commonly used, based on the similarities between polypeptide associations and chemical polymerization (Oosawa and Asakura, 1975). In practical terms, a protein assembly reaction is a polymerization reaction where successive addition of assembly-competent species results in the formation of an increasingly long polymer. The two models put forth differ in the description of their aggregation rates, on-rates (describing the addition of aggregation-competent species to a growing polymer) and off-rates (describing the dissociation of building blocks from the polymer), and tend to equilibrium.

1.2.1.1 Nucleation-independent (downhill) aggregation kinetics

The second model describing the assembly of proteins into amyloid fibrils is one where a nucleation step is not required, as all species in the pathway share the same on- and off-rates. According to this model, the reaction progress does not encounter any energy barriers, much like the elongation phase of the previous model, and is frequently referred to as a "downhill" reaction (Fig. 1.4 – B), in the sense that the free energy of species formed is always progressively lower. No critical concentration of protein exists and the reaction progresses at a steadily decreasing rate as it tends to equilibrium and the available monomer concentration decreases (Fig. 1.4 – D).

Importantly, a downhill reaction can display kinetics with a lag phase, not because nucleation is present, but due to secondary processes (such as multiple competing assembly pathways) that slow down the initial polymerization steps (Ferrone, 1999). Such reactions, however, do not display seeding effects (Fig. 1.4 – F) and can therefore be experimentally distinguished from nucleated assembly reactions by the addition of preformed nuclei to the initial steps of the reaction (compare Fig. 1.4 – E and F).
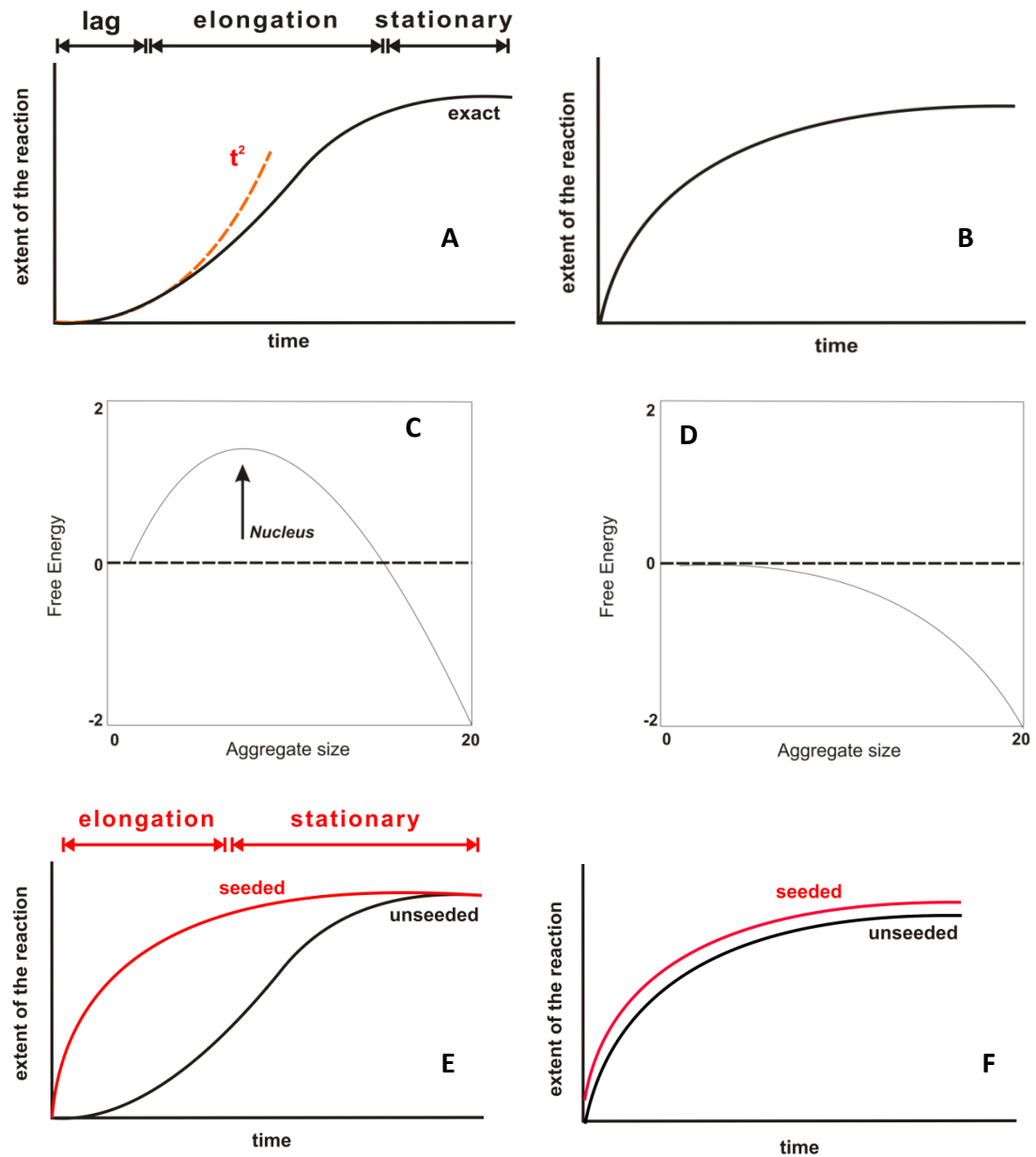
Fig. 1.4 - Schematic representation of important characteristics distinguishing aggregation models. A) and B) Representative aggregation kinetics of nucleated (A) or downhill (B) assembly. Dashed lines in A represent the describing second-order polinomial function, t^2. In A, aggregation phases for nucleated processes are also indicated. C) and D) Representative free energy plots for the assembly, with nucleus formation in C indicated on the plot. E) and F) Representative plots of seeding results for nucleated or downhill reactions. Seeds are added at the start of the reaction. A and C were adapted from Ferrone, 1999. B was adapted from Gosal, 2005 and E was adapted from Harper and Lansbury, 1997.

1.2.1.2 Nucleation-dependent aggregation kinetics

The first of these two models, and the most common in amyloid assembly (Ferrone, 1999), is referred to as nucleation-dependent polymerization. According to the model, the initial steps of aggregation are slower, with a period where the off-rates are higher than the on-rates. During this time, frequently referred to as the "lag phase" (Fig. 1.4 - A), the first species in the pathway to display higher on-rates than off-rates is formed, by overcoming the energy barriers to its formation. This species, the "nucleus" is highly unstable and consequently lowly populated. Its formation, nucleation, marks a turning point in the free energy of the system and bottlenecks the aggregation, with all subsequent steps being highly favourable until equilibrium is reached (Fig. 1.4 - C). The lag phase length is dependent on the initial monomer concentration and the nucleus size (number of monomeric units in the nucleus). As the initial protein concentration decreases, the length of the lag phase increases, possibly due to an entropy loss in the system (as the protein associates into ordered aggregates) that is felt more strongly at lower initial concentrations of monomer. This implies two key points: 1 – at very high initial concentration of monomer, the lag phase should not be visible, as nucleation would happen almost instantaneously; 2 – at very low initial concentration of monomer the lag phase persists for an infinite amount of time, establishing what some authors refer to as a "critical concentration", below which no polymerization occurs (within adequate timeframes). Another important characteristic of nucleated reactions is that the addition of preformed nuclei to the initial aggregating material reduces or abolishes the lag phase length, by bypassing the nucleation steps. This is commonly referred to as "seeding" of a reaction (Fig. 1.4 – E).

After the lag phase, the polymerization reaction in a nucleated process displays an abrupt change in rate, with on-rates far exceeding off-rates and resulting in increasingly longer polymers formed. This is known as the "elongation phase" of the reaction (Fig. 1.4 – A), and has been described to display a second-

order polinomial dependence versus time (Ferrone, 1999 and Fig. 1.4 – A). In some cases the elongation phase is much more pronounced and rapid than the expected dependence, showing an exponential increase instead (Xue, 2008). Inclusion of secondary mechanisms on the nucleation theory allows the explanation of these observations, where events such as fragmentation of fibrillar aggregates (Ruschak, 2007; Xue, 2008, 2009 and 2010), branching (Andersen, 2009; Milhiet, 2010) and heterogenic nucleation (Ferrone, 2002) create more elongation-competent species and lead to more rapid polymerization. Specifically, whereas fragmentation of fibrils leads to more fibrillar structures that can be extended (albeit smaller species), branching allows the elongation to occur outside the growing end of a fibril, with the fibrillar aggregate acting as a scaffold for elongation. On the other hand, a mechanism such as that shown for sickle-cell haemoglobin (Ferrone, 2002) implies the existence of an initial homogeneous nucleation step, with the resulting aggregates further catalysing the aggregation reaction on a "secondary nucleation event". The end of the elongation phase marks an equilibrium plateau between a finite amount of monomer and the polymers formed, as the on-rates match off-rates. Polymers in this final phase, the "stationary phase" (Fig. 1.4 – A) are implicitly not static species but rather dynamic ones, constantly losing building blocks and adding new ones without significantly changing their length (Fig. 1.6).

1.2.1.2.1 Thermodynamical nucleus estimation

An important parameter determining the aggregation rates in nucleated assemblies is the size of the nucleus, frequently referred to as the "thermodynamic nucleus" (since it depends on thermodynamic parameters). Larger nuclei tend to result in lower aggregation rates, as the structure itself is less likely to form. From this premise, two mathematical approaches have been put forth to estimate the size of the thermodynamical nucleus from experimental data: the Oosawa and the Ferrone method (Oosawa and Asakura, 1975 and Ferrone, 1999, respectively). While both approaches assume that the rates of aggregation are unchanged throughout the reaction, the Oosawa method relates the size of the nucleus to the

midpoint of the reaction (the midpoint of the experimental data), whereas the Ferrone method extracts the size of the nucleus from the initial rates of aggregation measured. No mathematical estimation is, by definition, completely correct, and both approaches have their pros and cons. The method developed by Ferrone assumes that at the end of the experimentally-monitored reaction, 100% of the initial monomer has been converted to end-products (polymers). This assumption is crucial to determine the initial rate of aggregation, defined experimentally as the rate at the first 10% of the reaction, and implies a steady-state plateau for the stationary phase. Nevertheless, the more accurate fitting of aggregation rates at a point where nucleation still holds back the progression of polymerization has the potential for greater accuracy in the calculations. On the other hand, the method developed by Oosawa and Asakura can only hold true for "low" protein concentrations since, with an increase in initial monomer concentration, the nucleus becomes progressively more stable, until eventually it reaches a point where it is more stable than the monomer, the "supercritical concentration". However, the estimation of nucleus size does not depend on a complete reaction, safeguarding the equilibrium nature of the stationary phase. Despite their inherently erroneous nature, both approaches have proven useful in the understanding of protein assembly mechanisms. The experimental details of both these approaches are described in Section 2.3.12.4.

1.2.2 Assembly pathways

The association process can be initiated from a natively folded monomer, holding a native-like fold throughout at least most of the assembly reaction (Soldi, 2005; Hamada, 2009 and reviewed in Chiti and Dobson, 2009), or, alternatively, requires unfolding of the monomer (at least partially), thus originating aggregates where each monomer has a conformation typically distinct from the native fold (Booth, 1997, Myers, 2006, Hammarström, 2002 and Rodrigues, 2010 for Lysozyme, β_2m and TTR, respectively). It is also important to mention that for any given system

(the aggregating protein and the aggregating conditions) multiple competing pathways exist (Gosal, 2005; Jahn, 2008; Shtainfeld, 2011), not all originating oligomers productive for fibril formation, leading instead to alternate aggregate structures ("off-pathway" species), including disordered aggregates (Fig. 1.6).
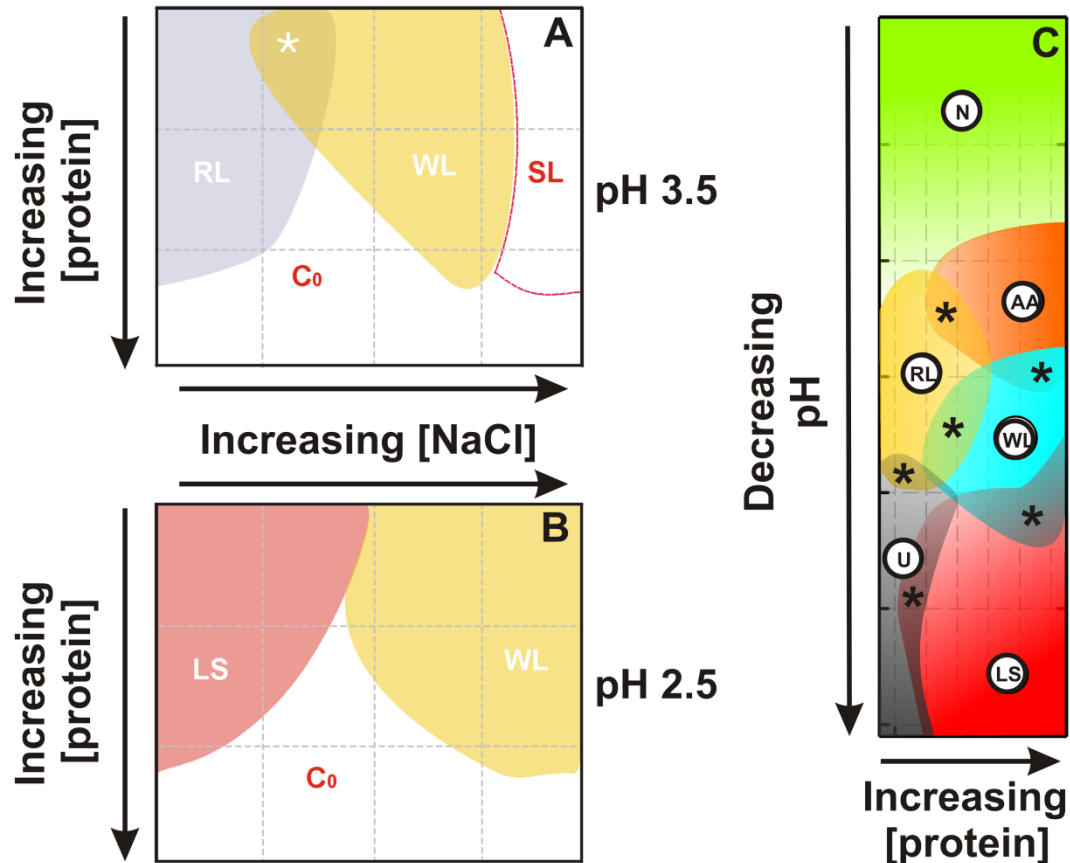

Fig. 1.5 – Schematic representation of the phase changes in β₂m assembly products *in vitro*. Parameters that influence the aggregation were titrated and phase boundaries for end-products of association drawn. A) relationship between NaCl concentration ([NaCl]) and protein concentration ([protein]) at pH 3.5; B) relationship between NaCl concentration ([NaCl]) and protein concentration ([protein]) at pH 2.5; C) relationship between pH and protein concentration ([protein]) at low NaCl concentration. RL, WL and LS refer to the morphology of fibrillar aggregates formed, respectively "rod-like", "worm-like" and "long-straight". AA represents the formation of amorphous aggregates while U denotes unfolded protein and N denotes "native-like" protein. C₀ represents the apparent critical concentration (below which no aggregates are formed) and SL is the solubility limit of the protein. While some overlap between boundaries is seen (labelled with an asterisk on A and C), the final morphology of the sample is conditioned by the environmental conditions. A and B were adapted from Gosal, 2005; C was adapted from Jahn and Radford, 2008.
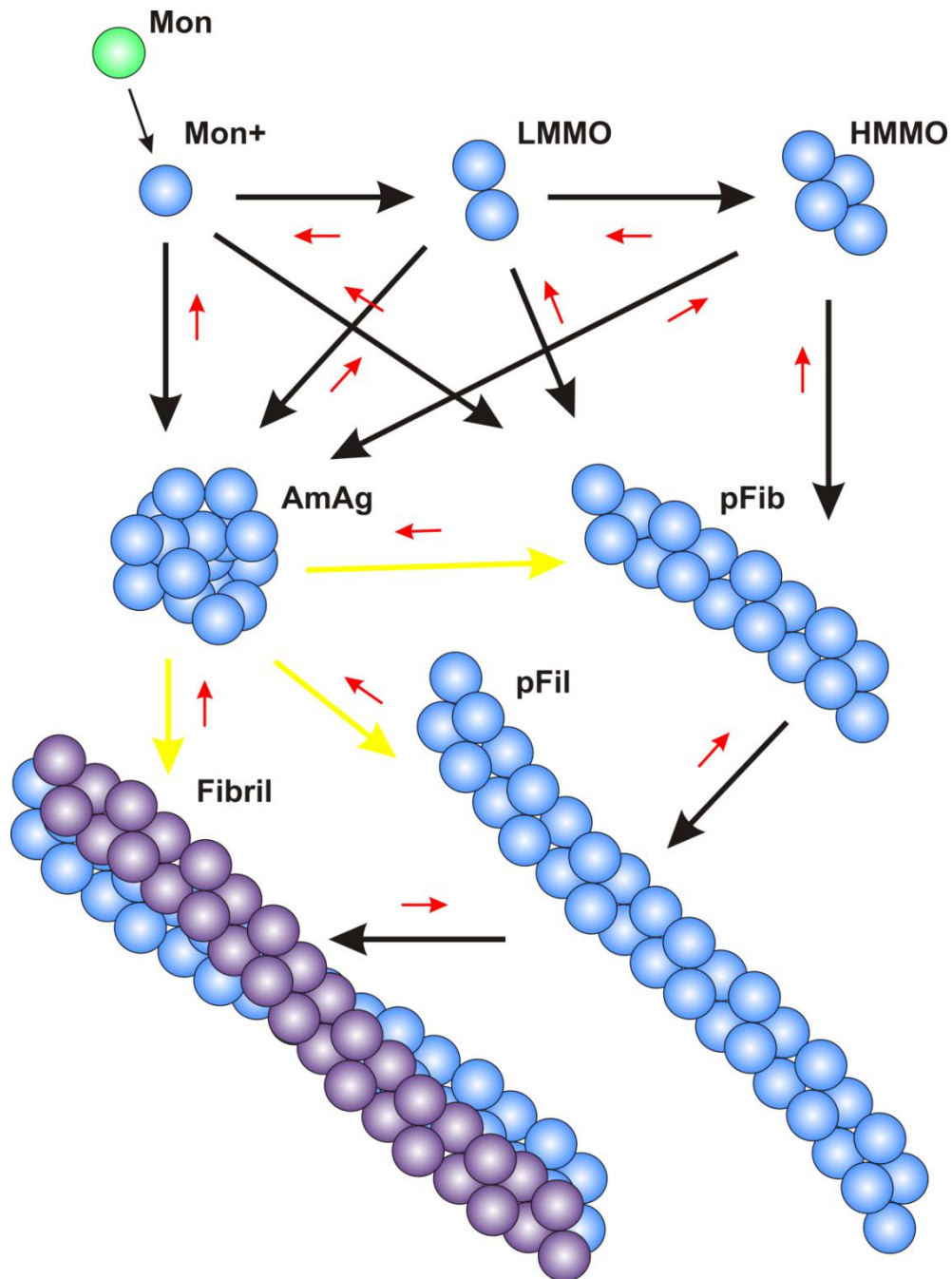
Fig. 1.6 – Schematic representation of the links between species formed during amyloid aggregation. Monomeric protein (Mon) undergoes "activation" to a competent state (Mon+) and further associates into low molecular mass oligomers (LMMO) and high molecular mass oligomers (HMMO). Either one of these species type can aggregate further into amorphous aggregates (AmAg), protofibrillar species (pFib), protofilaments (pFil) or directly onto the fibril. Protofibrillar species can further associate into protofilaments and these into fibrils. Fibrillar products can originate in amorphous aggregates, by conformational rearrangement directly into protofilament states or by dissociation into oligomers (shown in yellow arrows). Back reactions to all those shown by black or yellow arrows, are shown as red smaller arrows.

The balance between competing pathways is determined by intrinsic and extrinsic factors to the polymerizing molecules, and determines the final aggregate morphology (see Section 1.2.3 for an overview of these factors). Gosal and co-workers constructed phase diagrams for $\beta_2 m$ aggregation (Fig. 1.5), determining the interplay of some of these factors and assembly pathways (Gosal, 2005; Jahn and Radford, 2008). These competing pathways frequently have overlapping species, allowing the flux of assembly units between them in a search for stable assemblies (Jahn and Radford, 2006). Even end-products of aggregation are dynamic structures, with aggregation units constantly being released and repolymerized (Gosal, 2005 and Jahn and Radford, 2008; Waudby, 2010). In addition, disordered protein aggregates have been shown to be capable of reorganizing into species that lead to fibril formation ("on-pathway" species), under appropriate conditions (Chiti and Dobson, 2006; Li, 2008), stressing the cross-talk between protein aggregation pathways (Fig. 1.6).

1.2.3 Amyloid aggregation parameters

Despite showing a common architecture, likely dominated by properties of the peptide backbone, the different morphologies for fibrils of different proteins or peptides under similar conditions stress the importance of intrinsic factors in aggregation (Fig. 1.7 – A, C and E). On the other hand, different morphologies for fibrils of the same polypeptide under different conditions underline the importance of extrinsic factors in aggregation (Fig. 1.7 – B, D, F).

1.3.3.1 Intrinsic factors in protein aggregation

Intrinsic factors that play a role in determining the rate of fibril formation and the fibril structure generated include charge and hydrophobicity, patterns of

polar and non-polar residues, and the propensity to adopt certain secondary structure motifs, such as β-rich structures (reviewed by Chiti and Dobson, 2006). In solution, molecules tend to balance or shield charged or hydrophobic patches in order to establish stable solubility. It is no surprise then that charge, both global or in local patterns, and hydrophobicity (the sum of the hydrophobic contributions of each residue), are significant determinants of aggregation rates. Consistently, hydrophobic-hydrophobic patterns were shown to be uncommon occurrences in natural protein sequences (Schwartz, 2001), suggesting an evolutionary effort towards avoiding these aggregation-prone patterns.
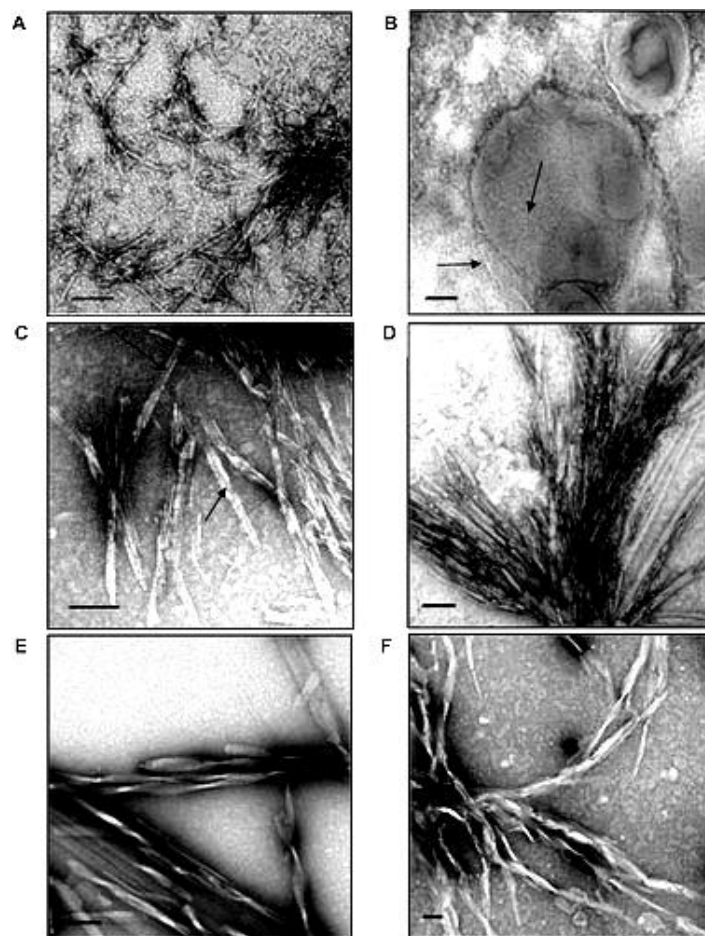
Fig. 1.7 – Electron micrographs of pentapeptides in 10mM Tris base (pH 7.2), 10 mM NaCl, and 0.02% NaN$_3$ with or without DMPC/DMPG/Cholesterol vesicles. A) DFNKF peptide in buffer. B) DFNKF peptide in vesicle solution. C) DANKA peptide in buffer. D) DANKA peptide in vesicle solution. E) DFNMF peptide in buffer. F) DFNMF peptide in vesicle solution. Arrows in B and C point to the fibrillar aggregates due to the low contrast and unusual morphology (respectively). Bars represent 100nm. Taken from Shtainfeld, 2010.

The propensity to aggregate is naturally anticorrelated with the stability of the native state (DuBay, 2004). Aggregation-prone sequences can be found in both α-helical stretches and β-sheets in the native proteins, with a higher probability of intermolecular contacts formed when these regions are solvent-exposed (reviewed in Jahn and Radford, 2008). To readily aggregate into amyloid fibrils *in vitro*, globular proteins are often found to require unfolding (Booth, 1997; Myers, 2006; Rodrigues, 2010), at least partially, possibly to expose aggregation-prone regions that were otherwise buried in the protein's fold (Jahn and Radford, 2008). In contrast, natively unfolded proteins, where nearly all regions are solvent-exposed, typically display lower propensity to form amyloid (Fig. 1.8).
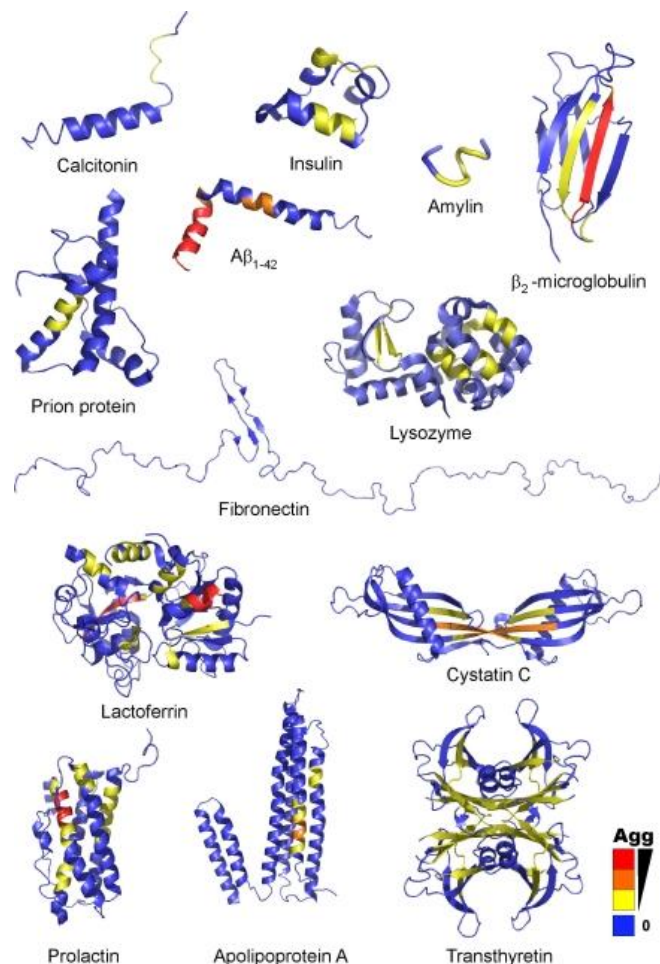

Fig. 1.8 – Cartoon representation of aggregation-prone stretches in disease-related amyloid-forming proteins, as calculated by TANGO (Fernandez-Escamilla, 2004). Warm colours represent high propensity to aggregate (red being the highest), while cool colours represent the opposite (blue for non-aggregation prone regions). Taken from Jahn and Radford, 2008.

1.2.3.2 Extrinsic factors in protein aggregation

Extrinsic factors known to influence the rates of polypeptide aggregation range from physicochemical parameters describing the environment (such as pH, ionic strength, temperature and protein concentration) to interacting reactions, namely with the quality control and protein clearance biomachinery of a cell (chaperones, ubiquitin-related proteins, among others).

1.2.3.2.1 Quality control at the cellular level

A battery of cellular components forms the cell's quality control machinery and ensures correct protein folding or rapid degradation of misfolded peptides (Hebert, 2007; Buchberger, 2010). Molecular chaperones, most of which are induced by stress, decrease the population of aggregation-prone species by enhancing the folding efficiency, and/or shielding aggregation-prone regions from intermolecular interactions (Bukau, 2006; Hebert, 2007; Buchberger, 2010). When a protein fails to pass the quality control machinery, the cell uses two major degradation systems: the ubiquitin-proteasome pathway, where proteins are marked for degradation by the proteasome proteases (Hershko and Ciechanover, 1998, reviewed by Haglung and Dikic, 2005), and autophagy, where vesicles engulf portions of the cytoplasm and deliver them to lysosomes for degradation (reviewed in Rabinowitz, 2010). A broader overview of a cell's quality control mechanism is discussed in Section 1.3.1.1. Both these strategies contribute to a decreased pool of aggregation-competent species, decreasing aggregation. In addition, the ubiquitin system also determines localization within a cell (reviewed in Schwartz and Hochstrasser, 2003 and Ikeda and Dikic, 2008), thus potentially changing the environmental factors to aggregation.

1.2.3.2.2 Environmental factors

The influence of environmental factors, particularly physico-chemical parameters such as pH, temperature, ionic strength and protein concentration, is typically not strictly local, with effects felt in patches at the surface of the protein, as well as globally throughout the protein as a whole. It is clear that pH affects the charge of the protein, by interfering with the protonation state of amino acid side-chains, which in turn affects the protein's solubility. Accordingly, Schmittschmitt and Scholtz showed that for ribonuclease Sa and two charge-reversal variants, fibril formation correlates with solubility at different pH values, but not with protein stability (Schmittschmitt, 2003). On the other hand, the environment's ionic strength interferes with electrostatic forces, such as those involved in protein interactions, as well as altering local charge patterns, working both as positive or negative reinforcers of aggregation, depending on the assembly pathway (Gosal, 2005). In addition, an increase in the environmental temperature results in an increase in the entropy of a system, an effect that needs to be balanced by a loss of enthalpy, possibly by denaturation of the polypeptides, resulting in exposed aggregation-prone regions (Jahn and Radford, 2008) and typically leading to an increased aggregation (Skerget, 2009). Furthermore, an increase in initial protein concentration, provided the pathway is maintained, leads to an increase in aggregation rates, due to an acceleration of nucleation rates, as discussed above (Section 1.2.1.2).

1.2.3.3 Prediction of protein aggregation

Increasing knowledge of the properties linking primary structures of polypeptides and aggregation propensity, have led to the development of several prediction algorithms (reviewed in Belli, 2011). Initially limited to unfolded polypeptides as the starting material (Chiti, 2003), these algorithms evolved to

predict the aggregative behaviour of natively folded polypeptides *in vitro* and, more recently, were shown to accurately predict aggregation parameters *in vivo* (reviewed in Belli, 2011), with even some degree of success in predicting phenotypic effects (Luheshi, 2007). The success of these algorithms in predicting aggregation parameters, whilst not perfect, shows that common principles rule protein aggregation. It is reasonable to assume that, in coming years, the accumulated data on protein aggregation will allow the continuous refinement of these prediction algorithms. A recent step in this direction was the creation of a database (located at www.unifi.it/scbio/bioinfo/AmyloBase.html) for the deposition of protein aggregation data.

1.3 Amyloid in the natural world

The link between amyloid and pathology is a well established one, from the original description of the aggregates by Virchow in 1854 to the present-day knowledge of the molecular aspects of disease. The International Society of Amyloidosis, specifically its Nomenclature Committee, took on the task of not only defining the term "amyloid" (as discussed in Section 1.1), but also list the diseases associated with these protein structures, globally referred to as "amyloidosis". Thirty amyloid-associated human disorders are accepted, according to recent listings (Sipe, 2010), detailed in Table 1.1.

An additional seventeen human disorders are considered amyloid-related (Chiti and Dobson, 2006; Williams, 2008) and are listed in Table 1.2. Within these disorders a group of nine stands out by its common associated cause, an expansion of CAG triplets in specific locations of coding genes. These are highlighted by a green background on Table 1.2 and will be discussed further in Section 1.3.1.2.

Table 1.1 – Listing of human amyloidosis and related proteins. Adapted from Sipe, 2010.

Disease	Aggregating protein or peptide
• AL amyloidosis	Immunoglobulin light chain
• AH amyloidosis	Immunoglobulin heavy chain
• Hemodialysis-related amyloidosis	β_2-microglobulin (β_2m)
• Senile systemic amyloidosis (wild-type) • Familial amyloidotic polyneuropathy (mutant)	Transthyretin (TTR)
• AA amyloidosis • Familial Mediterranean fever	Serum amyloid A protein
• ApoAI amyloidosis	Apolipoprotein AI
• ApoAII amyloidosis	Apolipoprotein AII
• ApoAIV amyloidosis	Apolipoprotein AIV
• Finnish hereditary amyloidosis	Gelsolin
• Lysozyme amyloidosis	Lysozyme
• Fibrinogen amyloidosis	Fibrinogen α-chain
• Icelandic hereditary cerebral amyloid angiopathy	Cystatin C
• Familial British dementia	ABri
• (LECT2)-associated renal amyloidosis	Leucocyte chemotactic factor 2 (LECT2)
• Familial Danish dementia	ADan
• Alzheimer's disease • Hereditary cerebral haemorrhage with amyloidosis Inclusion-body myositis	Amyloid β peptide (Aβ)
• Spongiform encephalopathies	Prion protein (PrP)
• Medullary carcinoma of the thyroid	Calcitonin
• Type II diabetes	Islet amyloid peptide (IAPP), also called Amylin
• Atrial amyloidosis	Atrial Natriuretic factor
• Pituitary prolactinoma	Prolactin
• Injection-localized amyloidosis	Insulin
• Aortic medial amyloidosis	Lactadherin
• Hereditary lattice corneal dystrophy	Kerato-epithelin
• Corneal amylodosis associated with trichiasis	Lactoferrin
• Calcifying epithelial odontogenic tumors	Odontogenic ameloblast-associated protein
• Senile seminal vesicle amyloid	Semenogelin I

Table 1.2 – Human diseases related to amyloid formation. Adapted from multiple sources: Chiti and Dobson, 2006 and Williams and Paulson, 2008. Pathologies resulting from CAG triplet expansion are highlighted with a green background.

Disease	Aggregating protein or peptide
Parkinson's disease Dementia with Lewy bodies	A-Synuclein (αSyn)
Frontotemporal dementia with Parkinsonism	Tau
Amyotrophic lateral sclerosis	Superoxide dismutase 1
Huntington's disease	Huntingtin (Htt)
Spinocerebellar ataxia type 1 (SCA1)	Ataxin-1
Spinocerebellar ataxia type 2 (SCA2)	Ataxin-2
Spinocerebellar ataxia type 3 (SCA3)	Ataxin-3
Spinocerebellar ataxia type 6 (SCA6)	P/Q-type calcium-channel subunit α1A
Spinocerebellar ataxia type 7 (SCA7)	Ataxin-7
Spinocerebellar ataxia type 17 (SCA17)	TATA-box-binding protein (TBP)
Dentatorubral-pallidoluysian atrophy (DRPLA)	Atrophin-1
Spinal and Bulbar Muscular Atrophy (SBMA)	Androgen receptor
Cataract	$_\gamma$-Crystallins
Pulmonary alveolar proteinosis	Lung surfactant protein C
Cutaneous lichen amyloidosis	Keratins

In addition to the 47 amyloid-forming proteins associated with pathological states, a much greater number of proteins has been shown to, under appropriate environmental conditions, adopt an amyloid aggregate conformation (Stefani and Dobson, 2003). The consistent generation of aggregates with structural similarity from such varied sources, suggests that the ability to self-associate into amyloid structures is an inherent property of polypeptides, a universal energy minimum for all peptide chains.

With this hypothesis, the study of amyloid gained increased interest, either in attempting to decrease or abolish the assembly for medical (as discussed above) and biotechnological reasons (for the production of soluble proteins), as well as in increasing, in a controlled manner, the formation of these structures to be used as biomaterials. In addition, accumulated data from multiple sources has revealed an evolutionarily-driven use of amyloid structures as biomaterials, with functional roles as varied as adhesion and protection, together with phenotype determination and physiological control of the organism. These amyloid structures justify an increased interest in amyloid also on the biological field, and are referred to as "functional amyloids", summarized in Table 1.3.

Table 1.3 – Summary of known "functional amyloid" proteins and their function. Adapted from multiple sources: Chiti and Dobson, 2006, Fowler, 2007 and Shewmaker, 2011.

Species	Protein	Amyloid function/characteristics
Bacteria		
Escherichia coli, Salmonella spp.	Curli	Biofilm formation, host invasion
Streptomyces coelicolor	Chaplins	Modulation of water surface tension (i.e. development of aerial structures)
Pseudomonas fluorescens	Fimbriae amyloid-like protein C (FapC)	Biofilm formation
Xanthomonas	Hypersensitive response protein G (HpaG)	Plant pathogen virulence factor
Klebsiella pneumoniae	Microcin E492 (Mcc)	Proposed to regulate Mcc toxicity
Bacillus subtilis	Translocation-dependent antimicrobial spore component A (TasA)	Biofilm component
		(continues)

Species	Protein	Amyloid function/characteristics
		(Table 1.3 - continued)
Protozoa		
Plasmodium falciparum	Merozoite surface protein 2 (MSP2)	Implicated in erythrocyte invasion
Fungi		
Podospora anserine	Small heterokaryon incompatibility protein (HET-s)	Regulation of heterokaryon formation (triggers apoptosis on incompatibility)
Saccharomyces cerevisiae	UREidosuccinate transport protein 2 (URE2p)	Regulation of nitrogen catabolism
S. cerevisiae	Supressor protein 35 (Sup35p)	Regulation of stop-codon read-through
S. cerevisiae	Nucleoporin 1 (Nsp1)	Control of nucleocytoplasmic mixing at nuclear pores
S. cerevisiae	Rich in asparagine and glutamine protein (Rnq1p)	Poorly understood role in [PIN+] phenotype determination
Neurospora crassa and most filamentous fungi	Hydrophobins	Fungal coat formation, modulation of adhesion and surface tension
Candida albicans, Saccharomyces cerevisiae	Fungal adhesins	Cell adhesion
		(continues)

Species	Protein	Amyloid function/characteristics
		(Table 1.3 – continued)
Animalia		
Bombyx mori and Antheraea polyphemus	Chorion proteins	Structural and protective functions in the eggshell
Aplisia californica	Neuron-specific isoform of Cytoplasmic polyadenylation element-binding protein (CPEB)	To promote long-term maintenance of synaptic changes associated with memory storage
Nephila spp.	Spidroins	Structural (i.e. spider silk)
Pseudopleuronectes americanus	Type I antifreeze protein (AFP)	May play role in ice inhibition
Homo sapiens	Premelanosome protein 17 (Pmel17)	Scaffolding and sequestration of toxic intermediates during melanin synthesis
H. sapiens	Peptide hormones	Storage of hormones in an amyloid-like conformation, for a controlled release of monomeric functional hormone peptides

1.3.1 Amyloid toxicity

A common amyloid structure in pathology suggests common origins of toxicity. Working on Alzheimer's disease (Table 1.1), Hardy and Allsop postulated that toxicity arose directly from the deposition of protein into amyloid plaques. These would then foster further aggregation, in a cascade of events leading to cellular disruption, the "amyloid cascade hypothesis" (reviewed in Hardy and Selkoe, 2002). Consistent with this theory is the observation that fibrillar Aβ is toxic

when injected in the cerebellar cortex of Rhesus monkeys and that the vast majority of mutations involved in the Alzheimer pathology increase the rates of fibril formation (reviewed in Tanzi and Bertram, 2005). However, a stronger correlation was found between the amounts of soluble Aβ aggregates and the severity of disease than with insoluble amyloid load (McLean, 1999). Accordingly, transgenic mice were found to develop symptoms before deposition of Aβ fibrils (Arendash, 2004). Together, these and other studies on several pathologies (discussed in Glabe, 2008 and reviewed in Stefani, 2010) forced a change in the views of toxicity, and a new hypothesis was postulated (Kayed, 2003 and 2007), the "toxic oligomer hypothesis", according to which the culprit of cellular toxicity in protein aggregation disorders is a soluble non-fibrillar intermediate, an oligomer. Accordingly, Kayed and co-workers developed an antibody, A11, able to recognize soluble forms of several amyloid-forming protein aggregates related to disease (Kayed, 2003 and 2007), suggesting a common mechanism of cytotoxicity. Furthermore, treating cells over expressing amyloidogenic proteins with the A11 antibody, results in a reduction or abolishment of toxicity (Kayed, 2003). Recent evidence has however blurred the distinction between these theories, with small soluble fibrillar aggregates of β_2m shown to have toxic effects in cells, implying a much more complex mechanism of toxicity. No single theory is likely to explain all origins of toxicity, despite shared characteristics of the aggregates. The patterns of histopathology are specific to the disease in question, varying from systemic to organ-specific diseases (discussed in Sipe, 2010), suggesting a modulation of toxicity by specific protein context.

1.3.1.1 Cellular quality control

Intracellular and extracellular quality control systems against misfolded proteins are frequently implicated in amyloidosis and amyloid-related diseases, particularly molecular chaperones and protein degradation systems (reviewed in Naiki and Nagai, 2009). Molecular chaperones assist in the folding of newly

synthesized proteins and refolding of misfolded or aggregated protein (discussed in Section 1.2.3.2.1 and reviewed in Muchowski and Wacker, 2005), and have been found to ameliorate the aggregation of amyloid proteins in cells (Cummings, 1998) and animal models (Warrick, 1999 and Cummings, 2001), as well as *in vitro* (reviewed in Nerelius, 2010). In addition, it has been proposed that chaperones may capture monomers or oligomers that depolymerise from fibrils (as *per* normal fibril dynamics, Section 1.2.2), effectively reducing the pool of aggregation-competent species and shifting the equilibrium towards depolymerisation (Waudby, 2010). On the other hand, protein degradation systems focus on clearing misfolded or aggregated protein from the cell by targeting them for degradation by the ubiquitin-proteasome or the autophagy-lysosome system. While the ubiquitin-proteasome system is thought of as inherently selective (by the specific tagging of molecules with a ubiquitin signal), the autophagic response was, until recently, thought of as non-selective. Recent evidence has, however, challenged this view and pointed to some selectivity in degradation (Kirkin, 2009 and Cuervo, 2011). Some evidence of a possible role by the autophagy-lysosome system has been accumulated for diseases such as Huntington's and MJD/SCA3 (Ravikumar, 2004; Nascimento-Ferreira, 2011), though the mechanism involved is still poorly known, as recent work on Alzheimer's disease by Boland and co-workers has shown, whereby evidence for autophagy was found, but not correlated to the metabolism of APP (Boland, 2008 and 2010, respectively). Conversely, evidence for the involvement of the ubiquitin-proteasome machinery in protein aggregation diseases is plentiful, particularly for CAG expansion diseases (see Section 1.3.1.2), where molecules of this quality control system co-localize with protein aggregates (reviewed by Naiki and Nagai, 2009).

Ubiquitin itself is a small globular protein that acts as a molecular tag in cells, by covalently binding to other proteins (ubiquitination), either as a single moiety or as a polymer, and determining such details as subcellular localization, protein turnover and enzymatic activity (reviewed in Schwartz and Hochstrasser, 2003 and Ikeda and Dikic, 2008). The binding of ubiquitin to either a protein substrate or to another ubiquitin molecule is a complex process involving several

other proteins: E1 ubiquitin-activating enzymes capture free ubiquitin and transfer it to E2 ubiquitin-conjugating enzymes. The E2-Ub complex then binds E3 ubiquitin ligases, which catalyze the covalent binding of ubiquitin moieties (Fig. 1.9). Plyubiquitin (polyUb) chains of different lengths are formed by an iterative process involving this machinery. Polymerization of ubiquitin can occur through any of seven lysine residues on its sequence (Lys6, Lys11, Lys27, Lys29, Lys33, Lys48 and Lys63), leading to chains of different morphology (Fig. 1.9) and ubiquitin tags of different effect (reviewed by Haglung and Dikic, 2005). The cellular role for all these homopolymers is not fully understood, although it is known that four or more ubiquitin moieties linked by Lys48, for example, are involved in the ubiquitin-proteasome pathways, forming a degradative tag that marks the ubiquitinated protein for degradation by the proteasome (Hershko and Ciechanover, 1998, reviewed by Haglung and Dikic, 2005). Polyubiquitin chains linked by Lys63, on the other hand, are known to be involved in DNA-repair processes, signal transduction and receptor endocytosis (reviewed in Haglung and Dikic, 2005 and Hayden and Ghosh, 2008).

Editing and disassembling polyUb chains and ubiquitin tags is carried out by deubiquitinating enzymes (DUBs), ubiquitin hydrolases essential for cellular homeostasis (Fig. 1.9). By proof-reading and editing ubiquitin signals, these enzymes are important in both positive and negative regulation of the ubiquitin pathways, as well as maintaining the levels of free ubiquitin in cells. Around 100 DUBs are encoded in the human genome, stressing their importance for the cell. Accordingly, regulation of DUB activity is key to cellular homeostasis and is performed both at the transcriptional and post-transcriptional level (reviewed in Nijman, 2005). Besides the co-localization of components of this system with growing amyloid aggregates *in vivo* and in cells, the link between the ubiquitin biomachinery and protein aggregation diseases can be a direct one, as with proteins such as parkin (an E3 ubiquitin ligase)(reviewed by Kahle and Haass, 2004) and ataxin-3 (a DUB) (reviewed by Nijman, 2005), respectively implicated in Alzheimer's disease and MJD/SCA3.
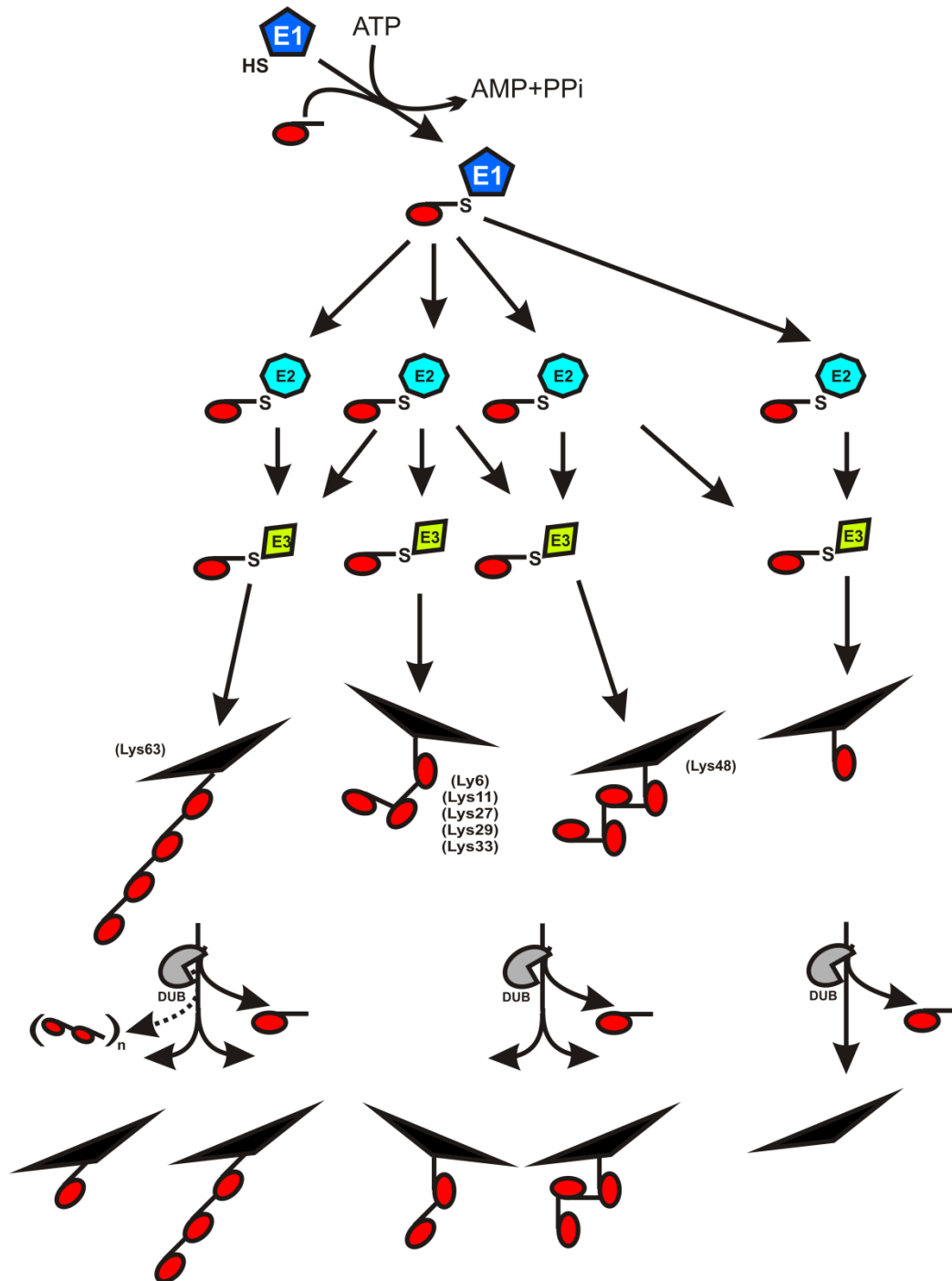
Fig. 1.9 – Schematic representation of some ubiquitin pathways. Ubiquitin conjugation, to either a substrate or itself (forming homopolymers linked by one of the Lysine residues labelled) is represented in the figure, along with the architecture of some of the resulting homopolymers. Ubiquitin molecules are shown in red, E1, E2 and E3 are shown in dark blue, light blue and green, respectively. Substrates are shown in black and deubiquitinating enzymes in grey. Deubiquitination enzymes (DUB) act in editing and disassembly of ubiquitin tags, in close and direct relation with the ubiquitin conjugation and polymerization reactions. Adapted from Wilkinson, 2005 and Dikic, 2009.

1.3.1.2 PolyQ expansion and disease

Roughly two decades ago, expansions of trinucleotide repeats directly correlated with intergenerational instability were described (Fu, 1991, LaSpada, 1991). The largest group of diseases associated with this instability are those in which an expansion of CAG repeats in coding regions of the affected gene are translated as polyglutamine (polyQ) tracts in corresponding proteins (reviewed by Williams and Paulson, 2008). They are referred to as polyglutaminopathies or polyQ disorders. Nine associated diseases have been described: Huntington's disease (HD), dentatorubral-pallidoluysian atrophy (DRPLA), spinocerebellar ataxias type 1, 2, 3 (MJD/SCA3), 6, 7 and 17, and spinobulbar muscular atrophy (SBMA). All except the latter (which is linked to sexual determination) are autossomal dominantly inherited. Table 1.4 summarizes the diseases and their associated proteins. All polyglutaminopathies are of late onset and slowly progressive, with severity of disease and age of onset correlated with the length of the polyQ tract. The proteins associated with polyQ disorders have an unstructured segment that contains the polyQ tract, typically N-terminally positioned, with the exception of the MJD/SCA3 and SCA6 protein products, where the polyQ stretch is at the C-terminal part of the protein (reviewed in Williams and Paulson, 2008). The proteins share no obvious homology apart from the expansion of polyglutamine tracts, and are ubiquitously expressed in neuronal (and some non-neuronal) tissues. Nevertheless, polyQ disorders are characterized by the development of intraneuronal aggregates of the expanded protein, typically nuclear, and co-localizing with cellular quality control biomachinery. Non-pathological polyQ-containing proteins (including non-expanded forms of the pathological protein) co-localize with these aggregates (Kawaguchi, 1994), suggesting a direct role for polyQ stretches in aggregation, strengthened by the potential for an added stability from the additional array of hydrogen-bonds between glutamine side-chains (Esposito, 2008; Berryman, 2009; Natalello, 2011). Models relating repeat-length dependent nucleation and disease

age-of-onset (Sugaya and Matsubara, 2009) are in concert with these observations. Interestingly, while for eight of the polyQ disorders a linear correlation was found between the age-of-onset and the affected protein's nucleation time, at different polyQ-length, MJD/SCA3 does not follow this linear trend (Sugaya and Matsubara, 2009), suggesting a pathology that is not completely dependent on the polyQ-repeat length. Increased expression of chaperones has been shown to mitigate polyQ toxicity, whereas inhibition of proteasomal degradation stimulates cell death, suggesting that the presence of expanded polyQ-containing proteins in aggregates may be an attempt by the cell to clear the aberrant protein. Nevertheless, each protein affects specific, though somewhat overlapping, regions of the brain, suggesting the importance of protein context in determining the pathology (reviewed in Havel, 2009 and Takahashi, 2010).

Table 1.4 – The nine polyQ disorders, their affected proteins and function. Taken from Williams and Paulson, 2008.

Disease	Affected gene	Affected protein	Relative protein size and polyQ position	Pathogenic polyQ length	Protein function
SBMA	AR	Androgen Receptor		40-62	Testosterone-activated steroid receptor
HD	HTT	Huntingtin		36-121	Possible scaffolding protein linked to diverse cellular pathways
DRPLA	ATN1	Atrophin-1		49-88	Possible transcriptional corepressor
SCA1	ATXN1	Ataxin-1		39-91	Transcriptional corepressor involved in transcription regulation, cell specification and synaptic activity
SCA2	ATXN2	Ataxin-2		32-200	Component of RNA processing and translational regulation pathways
SCA3	ATXN3	Ataxin-3		52-86	Deubiquitinating enzyme involved in protein quality control
SCA6	CACNA1A	P/Q-type Ca-channel subunit 1α		20-33	Voltage-sensitive calcium-channel subunit
SCA7	ATXN7	Ataxin-7		34-300	Component of histone acetyltransferase complex (TFTC/STAGA) and transcriptional regulation pathways
SCA17	TBP	TATA-box binding protein		47-63	Component of core transcriptional complex TFIID

1.4 Machado-Joseph's disease or Spinocerebellar ataxia type 3

One of the polyQ disorders, Machado-Joseph's disease (MJD) was originally described by Coutinho and Andrade in 1978 (reviewed in Bettencourt, 2011), named after the two families presenting the pathology that were initially studied. Later, and independently, the disease was also annotated by Harding (Harding, 1993), who named it Spinocerebellar ataxia type 3 (SCA3), based on the ataxia-like phenotype of the pathology. The disorders were later found to be the same pathology by genetic studies (Stevanin, 1994) and both designations persist to this day (referred to in this thesis as MJD/SCA3).

MJD/SCA3 is an autossomal dominantly-inherited disorder, and the most prevalent form of ataxia, accounting for 21-28% of all cerebellar ataxias, with a prevalence of 1-2:100000 (discussed in Albrecht, 2003). The pathology manifests with late onset, typically in the third or fourth decade of life, with characteristics of cerebellar ataxia, parkinsonian signs and peripheral neuropathy (Dürr, 1996; reviewed in Matos, 2011). The affected gene, *ATXN3*, is located on the long arm of chromosome 14 (14q32.1) (reviewed in Stevanin, 1994) and its transcription product is alternatively spliced into 56 known isoforms (Bettencourt, 2010), the translated product of which is ataxin-3. Ataxin-3 is widely expressed, but causes selective neurodegeneration of cerebellar dentate neurons, basal ganglia, brain stem and spinal cord (Dürr, 1996; reviewed in Matos, 2011). The protein is typically cytoplasmic, though the hallmark of the disease is the formation of nuclear inclusions (Paulson, 1997), composed of a core of expanded protein surrounded by a layer of non-expanded protein (Fujigasaki, 2000). These inclusions localize to regions of the brain roughly corresponding to the affected neurodegenerative pattern, and are heavily ubiquitinated (Paulson, 1997), suggesting an attempt by the cell to clear the aggregates.

Interestingly, ataxin-3 has been shown to co-localize with aggregates of proteins in other disease-states (Uchihara, 2001) and even with aggregates related

to the aging process, such as Marinesco bodies (Fujigasaki, 2001; Kettner, 2002). The significance of these observations is currently unclear, though they suggest a strong tendency for ataxin-3 to form aggregates. In fact, the presence of non-expanded protein in intracellular aggregates (Fujigasaki, 2000), suggests that while it is not responsible for the onset of disease, it may contribute to pathogenesis.
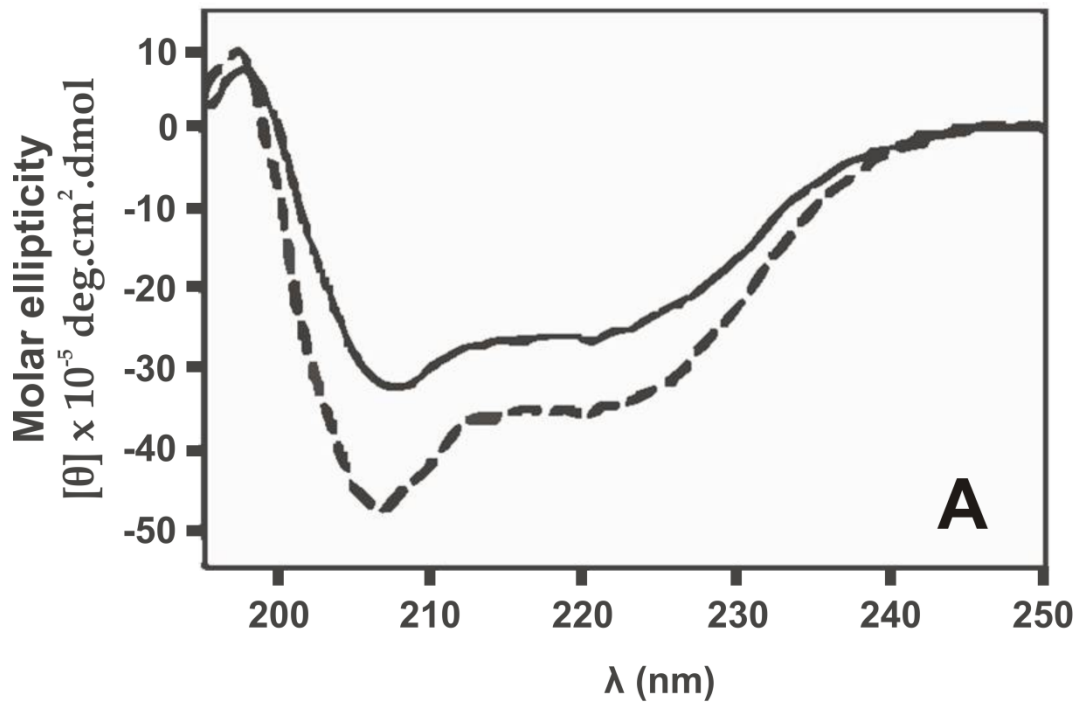
1.4.1 Ataxin-3 architecture

Kawaguchi and co-workers (Kawaguchi, 1994) originally isolated a transcript from human brain, encoding a 360 residue-long isoform of ataxin-3. Goto and colleagues later isolated another transcript for ataxin-3, differing from the previous one in the alternative splicing of the last exon, encoding a protein isoform that is slightly longer, with 361 residues (Goto, 1997). The general architecture of the protein was described by Masino and co-workers to be comprised of a compact and structured N-terminal domain (Josephin domain), partially resistant to proteolytic cleavage, followed by a relatively unstructured and flexible C-terminal region (Masino, 2003), in agreement with a predicted elongated shape based on analytical centrifugation measurements (Bevivino, 2001). The far-UV CD spectrum of ataxin-3 shows a high α-helical content, dominated by the secondary structure of the Josephin domain (Fig. 1.10, taken from Chow, 2004), and implies a lack of significant interaction between the N- and C-terminal portions of the protein (Masino, 2003; Chow, 2004; Gales, 2005).

Two functional ubiquitin interacting motifs (UIMs) where identified, C-terminal to the Josephin domain (Donaldson, 2003; Berke, 2005; Song, 2010), followed by a polyQ region (Fig. 1.10). The alternative splicing of the last exon in the transcript, determines the nature of the C-terminal region to the polyQ stretch. While with the shorter isoform (Kawaguchi, 1994) this region lacks a known structural motif, comprising a hydrophobic-rich amino acid sequence, the longer isoform (Goto, 1997) contains a third UIM in its C-terminal sequence, whose

functionality was also demonstrated (Donaldson, 2003; Berke, 2005). These protein isoforms are thus referred to as the 2UIM and the 3UIM ataxin-3 isoform, for the shorter and longer splice variant, respectively, and are considered to represent full-length ataxin-3 (Fig. 1.10). Both have been the object of study by several authors (Donaldson, 2003; Chow, 2004 and 2006; Gales, 2005; Berke, 2005; Ellisdon, 2006 and 2007), though recent work has indicated the 3UIM ataxin-3 isoform as the major isoform expressed in brain tissue (Harris, 2010), regardless of age or polyQ expansion, and despite both splice variants being detected at the mRNA level. The 3UIM isoform was used for the experiments described in this thesis.

The three-dimensional structure of ataxin-3 has not yet been determined, but work by two groups has, independently, led to a structural determination of the Josephin domain, shown in Fig. 1.11 (Nicastro, 2005 and 2006; Mao, 2005). This domain has a globular papain-like fold, with a high α-helical content, as suggested by far-UV CD measurements (Masino, 2003; Chow, 2004; Gales, 2005). A catalytic triad is formed in the N-terminal Josephin domain, by Cys14, His119 and Asn134, where the cysteine is responsible for the nucleophilic attack, classifying the enzyme as a cys-protease (Fig. 1.11). A detailed look at the catalytic triad on both NMR structures published (Nicastro, 2005; Mao, 2005) shows that while on one structure (Nicastro, 2005) the residues are close enough to imply an active enzyme, on the other (Mao, 2005) the cysteine residue is turned away from the active-site, too far from the histidine residue of the triad, implying an enzyme that requires conformational rearrangement (Fig. 1.11 - bottom). Further work by Nicastro and co-workers with SAXS measurements, allowed evidence consistent with the first structure to be collected (Nicastro, 2006), strongly suggesting an open conformation, as published previously by the author (Nicastro, 2005).
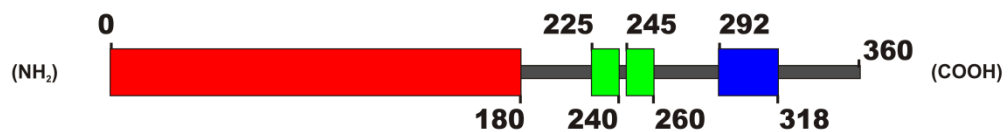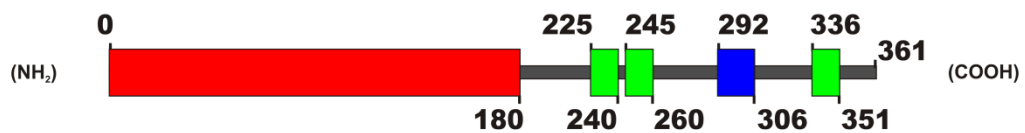
Fig. 1.10 – Architecture of ataxin-3. A) far-UV CD spectra of full-length ataxin-3 (2UIM isoform) and a truncation comprising the Josephin domain alone, respectively in a solid or dashed line. B) Schematic representation of the two full-length isoforms of ataxin-3 described. The 2UIM isoform was originally described by Kawaguchi, 1994. It contains two ubiquitin interacting motifs (in green), a polyQ tract (in blue) and a catalytic domain, Josephin (in red). It differs from the 3UIM isoform (Goto, 1997) in that it does not have a third UIM, C-terminal to the polyQ tract. A was taken from Chow, 2004.
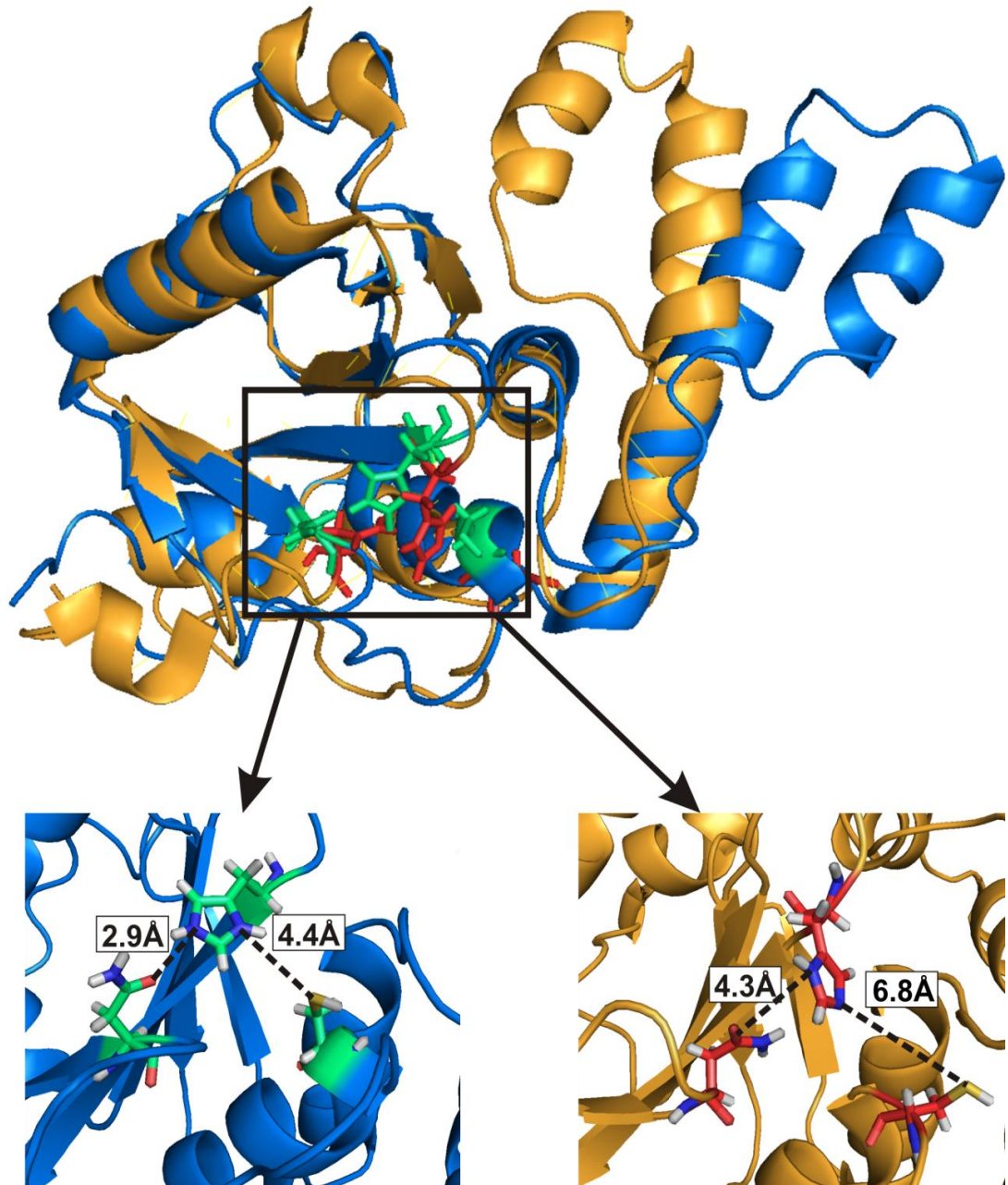
Fig. 1.11 – Cartoon representation of the two structures calculated by solution NMR of the Josephin domain. In blue is the structure published in Nicastro, 2005 (PDB access code: 1YZB) and in orange the structure in Mao, 2005 (PDB access code: 2AGA). The catalytic residues for each structure are shown in green and red, respectively. A magnified view of the catalytic site on both structures is shown below, with the distances between the catalytic residues annotated.

1.4.2 Ataxin-3 activity

Highly conserved residues in the Josephin domain allowed its original putative identification as a cysteine protease with distant homology to deubiquitinating enzymes (DUBs) (Albrecht, 2004). While the Cys-protease character was confirmed by the NMR determination of the domain (Nicastro, 2005; Mao, 2005), ubiquitin hydrolase activity was shown by Burnett and co-workers (Burnett, 2003) against a model DUB substrate, Ubiquitin-AMC (Ub-AMC). Furthermore, the ability to bind homopolymers of ubiquitin (polyUb) was shown, with a clear preference for tetra-ubiquitin (4Ub) polymers or longer.

Following on these findings, other authors published a deubiquitinase activity for ataxin-3 or a truncation variant comprising the Josephin domain (Chow, 2004; Mao, 2005), as well as an interaction with ubiquitin through the UIM regions (Donaldson, 2003; Berke, 2005; Winborn, 2008). Based on the structural determination of Josephin (Nicastro, 2005; Mao, 2005) and data reporting the absence of significant interaction between this domain and the highly flexible C-terminal portion of ataxin-3 (Masino, 2003), Mao and colleagues (Mao, 2005) proposed a model for catalysis where a ubiquitin-binding site on Josephin is arranged in a linear order with the catalytic pocket, the first two UIMs in tandem, the polyQ stretch and the alternatively-spliced C-terminus (Fig. 1.12). Interaction of polyUb chains with this linear arrangement on ataxin-3 would position the substrate with the isopeptide bond of a distal ubiquitin over the catalytical site, allowing the enzyme to trim polyUb chains. Work by Donaldson and co-workers initially showed that the activity of ataxin-3 on polyUb chains was linked to the UIM and, through mutation strategies designed to disrupt UIM interaction with Ub, the relative importance of each UIM was assessed (Donaldson, 2003). Specifically, the first UIM was shown to have the most determining contribution to activity, followed by the second UIM. The third UIM, however, was not shown to have a

marked influence on activity, consistent with recent data indicating that the enzymatic activity of the 2UIM and 3UIM isoforms of ataxin-3 is identical. Furthermore, work by Winborn and co-workers showed that ataxin-3 has a preference not just for tetra-Ub chains or longer, but also for those linked by Lys63 (Winborn, 2008). More importantly, the preference for binding and hydrolase activity was linked to the UIM regions, without which the enzyme's preferential substrate changes to a Lys48-linked dimer (Winborn, 2008).
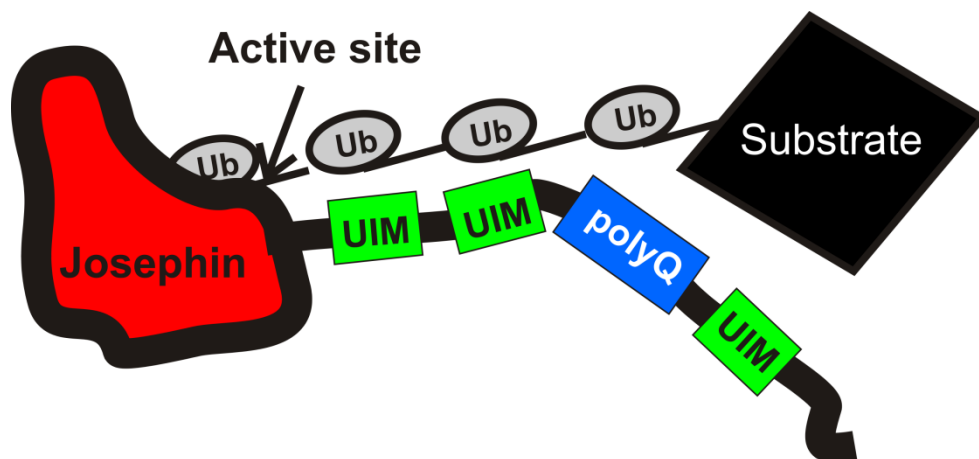
Fig. 1.12 – Schematic representation of the proposed catalytic activity mechanism for ataxin-3. Accumulated data shows the importance of the UIM sequences in determining specificity of cleavage, consistent with the model proposed by Mao and co-workers (Mao, 2005), whereby the interaction between the UIMs and ubiquitin moieties anchors the substrate for optimal positioning of the scissile bond. Adapted from Mao, 2005.

Molecular docking of the Josephin structure (Nicastro, 2005) suggested the existence of two patches at the surface of the protein that could interact with ubiquitin or ubiquitin-like peptides, one close to the active site (site1) and one opposite it (site2). Further work by the same author successfully determined the structure of a Josephin mutant variant, in complex with ubiquitin at either one of these sites (Nicastro, 2009), later shown to be consistent with the structure of a Lys48-linked dimer (Winborn, 2008).
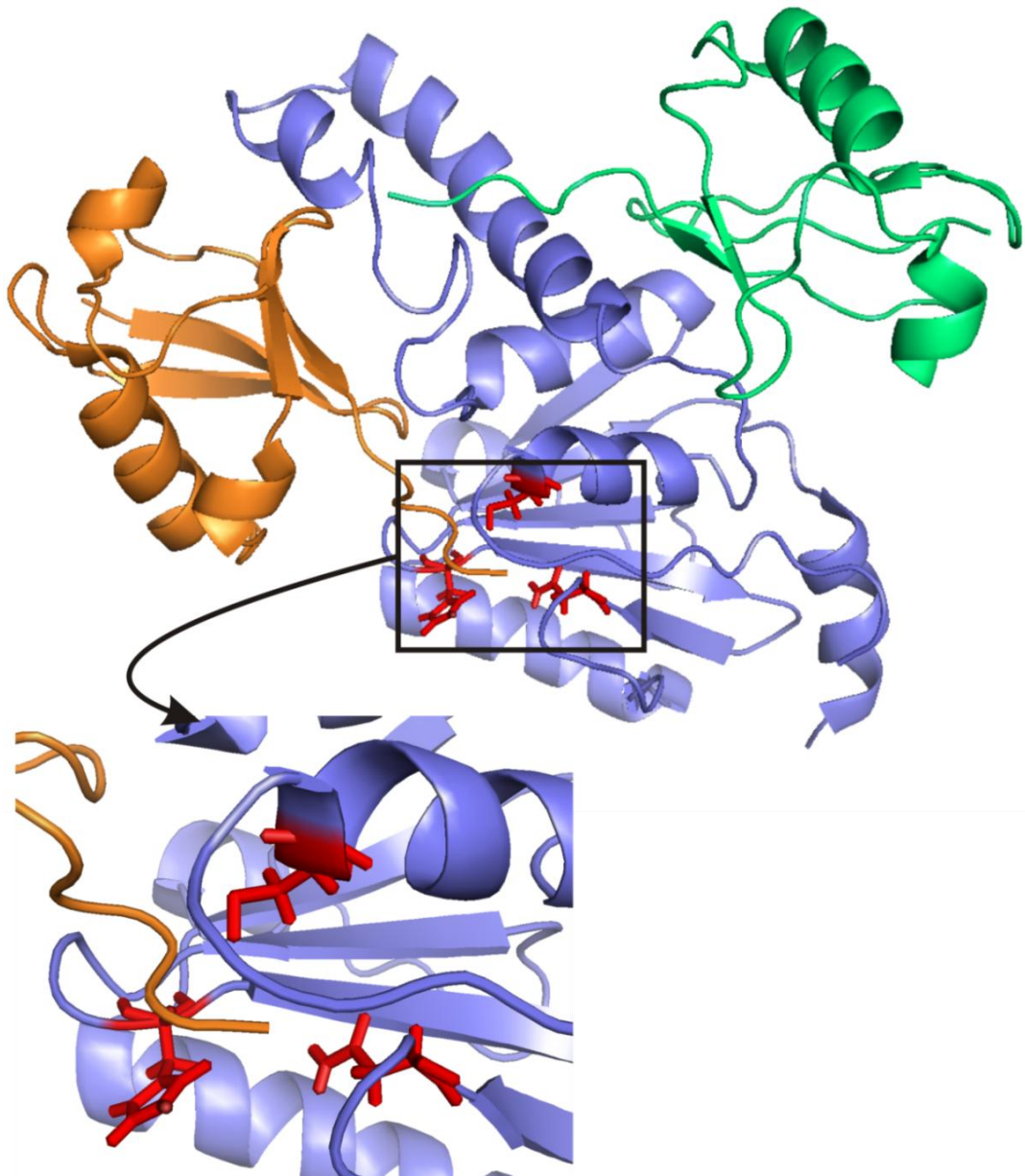
Fig. 1.13 – Cartoon representation of the Josephin domain interacting with ubiquitin. Two predicted sites for the interaction with substrate (Nicastro, 2005) were mutated and the structures of Josephin in complex with ubiquitin at each of the sites were determined (Nicastro, 2009). Ubiquitin molecules, shown in orange and green, occupy site1 and site2, respectively. Josephin is shown in blue (PDB access code: 2JRI), with its catalytic residues in red. Below is shown a closer view of the catalytic site and the position of site1-occupying ubiquitin's C-terminal.

While site 1 is necessary for cleavage to occur, possibly anchoring the ubiquitin molecule for optimal positioning of the scissile bond, site 2 overlaps with the modelled site for interaction with the ubiquitin-like domain of HHR23B (Nicastro, 2005), raising the question of a role by ubiquitin-like domains in wild-type

full-length ataxin-3. No obvious structural elements or loops are present in the structure of Josephin that could block access to the active site (Fig. 1.11), suggesting that substrate binding does not require major structural rearrangements, unlike the ubiquitin C-terminal hydrolases (UCH) and ubiquitin specific proteases (USP) families of DUBs (reviewed in Nijman, 2005). More recently, a new class of DUBs was created to include ataxin-3 and Josephin-containing proteins, the MJD family of DUBs (reviewed in Nijman, 2005). Interestingly though, recent work by Paulson's group (Todi, 2009) suggests that ataxin-3 is an inefficient protease when ataxin-3 itself is non-ubiquitinated (at least compared with the level of activity when ubiquitinated), consistent with reports by Chow and co-workers of low activity of ataxin-3 against the monoubiquitin substrate GST-Ub$_{52}$ (Chow, 2006), though low affinity interaction with monoubiquitin is expected for ataxin-3, as shown by Winborn and co-workers (Winborn, 2008).

1.4.3 Ataxin-3 aggregation and toxicity

Intracellularly, ataxin-3 is widely cytoplasmic, with the pathogenic isoform leading to the formation of neuronal inclusions (NI) *in vivo* (Paulson, 1997). Fujigasaki and co-workers dissected the architecture of the NI formed in cells, showing them to be comprised of a core of pathogenically expanded ataxin-3 surrounded by non-expanded protein and other polyQ-containing proteins, in turn surrounded by cellular machinery linked to stress response and protein clearance pathways (Fujigasaki, 2000). This recruitment of non-expanded ataxin-3 and other polyQ-containing proteins suggested a polyQ-mediated aggregation and a possible "loss of function" mechanism for toxicity, as sequestered proteins would be unable to perform their normal function, disrupting the cellular homeostasis and eventually leading to cell death. Furthermore, proteins such as Sup35p, Ure2p and HET-s are, like ataxin-3, comprised of a globular domain and an unstructured domain in their soluble states, the latter shown to be the region forming the cross-beta core of aggregation products (Paushkin, 1996; Balguerie, 2003; Baxa, 2005).

The flexible C-terminal tail of ataxin-3 holds the polyQ region, whose expansion is unequivocally linked to pathogenesis (Paulson, 1997, reviewed in Williams and Paulson, 2008 and Takahashi, 2010), making it an ideal candidate for the molecular instigator of ataxin-3 aggregation.

Surprisingly, *in vitro* studies have shown that ataxin-3 self-associates under physiologically-relevant conditions, regardless of polyQ tract length (Gales, 2005; Ellisdon, 2006). Both pathologically expanded and non-expanded forms of ataxin-3 were shown to aggregate through a nucleated assembly process, with a monomeric thermodynamical nucleus. Polymerization is accompanied by a conformational change of the globular domain (Josephin domain), with an increase of beta-sheet content concomitant with a loss of α-helical character (Ellisdon, 2006). Expanded and non-expanded recombinant ataxin-3 polymerizes into small spherical or spherical-like aggregates, 5-20nm in diameter, and short curvilinear aggregates of 10-12nm width and variable length that increases with incubation time (Fig. 1.14), consistent with reports for other polyQ-containing proteins and amyloid-related proteins (Ross and Poirier, 2004; Gosal, 2005; Xue, 2008). As the aggregation of ataxin-3 under physiologically-relevant conditions progresses, these curvilinear aggregates increase dramatically in length, reaching up to several hundred nanometers (nm), while retaining their width. The final morphology resembles that of worm-like fibrils described for the polymerization of proteins such as TTR and β_2m, formed under physiologically-relevant conditions and mild-denaturing conditions, respectively (Cardoso, 2002 and Gosal, 2005, respectively). However, polyQ tract length determines the properties of the end-products of aggregation, with pathologically expanded ataxin-3 leading to the formation of SDS-stable aggregates (Ellisdon, 2006), morphologically larger than the worm-like fibrils formed by non-expanded ataxin-3 and more fibrous (Fig. 1.14 - F).
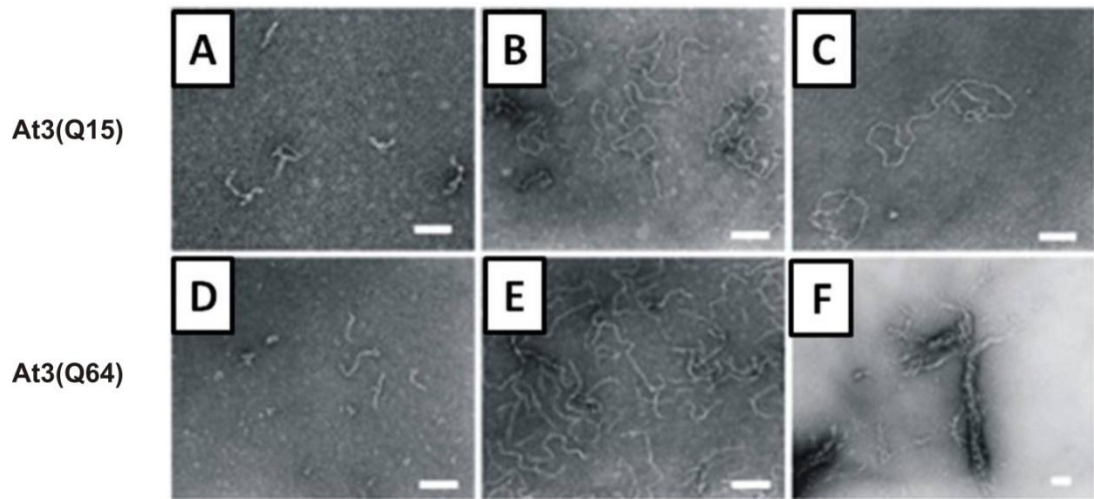
Fig. 1.14 – Negative-stain electron microscopy of aggregates formed during the aggregation of expanded (At3(Q64)) and non-expanded (At3(Q15)) ataxin-3 (2UIM isoform). A, B and C are representative of non-expanded ataxin-3 incubated at 37ºC under physiologically-relevant conditions for 11h, 48h and 100h, respectively. D, E and F are representative of aggregation of a pathologically expanded isoform of ataxin-3, with 64 glutamines, after 11h, 48h and 100h incubation under identical conditions to At3(Q15), respectively. The aggregates formed by the expanded isoform in the early steps of aggregation (D and E) resemble those formed by the non-expanded isoform (A, B and C). Later steps display aggregates of different morphology, that are thicker and longer (F). Bars represent 100nm. Taken from Ellisdon, 2006.

These aggregates are 50-100nm in width and up to many microns in length, with a tendency to further self-associate into large insoluble aggregates (Ellisdon, 2006), making their further characterization a technical challenge (Fig.1.14 - F). The formation of these SDS-resistant aggregates can be blocked by incubation with the "glutamine-breaker" peptide (QBP1), pointing to a direct role in assembly by the polyQ region (Ellisdon, 2006).

Based on these observations, a two-stage model for the self-association of ataxin-3 was proposed by Ellisdon and co-workers (Ellisdon, 2006) where a first stage of aggregation is polyQ-independent, followed by a polyQ-dependent stage that confers the aggregates' increased stability to SDS (Fig. 1.15). More recently, protein engineering studies have shown this model of aggregation for other naturally occurring polyQ-containing proteins, such as Htt (Ignatova, 2007), as well as engineered polyQ-containing B-domain (Robertson, 2008). An important consequence of this model is that the initial aggregation events cannot localize to

the polyQ tract, giving ataxin-3 at least two regions potentially involved in self-association. Furthermore, the presence of non-expanded ataxin-3 in pathological aggregates (Fujigasaki, 2000), together with the model, suggests an active aggregative role for the non-expanded protein in disease-states. Several possibilities for this putative second region linked to aggregation are suggested by data in literature, specifically the N-terminal Josephin domain and the region of the protein C-terminal from the polyQ tract.

Work by Gales and co-workers using a truncated form of ataxin-3 comprised solely of the N-terminal Josephin domain, indicated this domain's ability to self-associate into oligomers (dimers) under physiologically-relevant conditions. Like the oligomers formed by full-length ataxin-3, those formed from this truncation variant are positive for A11 binding, indicating the presence of a shared epitope and suggesting a role for this domain in early interactions between ataxin-3 molecules, possibly representing the second region of the protein involved in aggregation (as above). More recently, work by Masino and co-workers has shown that the Josephin domain alone is capable of self-association into fibrillar aggregates (Masino, 2011), undergoing structural rearrangement not unlike that which has been shown for the full-length protein (Ellisdon, 2006), including an increase in beta-sheet content upon polymerization. Morphological characterization of the aggregation products from a truncation mutant of ataxin-3 containing residues 1-180 of ataxin-3 the Josephin domain alone, revealed a heterogeneous population of curvilinear aggregates, with beaded morphology, 7-12nm in width and up to 170nm long, as well as globular structures of 7-12nm diameter (Masino, 2011), resembling early aggregation time-points of both expanded and non-expanded ataxin-3 (Ellisdon, 2006, discussed above). Mutational and *in silico* analysis of the Josephin domain identified two aggregation-prone hydrophobic patches on the surface of the protein (Fig. 1.13), that are also involved in substrate recognition (Nicastro, 2009; Masino, 2011), suggesting a link between function and self-association. Notably, neither competition for these sites by the substrate ubiquitin, nor mutationally disrupting the hydrophobicity of the patches completely abolishes

aggregation (Masino, 2011), indicating the need for further studies to fully appreciate the role of these regions in both function and aggregation.

Work by Harris and co-workers (Harris, 2010) has indicated that both in cell lines and brain tissue the 2UIM isoform of ataxin-3 is more aggregation-prone than the 3UIM isoform, suggesting the C-terminal portion of ataxin-3, downstream from the polyQ tract, is involved in mediating aggregation, unveiling another possible aggregation-prone region of the protein. Changes to the protein context can alter subcellular localization or even affect protein-protein interactions related to aggregation, justifying a closer look at the C-terminal portion of the protein. The UIM regions of ataxin-3 have been implicated in modulating protein aggregation, with work by Miller and co-workers showing that a construct comprising the 3UIM regions of ataxin-3 was able to, when co-expressed with Htt, decrease the formation of intracellular aggregates by the latter protein. It is reasonable to suggest a role by the third UIM of the 3UIM isoform in decreasing the aggregation propensity of the protein, when compared to the 2UIM isoform.

In addition, adding an exogenous nuclear export signal to ataxin-3 prevents the aggregation into inclusions, but also prevents aggregation (Bichelmeier, 2007). Accordingly, adding an exogenous nuclear localization signal to ataxin-3 increases phenotypical severity. A study of the relative influence of nuclear export signals and nuclear localizing signals in determining ataxin-3 localization, suggested the N-terminal portion of the protein to be involved in nuclear import, particularly effective when the portion of the protein upstream of the polyQ regions is present (Macedo-Ribeiro, 2009). It is interesting to consider the nuclear environment itself as having an active role in the pathology, supported by the previously described interaction of ataxin-3 with the nuclear matrix (Perez, 1999). This interaction was not, however, mapped to a specific region of the protein.
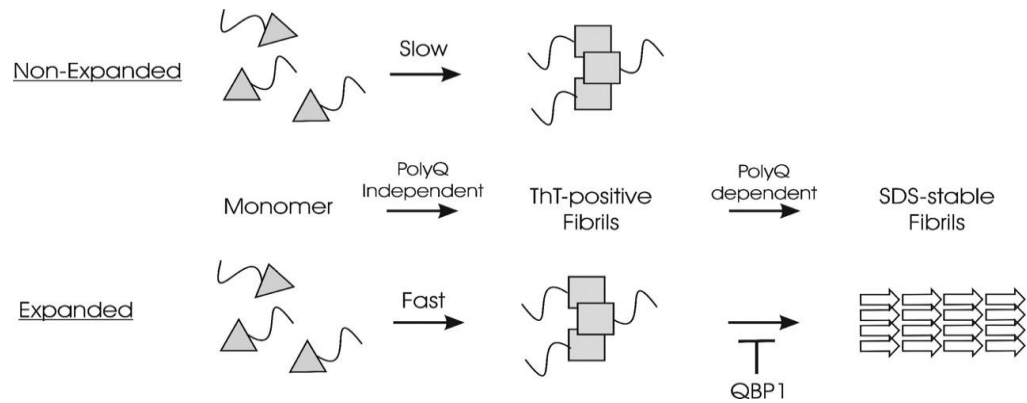
Fig. 1.15 – Schematic representation of the proposed two-step pathway for fibril formation by ataxin-3. The initial steps of the aggregation are similar with both expanded and non-expanded forms of the protein, and lead to the formation of fibrils, positive for Thioflavin T (ThT). A second step takes the aggregation process further, by the formation of SDS-stable fibrils, mediated by the polyQ-repeat length, as shown by the polyglutamine binding peptide (QBP1)-dependent blocking of this step. Taken from Ellisdon, 2006.

1.5 Objectives

The increase in life expectancy over the last four decades has lead to an increase in the numbers of patients with protein-aggregation diseases, particularly for non-hereditary diseases. The common structural ground of protein-aggregation diseases provides hope for common therapeutic strategies. However, the molecular events at the initial steps of pathology are still poorly known, very much due to their complexity and heterogeneity. Furthermore, despite commonalities in the protein aggregates of pathological states, the histopathology of disease varies greatly, even for ubiquitously expressed proteins, suggesting a modulation of toxicity by the specific protein context of pathology and stressing the need to address each disease individually. In fact, interactions with other molecules have been shown to modulate the aggregation of amyloid-forming proteins *in vivo*, cell cultures and *in vitro*. In addition, the accumulated knowledge of assembly mechanisms has allowed the recognition of peptide sequences prone to amyloid aggregation, though it is becoming clear that the environment plays just as

important a role, including the peptide sequences flanking these amyloid-prone stretches. This is particularly evident for polyQ disorders, where the related proteins share an expansion of glutamine repeats in their sequences. Region and domains flanking these amyloid-prone polyQ stretches are crucial in modulating at least the early steps of aggregation as shown both for naturally-occurring sequences and engineered constructs.

For MJD/SCA3, the N-terminal Josephin domain of ataxin-3 is proposed to mediate early interactions, with the polyQ-expanded tract directly involved in a second step of aggregation that determines the properties of the final aggregates. However, a 3UIM isoform of ataxin-3 was shown to be less aggregation-prone than its 2UIM counterpart, suggesting the C-terminal portion of the protein is a mediator of aggregation. Though the 3UIM isoform is the most prevalently expressed in brain tissue (where cellular degeneration leads to pathology), the majority of the *in vitro* work on ataxin-3 has focused on the 2UIM isoform, leaving a gap of knowledge in the aggregation details of the 3UIM isoform.

The first objective of this thesis is to fill that gap, providing a systematic study of the aggregation parameters for the 3UIM isoform of ataxin-3. Supported by the hypothesis of polyQ-independent initial steps in aggregation, the aggregation details of the non-expanded 3UIM ataxin-3 are detailed, namely the influence of environmental conditions and biochemical and biophysical parameters describing the assembly.

Fulfilment of this objective provides a reference point for further study of aggregation determinants in the protein's sequence, the second objective of this thesis. The aggregation details obtained for non-expanded 3UIM ataxin-3 are compared to those of truncated forms of the protein, narrowing down the region of the protein's sequence responsible for the early interactions.

Chapter 2

2 – Materials and Methods

2.1 Materials and Equipment

2.1.1 Molecular Biology

	Manufacturer:
Plasmid prep kits	
• HiSpeed Plasmid Midi kit	Qiagen
• FastPlasmid Mini kit	Eppendorf

	Manufacturer:
Small-volume orbital incubators	
• SI50	Stuart Scientifc
Small-volume centrifuges (bench-top)	
• 5430R	Eppendorf
• GenFuge 24D	Progen

2.1.1.1 Expression vectors

All plasmid vectors used in this thesis have the commercially available pDEST17 plasmid as parent vector (Table 2.1). The ataxin-3 nucleotide sequence described by Goto and co-workers (Goto, 1997) was cloned into the vector, under the control of a T7 promotor, using the Gateway technology developed by the manufacturer of the parent vector (Invitrogen). Also upstream of the cloning site, a region coding for a hexaHistidine tag (His6) on the resulting protein is present, followed by a 12-residue linker (Fig. 2.1). The sequence of this linker was mutated to incorporate a TEV protease cleavage site in the corresponding protein, though the experiments performed in this thesis used uncleaved recombinant ataxin-3 protein.

```
atgtcgtactaccatcaccatcaccatcacctcgaaaacctgtattttcagtctaaagca
 M  S  Y  Y  H  H  H  H  H  H  L  E  N  L  Y  F  Q ↓ S  K  A
ggctggatggagtccatcttccacgagaaacaagaaggctcactttgtgctcaacattgc
 G  W  M  E  S  I  F  H  E  K  Q  E  G  S  L  C  A  Q  H  C
ctgaataacttattgcaaggagaatattttagccctgtggaattatcctcaattgcacat
 L  N  N  L  L  Q  G  E  Y  F  S  P  V  E  L  S  S  I  A  H
cagctggatgaggaggagaggatgagaatggcagaaggaggagttactagtgaagattat
 Q  L  D  E  E  E  R  M  R  M  A  E  G  G  V  T  S  E  D  Y
cgcacgtttttacagcagccttctggaaatatggatgacagtggttttttctctattcag
 R  T  F  L  Q  Q  P  S  G  N  M  D  D  S  G  F  F  S  I  Q
gttataagcaatgccttgaaagtttgggggtttagaactaatcctgttcaacagtccagag
 V  I  S  N  A  L  K  V  W  G  L  E  L  I  L  F  N  S  P  E
tatcagaggctcaggatcgatcctataaatgaaagatcatttatatgcaattataaggaa
 Y  Q  R  L  R  I  D  P  I  N  E  R  S  F  I  C  N  Y  K  E
cactggtttacagttagaaaattaggaaaacagtggtttaacttgaattctctcttgacg
 H  W  F  T  V  R  K  L  G  K  Q  W  F  N  L  N  S  L  L  T
ggtccagaattaatatcagatacatatcttgcacttttcttggctcaattacaacaggaa
 G  P  E  L  I  S  D  T  Y  L  A  L  F  L  A  Q  L  Q  Q  E
ggttattctatatttgtcgttaagggtgatctgccagattgcgaagctgaccaactcctg
 G  Y  S  I  F  V  V  K  G  D  L  P  D  C  E  A  D  Q  L  L
cagatgattagggtccaacagatgcatcgaccaaaacttattggagaagaattagcacaa
 Q  M  I [R]↓ V  Q  Q  M  H  R  P  K  L  I  G  E  E  L  A  Q
ctaaaagagcaaagagtccataaaacagacctggaacgagtgttagaagcaaatgatggc
 L  K  E  Q  R  V  H  K  T  D  L  E  R  V  L  E  A  N  D  G
tcaggaatgttagacgaagatgaggaggatttgcagagggctctggcactaagtcgccaa
 S  G  M  L  D  E  D  E  E  D  L  Q  R  A  L  A  L  S  R  Q
gaaattgacatggaagatgaggaagcagatctccgcagggctattcagctaagtatgcaa
 E  I  D  M  E  D  E  E  A  D  L  R  R  A  I  Q  L  S  M  Q
ggtagttccagaaacatatctcaagatatgacacagacatcaggtacaaatcttacttca
 G  S  S [R]↓ N  I  S  Q  D  M  T  Q  T  S  G  T  N  L  T  S
gaagagcttcggaagagacgagaagcctactttgaaaaacagcagcaaaagcagcaacag
 E  E  L  R  K  R  R  E  A  Y  F  E  K  Q  Q  Q  K  Q  Q  Q
cagcagcagcagcagcagcaggggggacctatcaggacagagttcacatccatgtgaaagg
 Q  Q  Q  Q  Q  Q  Q  G  D  L  S  G  Q  S  S  H  P  C  E  R
ccagccaccagttcaggagcacttgggagtgatctaggtgatgctatgagtgaagaagac
 P  A  T  S  S  G  A  L  G  S  D  L  G  D  A  M  S  E  E  D
atgcttcaggcagctgtgaccatgtctttagaaactgtcagaaatgatttgaaaacagaa
 M  L  Q  A  A  V  T  M  S  L  E  T  V  R  N  D  L  K  T  E
ggaaaaaaataa
 G  K  K  *
```

Fig. 2.1 – Nucleotide sequence and the translated amino acid residue sequence for ataxin-3 proteins studied in this thesis. The N-terminal tag and linker to aid purification (highlighted in yellow) is part of the construct and the engineered recognition site for TEV protease is underlined, with the expected cleavage site labelled with a black arrow. The stop codon for the full-length protein is shown as an asterisk (*) and the nucleotide at the C-terminus of each truncation mutant is boxed in red or green, for the shortest or the longest truncation mutant prepared, respectively. Including the purification tag and linker, the smallest truncation mutant comprises 204 residues, while the longest truncation mutant comprises 284. The full-length protein is 383 residues long. Highlighted in red is the sequence corresponding to the N-terminal Josephin domain, while in green the ubiquitin interacting motifs (UIM) are highlighted. The polyglutamine tract is highlighted in blue. The nucleotide sequence, as obtained and confirmed by sequencing, was translated using the tools available on the Expasy proteomics server (http://expasy.org). The UIM sequences highlighted follow the work by Donaldson and co-workers (Donaldson, 2003).

Table 2.1 – Plasmid vectors mentioned in this thesis

Plasmid vector	Supplied by:
pDEST17 (parent vector)	Invitrogen
pDEST17 – MJD 1N (encoding the AT3[U3/14Q] protein)	Dr. Patricia Maciel (ICVS, Braga, Portugal)
pDEST17 – MJD D1 (encoding the AT3[2U/Jos] protein)	Sandra Macedo-Ribeiro (IBMC, Porto, Portugal)
pDEST17 – MJD J1 (encoding the AT3[Jos] protein)	Sandra Macedo-Ribeiro (IBMC, Porto, Portugal)

Truncation mutants were obtained by the in-frame mutation of a nucleotide triplet to a stop codon, corresponding to the residue after the intended truncation in the ataxin-3 protein. All resulting plasmids were sequenced to ensure their correct sequence (Section 2.3.2.2). The final sequence of the cloned genetic material is shown in Fig. 2.1, together with the translated protein sequence.

2.1.2 Bacterial Strains

	Manufacturer:
E. coli BL21 SI Genotype: (F- ompT lon hsdS$_B$(r$_B^-$m$_B^-$) gal dcm endA1 proUp::T7 RNAP::malQ-lacZ Tet^s)	Invitrogen
E. coli DH5α Genotype: (F- endA1 glnV44 thi-1 recA1 relA1 gyrA96 Φ80dlacZΔM15 deoR Δ(lacZYA-argF)U169 nupG hsdR17(rK-mK+) λ–)	Stratagene

2.1.3 Bacterial Cultures and Protein Expression

	Manufacturer:
Agar	Melford or Merck
Ampicillin (sodium salt)	Melford or USB
Calcium chloride ($CaCl_2$)	Fisher or Sigma
D-glucose	Sigma or Merck
Glycerol	Fisher or Panreac
Luria-Bertani (LB) complete mix	VWR or Merck
Lysozyme (hen egg white)	Sigma
Sodium chloride (NaCl)	Fisher or Merck
Tryptone	Melford or Fluka
Yeast Extract	Melford or Merck

	Manufacturer:
Large-volume orbital shaking incubators	
• Gallenkamp	Sanyo
• Minitron	INFORS AG
Small-volume orbital shaking incubators	
• SI50	Stuart scientific
Small-volume centrifuges (bench-top)	
• 5430R	Eppendorf
• GenFuge 24D	Progen
Avanti® J-26 XP large-volume centrifuge	Beckman-Coulter
Rotor J-LITE® JLA 8.1000	Beckman-Coulter
Rotor J-LITE® JA 25.50	Beckman-Coulter

2.1.4 Protein purification and chromatography

	Manufacturer:
2-propanol	Merck or Panreac
Dithiothreitol (DTT)	Melford or USB
DNAse	Sigma
Ethanol	Fisher or Merck
Ethylenediaminetetraacetic acid (EDTA)	Sigma or Merck
Glycerol	Fisher or Panreac
HEPES acid	Melford or Calbiochem
Imidazole	Fisher or USB
Magnesium chloride (MgCl$_2$)	Sigma or Merck
Nickel chloride (NiCl$_2$)	Merck or Sigma
Phenylmethylsulfonyl fluoride (PMSF)	Sigma or Roche
Sodium chloride (NaCl)	Fisher or Merck
Sodium hydroxide (NaOH)	Merck or Panreac
Sodium phosphate dibasic	Sigma or Merck
Sodium phosphate monobasic	Sigma or Merck
Tris base	Melford or Merck

	Manufacturer:
Purification columns	
• Sephacryl 200 (36/60) Hi Load	GE Healthcare
• Superdex 200 (10/300) GL	GE Healthcare
• HiPrep iMAC FF	GE Healthcare
• HisTrap HP (5ml)	GE Healthcare
FPLC systems	
• ÄKTA™ prime and prime plus	GE Healthcare
• ÄKTA™ purifier	GE Healthcare
• BioLogic DuoFlow™	Bio-Rad
Peristaltic pumps	
• Pump P-1	GE Healthcare
• EconoPump EP-1	Bio-Rad
Tubing and sample loops (includes 50ml Superloop™)	GE Healthcare

2.1.5 Protein concentration

	Manufacturer:
Concentration determination	
• 5X Protein Assay (Bradford)	Bio-Rad
• BCA protein assay (Lewy)	Pierce scientific (Thermo Inc.)
Amicon Ultra – 15 (10kDa cut-off) centrifugal concentrators	Millipore
UVette®, disposable cuvettes	Eppendorf
Bovine Serum Albumin (BSA)	Sigma

	Manufacturer:
Small-volume centrifuges (bench-top)	
• 5430R	Eppendorf
• GenFuge 24D	Progen
Avanti® J-26 XP large-volume centrifuge	Beckman-Coulter
Rotor J-LITE® JS 5.3	Beckman-Coulter
Spectrophotometers	
• UV-2401 PC	Shimadzu
• Ultrospec 2100	GE Healthcare

2.1.6 Protein analysis

2.1.6.1 Electrophoresis

	Manufacturer:
Tris-Tricine-SDS 10x buffer	Fermentas
Tris base	Calbiochem
Bicine	Sigma
Sodium dodecyl sulphate (SDS)	Bio-Rad
Coomassie Brilliant Blue	MP Biomedicals or Sigma
Tetramethylethylenediamine (TEMED)	Sigma or Merck
Ammonium persulfate (APS)	Sigma
2-propanol	Merck or Panreac
Acrylamide/Bis-acrylamide solution	
• (30% (37:1)	Severn Biotech
• (30% (37.5:1)	Bio-Rad
Protein markers	
• Benchmark Mark12™ unstained	Invitrogen
• Precision Plus unstained	Bio-Rad
Gel Staining	
• ProtoBlue Safe	Geneflow
• PageBlue™	Fermentas

	Manufacturer:
Electrophoretic apparatus	
• Mini-Protean® system III and IV	Bio-Rad
• miniPAGE electrophoresis	Atto
Gel documentation	
• InGenius	SynGene
• GelDoc™	Bio-Rad

2.1.6.2 Protein aggregation

	Manufacturer:
Dithiothreitol (DTT)	Melford or USB
EM grids	Agar or Tabaar
Ethylenediaminetetraacetic acid (EDTA)	Sigma or Merck
Glycerol	Fisher or Panreac
Glycine	Merck
HEPES acid	Melford or Calbiochem
MES acid	Sigma or Merck
Microplate coverslip	Thistle scientific
Phenylmethylsulfonyl fluoride (PMSF)	Sigma or Roche
Sodium Azide (NaN$_3$)	Sigma
Sodium chloride (NaCl)	Fisher or Merck
Thioflavin T (ThioT)	Sigma or Calbiochem
Tris base	Melford or Merck
Uranyl acetate	Agar or Tabaar
β-mercaptoethanol (BME)	Sigma
Microplates	
• P8741 FluoroNunc™ 96 well plates	Sigma
• C3915 black 96well plates	Corning Scientific
• 781076 black 384well microplates	Greiner Bio-One
• 3821BC black 384well microplates	Corning Scientific
Buffer exchange	
• NAP5 or NAP10 columns	GE Healthcare

	Manufacturer:
Microplate readers	
• SPECTRAmax® GEMINI-XS	Molecular Devices
• FluoStar Optima	BMG
• FluoStar Omega	BMG
• PolarStar Galaxy	BMG
Electron microscopes (EM)	
• EM 902	Zeiss
• CM-10	Philips
Gemini PX Ultra, X-ray Diffractometer	Oxford Diffraction
Biomek® 2000 series robot mixer and dispenser	Beckman-Coulter
Small-volume centrifuges (bench-top)	
• GenFuge 24D	Progen
Avanti® J-26 XP large-volume centrifuge	Beckman-Coulter
Rotor J-LITE® JS 5.3	Beckman-Coulter

2.1.7 Others

	Manufacturer:
Syringe filters (0.2μm pore size)	
• GyroDisc® Polyethersulfone (PES) membrane	Orange Scientific
• Minisart® Polyethersulfone (PES) membrane	Sartorius
Filters (0.2μm pore size)	
• Steritop™ Filter Units	Millipore
• Whatman® Mixed Ester (ME) membranes	Schleicher-Schuell
Urea	MP Biomedicals or Sigma
Alexa™ 488 Fluor dye (succinimidyl ester)	Invitrogen
Hydrochloric acid (HCl) solution	Fisher or J.T.Baker
Sodium hydroxide (NaOH) pellets	Sigma or Merck
Ammonium acetate	Sigma
Ubiquitin-aminomethyl coumarin (Ub-AMC)	Boston Biotech or Sigma

	Manufacturer:
Spectropolarimeters	
• Olis spectropolarimeter	On line Instrument Systems, Inc.
Spectrofluorometer	
• Quantamaster™	PTI
Mass spectrometer	
• LCT Premier	Waters
Micropipettes	Gilson (all)
• P5000 (1000-5000μl)	
• P1000 (200-1000μl)	
• P200 (20-200μl)	
• P20 (2-20μl)	
• P2 (0.2-2μl)	

2.2 Stocks and solutions

For practicality and reproducibility, concentrated stocks of reagents were prepared and stored, whenever possible.

2.2.1 Preparation of stock solutions

2.2.1.1 Ampicillin

Ampicilin solutions were prepared by dissolution of ampicilin sodium salt (Section 2.1.3) in RO-purified water to a final concentration of 100 mg/ml, followed by sterilisation by filtration through a sterile 0.2 μm pore polyethersulphone (PES) syringe filter (Section 2.1.7). Sterile ampicilin solutions were used at 100μg/ml final concentration, either immediately or aliquoted and stored for up to a month at -20ºC.

2.2.1.2 DNAse

DNAse in powder form (Section 2.1.4) was dissolved to a final concentration of 2mg/ml in 10 mM Tris-HCl pH 7.5, 50 mM NaCl, 10 mM $MgCl_2$, 1 mM DTT and 50% (v/v) glycerol (prepared as described in Section 2.2.1.11). The solution was next filter-sterilized (Section 2.1.7), and aliquots prepared and stored at -20ºC.

2.2.1.3 DTT

To prepare this stock, reagent powder (Section 2.1.4) was dissolved in RO-purified water to a final concentration of 1M, filter-sterilized (Section 2.1.7), aliquoted and stored at -20ºC. Due to this reagent's ease of oxidation at room temperature, only freshly prepared or freshly thawed solutions were used.

2.2.1.4 EDTA

To prepare this stock, EDTA (Section 2.1.4) was dissolved to a final concentration of 500mM in pH 8.0 RO-purified water. Concentrated NaOH (Section 2.2.1.7) was used to set the pH. The solution was next filter-sterilized (Section 2.1.7) and stored at room temperature.

2.2.1.5 HEPES

Buffer stock was prepared by dissolving HEPES acid (Section 2.1.4) in RO-purified water to a final concentration of 500mM. pH was set to 7.5 or 7.0 (as required) with concentrated NaOH (Section 2.2.1.7). The solution was next filter-sterilized (Section 2.1.7) and stored at room temperature.

2.2.1.6 MES

Buffer stock was prepared by dissolving MES acid (Section 2.1.6.2) in RO-purified water to a final concentration of 500mM. pH was set to 6.5 or 6.0 (as required) with concentrated NaOH (Section 2.2.1.7). The solution was next filter-sterilized (Section 2.1.7) and stored at room temperature.

2.2.1.7 Single salt solutions

Reagent salt (Section 2.1.4 and 2.1.6.2) was dissolved in RO-purified water to the final concentration required. The solution was next filter-sterilized (Section 2.1.7) and stored at room temperature. This stock preparation method was applied to the following solutions:

- $MgCl_2$ (1M final concentration)
- $NiCl_2$ (0.1M final concentration)
- NaN_3 (1% (w/v) final concentration)
- NaCl (5M final concentration)
- NaOH (5M, 1M and 0.1M final concentration)

2.2.1.8 Phosphate buffer

Phosphate buffer stock at pH 7.5 was prepared from the mixture of two phosphate solutions: 1 – Sodium phosphate monobasic (Section 2.1.4), dissolved in RO-purified water to a final concentration of 200mM; 2 – Sodium phosphate dibasic (Section 2.1.4), dissolved in RO-purified water to a final concentration of 200mM.

Solutions 1 and 2 were mixed to achieve a pH of 7.5, after which the resulting buffer solution was filter-sterilized (Section 2.1.7) and stored at room temperature.

2.2.1.9 PMSF

To prepare a stock of this protease inhibitor, reagent powder (Section 2.1.4) was dissolved in 2-propanol to a final concentration of 200mM and stored at room temperature for no longer than a month, as described by the supplier.

2.2.1.10 Tris-HCl buffer

Buffer stock was prepared by dissolving Tris base (Section 2.1.4) in RO-purified water to a final concentration of 400mM or 500mM, and pH set to 8.8, 7.5 or 6.8 (as required) with concentrated HCl solution (Section 2.1.7). The buffer was then filter-sterilized (Section 2.1.7) and stored at room temperature.

2.2.1.11 DNAse buffer

Stock solutions of Tris-HCl (pH 7.5), NaCl, $MgCl_2$ and DTT (Sections 2.2.1.10, 2.2.1.7 and 2.2.1.3) were diluted with RO-purified water to final concentrations of 20mM, 100mM, 20mM and 2mM concentrations, respectively. This solution was further diluted 2-fold with 100% (v/v) glycerol, for final concentrations of 10mM Tris-HCl pH 7.5, 50mM NaCl, 10mM MgCl2, 1mM DTT and 50% (v/v) glycerol, filter-sterilized and stored at -20ºC.

2.2.1.12 Sterile glycerol solution

To prepare sterile glycerol stock, glycerol reagent (Section 2.1.4) at 100% (v/v) was diluted to 87% (v/v) final concentration, and aliquoted into 5ml aliquots. These were next autoclaved at 121ºC and 15 p.s.i. for 20min and, after cooling down, stored at room temperature for up to 3 months.

2.2.1.13 5TXB buffer

Concentrated TXB buffer (5TXB) was prepared from stocks of its components (Sections 2.2.1.3, 2.2.1.4, 2.2.1.5 and 2.2.1.7) in RO-purified water, to final concentrations of 100mM, 1M, 5mM and 0.05% for HEPES, NaCl, EDTA and NaN$_3$, respectively, and filter-sterilized (Section 2.1.7), prior to storage at room temperature, shielded from light. The volume prepared was sufficient for all experiments using this buffer in the present thesis.

2.2.1.14 Calcium chloride

To prepare stocks of calcium chloride solutions, salt reagent was dissolved in RO-purified water with or without 30% (v/v) glycerol (final concentration), to a final concentration of 100mM. The solution was next autoclaved at 121ºC and 15 p.s.i. for 20 min and stored at room temperature.

2.2.1.15 ThioT

Thioflavin T reagent (Section 2.1.6.2) was dissolved in RO-purified water and filter-sterilized (Section 2.1.7). Concentration of the dye was determined as described by Wall and co-workers (Wall, 1999), by measuring the solution's absorbance at 416nm, knowing the extinction coefficient of the reagent: (ε^{mM}_{416} = 26.620). Stock was stored at 4ºC.

2.2.1.16 SDS-PAGE loading buffer

Two-fold concentrated stocks were prepared by diluting Tris-HCl (pH 6.8) and DTT stocks (Section 2.2.1.10 and 2.2.1.3, respectively) to final concentrations of 100mM and 200mM, respectively, with RO-purified water supplemented with SDS and glycerol (Section 2.1.6), to obtain final concentrations of 4% (w/v) and 20% (v/v), respectively. Coomassie brilliant blue (Section 2.1.6) was dissolved in the solution to final concentration of 0.2% (w/v) and aliquots frozen at -20ºC until required.

2.2.1.17 Tris-Bicine-SDS

Tris base and bicine reagents (Section 2.1.6.1) were dissolved in RO-purified water to final concentrations of 25mM and 250mM, respectively, and supplemented with SDS (Section 2.1.6.1) to 0.1% (w/v) final concentration. pH was set to 8.3 and the buffer stored for short periods of time (up to 1 week) at room temperature.

2.2.1.18 Tris-SDS

Tris base reagent (Section 2.1.6.1) was dissolved to final concentrations of 1M or 1.5M in RO-purified water supplemented with SDS to a final concentration of 0.1%(w/v) (Section 2.1.6.1). pH was next set to 6.8 or 8.8, respectively, with concentrated HCl solution (Section 2.1.7) and the buffers stored at room temperature. A third buffer stock was prepared the same way, albeit at 3M Tris and 0.3% (w/v) SDS final concentrations, and its pH set to 8.45.

2.2.2 Preparation of ready-use solutions

The following solutions were never stored but rather used immediately, once prepared. All solutions used for chromatographic studies were degassed by applying vacuum for 1h/L of solution.

2.2.2.1 Buffer A (low imidazole concentration)

Imidazole sodium salt (Section 2.1.4) was dissolved to a final concentration of 10mM in RO-purified water, in which NaCl stock (Section 2.2.1.7) and phosphate buffer stock (Section 2.2.1.8) were diluted to 500mM and 20mM final concentrations, respectively. pH was set to 7.5 with HCl solution (Section 2.1.7) and the buffer was filter-sterilized (Section 2.1.7) and shielded from direct light.

2.2.2.2 Buffer B (high imidazole concentration)

Imidazole sodium salt (Section 2.1.4) was dissolved to a final concentration of 500mM in RO-purified water, in which NaCl stock (Section 2.2.1.7) and phosphate buffer stock (Section 2.2.1.8) were diluted to 500mM and 20mM final concentrations, respectively. pH was set to 7.5 with HCl solution (Section 2.1.7) and the buffer was filter-sterilized (Section 2.1.7) and shielded from direct light.

2.2.2.3 Storage buffer

Stock solutions of HEPES, NaCl and EDTA (Sections 2.2.1.4, 2.2.1.5 and 2.2.1.7) were diluted to final concentrations of 20mM, 500mM and 1mM, respectively, in RO-purified water with a final concentration of 5% glycerol (v/v) (Section 2.1.4). Thawed DTT stock (Section 2.2.1.3) was added to the buffer, to a final concentration of 1mM. The pH of the buffer was next set to 7.5 with concentrated NaOH (Section 2.2.1.7) and the solution filter-sterilized (section 2.1.7).

2.2.2.4 Circular dichroism buffer

To prepare the buffer used in circular dichroism experiments (CD), stocks of phosphate buffer (pH 7.5) and EDTA were diluted to 50mM and 1mM EDTA final concentration, respectively. The solution was next filter-sterilised prior to use.

2.2.2.5 Ubiquitin protease activity buffer

Activity buffer was prepared by dilution of Tris-HCl (pH 7.5), EDTA and DTT stocks (Sections 2.2.1.3, 2.2.1.4 and 2.2.1.10) with RO-purified water, to final concentrations of 50mM, 1mM and 10mM, respectively. BSA was added from a 2mg/ml intermediate stock prepared in RO-purified water, to a final concentration of 0.1mg/ml. The buffer was filter-sterilized (Section 2.1.7) prior to use.

2.2.2.6 Glycine-ThioT

Following the guidelines published by LeVine III in 1999, glycine buffer was prepared by dissolving the reagent (Section 2.1.4) in RO-purified water to a final concentration of 50mM. pH was set to 9.0 with concentrated NaOH solution (Section 2.2.1.7) and the buffer filter-sterilized (Section 2.1.7). Thioflavin T was then added (Section 2.1.6.2) and its concentration determined as described in Section 2.2.1.15.

2.2.2.7 Aggregation buffers

2.2.2.7.1 TX buffer

Prepared by the dilution of stock solutions of HEPES, NaCl, DTT, NaN_3 and EDTA (Sections 2.2.1.3, 2.2.1.4, 2.2.1.5 and 2.2.1.7) with RO-purified water, supplemented with glycerol (to a final concentration of 5% (v/v)), to final concentrations of 20mM, 200mM, 1mM, 0.01% (w/v) and 1mM, respectively. The final solution was filter-sterilized (Section 2.1.7) before use.

2.2.2.7.2 TX buffer variations

Prepared as described above (Section 2.2.2.7.1), albeit omitting the component tested, as mentioned (Section 4.2). The pH was set to 7.5 and the buffer filter-sterilized (Section 2.1.7), prior to use.

2.2.2.7.3 Storage buffer

(see Section 2.2.2.3)

2.2.2.7.4 TXB

Prepared from stock solutions of HEPES, NaCl, DTT, NaN$_3$ and EDTA (Sections 2.2.1.3, 2.2.1.4, 2.2.1.5 and 2.2.1.7) diluted with RO-purified water to final concentrations of 20mM, 200mM, 1mM, 0.01% (w/v) and 1mM, respectively, this solution was next filter-sterilized (Section 2.1.7) prior to use.

2.2.2.7.5 TXBa

Buffer stock 5TXB (Section 2.2.1.13) was diluted 5-fold with RO-purified water containing DTT, freshly thawed (Section 2.2.1.3) and at a final concentration of 1mM. The buffer was filter-sterilized (Section 2.1.7) prior to use.

2.2.3 Media preparation

2.2.3.1 Liquid media

2.2.3.1.1 Luria-Bertani (LB) medium

Bacterial cells of the *E. coli* DH5$_\alpha$ strain (Section 2.1.2) were cultured in LB media, prepared as follows: 10g tryptone, 5g yeast extract and 10g NaCl (Section 2.1.3) were dissolved to a final volume of 1L in RO-purified water and sterilised by autoclaving at 121ºC and 15 p.s.i. for 20min. After cooling, ampicilin to a final concentration of 100µg/ml was added when appropriate, from thawed stock (Section 2.2.1.1).

2.2.3.1.2 Luria-Bertani medium with low osmolarity (LBON)

E. coli bacterial cells of the BL21 SI strain (Section 2.1.2*)* were cultured in LB media without NaCl (LBON), composed of 10g tryptone and 5g yeast extract (Section 2.1.3), dissolved *per* 1L RO-purified water and sterilised by autoclaving at 121ºC and 15 p.s.i. for 20min. After cooling, ampicilin to a final concentration of 100 µg/ml was added when appropriate from thawed stock (Section 2.2.1.1).

2.2.3.2 Solid media

Media prepared as described above (Section 2.2.3.1) was supplemented with 10 g of agar (Section 2.1.3) *per* liter, prior to sterilisation by autoclaving. After the sterilisation procedure the medium was cooled and ampicilin was added from thawed stock (Section 2.2.1.1) to a final concentration of 100 µg/ml, when appropriate, and gently mixed. Approximately 20-25 ml of the molten mixture was dispensed *per* sterile petri dish and allowed to fully cool at room temperature before storage at 4ºC (up to 1 week).

2.3 – Methods

2.3.1 Bacterial cultures

2.3.1.1 Starter cultures

Bacterial starter cultures (5ml-500ml) were inoculated with a single bacterial colony, picked from solid media (Section 2.2.3.2) with a sterile micropipette tip. This tip was dropped into autoclaved liquid media (Section 2.2.3.1) containing ampicilin (Section 2.2.1.1) when appropriate. Alternatively, inoculation of starter cultures was achieved by dipping a flame-sterilized wire loop (allowed to cool) in a glycerol stock of bacterial cells (Section 2.3.1.2), and rinsing in the culture medium with appropriate antibiotic. The culture was next incubated overnight (ca. 14 h) at 37ºC in an orbital incubator shaking at 200 rpm.

2.3.1.2 Glycerol stocks

Glycerol stocks of bacterial cells, both transformed with the plasmids mentioned in this thesis (Table 2.1) or without them, were stored at -80ºC. To prepare the glycerol stocks, 500µl liquid bacterial culture grown overnight was combined with 250µl fresh media (Section 2.2.3.1) and 250µl of 87% (v/v) glycerol solution (Section 2.2.1.12). After gentle mixing, stocks were flash-frozen in liquid nitrogen and stored at -80ºC.

2.3.2 Molecular biology

2.3.2.1 Plasmid preparation

Liquid cultures inoculated with plasmid-transformed DH5$_\alpha$ cells (obtained either by fresh transformation as described in Section 2.3.3.2 or from previous glycerol stocks prepared as described in Section 2.3.1.2) were grown overnight (ca. 14h) in 150ml LB liquid media with antibiotic selection (Section 2.2.3.1). This culture was then used to isolate and purify plasmid DNA with plasmid prep kits (Section 2.1.1), according to manufacturer's instructions.

2.3.2.2 DNA sequencing

When required, plasmid DNA was sequenced from the T7 promoter, by StabVida Lda. (Caparica, Portugal), ensuring the correct nucleotide sequence was present. The hybridizing primers used in sequencing are detailed in Table 2.2.

Table 2.2 – Primers used in sequencing the nucleotide sequences utilized in this thesis.

Primer name	Primer sequence	Supplier
T7 Fwd	5' – TAA TAC GAC TCA CTA TAG G – 3'	StabVida Lda.
T7 Rev	5' – GCT AGT TAT TGC TCA GCG G – 3'	StabVida Lda.

2.3.3 Transformation procedures

2.3.3.1 Preparation of competent cells

Starter cultures grown overnight (Section 2.3.1.1) were diluted 100-fold into fresh liquid media (Section 2.2.3.1), lacking antibiotic, and incubated at 37ºC with 200rpm shaking, until an optical density of 0.4-0.45 was reached at 600nm (OD_{600}). Cultures were next placed on ice for 10min, prior to being centrifugated at 4000rpm (Beckman-Coulter JA 25.50) for 10min at 4ºC (Section 2.1.3). The supernatant was discarded and the cell pellets resuspended gently with 10ml pre-chilled $CaCl_2$ solution without glycerol (Section 2.2.1.14). The resuspended cells were next incubated on ice for 10min and centrifugated once more, as described above. Cell pellets collected after discarding the supernatant were resuspended in 2ml $CaCl_2$ solution containing 30% (v/v) glycerol (Section 2.2.1.14) and 100µl aliquots were transfered into sterile microcentrifuge tubes, flash-frozen in liquid nitrogen and stored at -80ºC.

2.3.3.2 Bacterial cell transformation

Aliquots of competent cells (Section 2.3.3.1) were thawed on ice for at least 2h, after which they were mixed gently with 20-200ng of the relevant plasmid DNA and incubated on ice for a further 30min. At the end of this incubation the inoculated aliquots were heat shocked at 42ºC for 55sec and returned to ice for a further 5min. 250µl of fresh liquid media (Section 2.2.3.1) was added and the mixture placed at 37ºC with shaking at 200rpm in an orbital incubator for 1h

(Section 2.1.3). The resulting cell solution was next spread onto solid media (Section 2.2.3.2) with the appropriate antibiotic and incubated at 37ºC for ~14h.

2.3.4 Electrophoretic techniques

Gel electrophoresis was used to analyse protein samples based on their molecular weight, using discontinuous sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) with a stacking (at 4% (w/v) acrylamide) and a resolving gel (at 15% (w/v) acrylamide). Two different protocols were followed, depending on availability of equipment.

2.3.4.1 Sample preparation

Protein samples were diluted 2-fold with loading buffer (Section 2.2.1.16) and heated to 95ºC for 5min before loading in the gel wells. At least one well per gel was loaded with protein marker (Section 2.1.6.1). When analysing bacterial cultures, these were first centrifugated at 16000x g for 30min, and the pellet ressupended in imidazole buffer (Buffer A, Section 2.2.2.1). This mixture was next prepared as above, prior to loading in the gel wells.

2.3.4.2 Protocol A – Tris-bicine buffered gels

2.3.4.2.1 Preparing the gel

Stacking and resolving gel solutions were prepared from stocks (Section 2.2.1.17) and reagents (Section 2.1.6.1) according to Table 2.3. Solutions were poured between specifically designed 0.75mm-spaced glass plates, assembled in

cast frames and placed in cast stands, as described by the manufacturer of Mini-Protean® systems (Section 2.1.6.1). The resolving gel solution was poured first, followed by a solution of 50% (v/v) 2-propanol (Section 2.1.6.1) in RO-purified water. After setting, the 2-propanol solution was rinsed off the top of the gel using RO-purified water, and the stacking gel solution poured over the polymerised resolving gel. A well comb was next inserted at the top, and the gel allowed to set at room temperature.

Table 2.3 – Mixing amounts to prepare acrylamide solutions for SDS-PAGE electrophoresis.

Solution component	Resolving gel Volume to add (ml)	Stacking gel Volume to add (ml)
30% (w/v) Acrylamide (Bio-Rad)	7.5	1.0
1.5M Tris-HCl pH 8.8, 0.4% (w/v) SDS	3.8	
1M Tris-HCl pH 6.8, 0.8% (w/v) SDS		0.75
H_2O (RO-purified)	3.65	4.2
10% (w/v) ammonium persulphate in RO-purified water	0.10	0.06
TEMED	0.010	0.006

2.3.4.2.2 Running the gel

Sample and standard was applied to the gel wells, and electrophoresis performed on a Mini-Protean® Cell, with inner and outer reservoir filled with Tris-bicine-SDS buffer (Section 2.2.1.17). A constant current of 100V was applied to the system until samples reached the end of the stacking gel. At this point current was increased to 150V until adequate separation of the samples was achieved. Electrophoresed gels were stained with PageBlue™ (Section 2.1.6.1) and destained in RO-purified water as described by the manufacturer. Gel images were taken, using a gel documentation equipment (Section 2.1.6.1), and stored in digital format.

2.3.4.3 Protocol B – Tris-tricine buffered gels

2.3.4.3.1 Preparing the gel

Stacking and resolving gel solutions were prepared from stocks (Section 2.2.1.18) and reagents (Section 2.1.6.1) according to Table 2.4 and poured between 2 Atto miniPAGE kit glass plates assembled with a 1.5mm spacer as described by the manufacturer (Section 2.1.6.1). The resolving gel solution was poured first, immediately followed by the stacking gel solution. A well comb was inserted at the top and the gel allowed to set at room temperature.

Table 2.4 – Mixing amounts to prepare acrylamide solutions for SDS-PAGE electrophoresis.

Solution component	Resolving gel Volume to add (ml)	Stacking gel Volume to add (ml)
30% (w/v) Acrylamide (Severn Biotech)	7.5	0.83
3 M Tris, 0.3% (w/v) SDS pH 8.45	5.0	1.55
H_2O (RO-purified)	0.44	3.72
Glycerol	2.0	-
10% (w/v) ammonium persulphate in RO-purified water	0.05	0.10
TEMED	0.005	0.005

2.3.4.3.2 Running the gel

Electrophoresis was performed on an Atto miniPAGE system, with the inner reservoir filled with Tris-Tricine-SDS buffer (Section 2.1.6.1) diluted 10-fold in RO-purified water, and the outer reservoir filled with 400mM Tris-HCl, pH 8.8 (Section 2.2.1.10). After loading the samples and protein markers in the gel wells, a constant current of 100V was applied to the system until the dye front reached the bottom of the stacking gel, at which point current was increased to 150V until adequate separation of the samples was achieved. Electrophoresed gels were stained with ProtoBlue Safe solution (Section 2.1.6.1) and destained in RO-purified water as described in the manufacturer's instructions. Gel images were taken using a Syngene InGenius gel documentation system and stored in digital format (Section 2.1.6.1).

2.3.5 Expression procedures

2.3.5.1 Preparation of induction media

Induction media for the expression of ataxin-3 and constructs thereof was prepared as follows: 20g of Tryptone digest from casein, 10g yeast extract and 175.3g NaCl (Section 2.1.3) were dissolved in RO-purified water, autoclaved at 121ºC and 15 p.s.i. for 20min and allowed to cool down before use.

2.3.5.2 Overexpression of recombinant ataxin-3

To overexpress recombinant ataxin-3 protein, a starter culture of BL21 SI bacterial cells (Section 2.1.2) transformed with the relevant plasmid vector (Section 2.3.3.2) (Table 2.1) was grown overnight (ca. 14h) in LBON (Section 2.2.3.1.2), in the presence of a final concentration of 100µg/ml ampicilin (Section 2.2.1.1). Aliquots of this starter culture were next diluted 25-fold into 2.5L flasks containing 500ml sterilised LBON media. To each flask ampicilin (Section 2.2.1.1) was added to a final concentration of 100ug/ml, as well as D-glucose (Section 2.1.3) to a final concentration of 0.3% (w/v), (from a 30% (w/v) glucose stock in RO-purified water previously autoclaved at 121ºC at 15 p.s.i. for 20min) (Table 2.5). Culture flasks were next incubated at 37ºC and 200rpm agitation in an orbital incubator until OD_{600} measured 0.6-0.7. At this point fresh glucose and ampicilin solutions where added, just as before (Table 2.5), and protein expression was induced by the addition of 50ml induction media (Section 2.3.5.1) to a final concentration of 300mM NaCl. Culture flasks were returned to the orbital incubator and placed at 30ºC with shaking at 200rpm for 3h (Section 2.1.3), after which cells were harvested

by centrifugation at 6500rpm (Beckman JLA 8.1000) for 20min at 4ºC and the supernatant discarded. Cell pellets were resuspended in 10ml buffer A (Section 2.2.2.1) per liter of culture, and lysozyme (Section 2.1.3) added to a final concentration of 50µg/ml, prior to storage at -20ºC until required.

Table 2.5 – Volumes added to each culture flask during the expression procedures for ataxin-3.

	Starter culture	*Glucose*	*Ampicilin*	*Induction media*
Growth step	20ml	5ml	500µl	-.-
Induction step	-.-	5ml	500µl	50ml

2.3.6 Purification procedures

2.3.6.1 Extracting soluble protein

Frozen cell pellets from expression of recombinant ataxin-3 in BL21 SI cells (Section 2.2.8.1.2) were thawed in a sealed vessel under running water. DNase, MgCl$_2$ and PMSF were added from stocks (Section 2.2.1) to final concentrations of 10ug/ml, 20mM and 2mM, respectively. The mixture was next incubated at 4ºC with mild stirring for 90min, after which it was centrifugated at 12000rpm (Beckman JA 25.50 rotor) for 40min at 4ºC, (Section 2.1.3), and the supernatant collected.

2.3.6.2 Affinity chromatography

All buffers used were degassed under vacuum for 1h/L prior to use. The supernatant obtained in Section 2.3.6.1 was loaded (with peristaltic pumps) onto a 5ml Histrap HP column or a 20ml HiPrep IMAC FF column (Section 2.1.4), previously charged with Ni^{2+} ion (Section 2.2.1.7) and equilibrated with 10 column volumes of buffer A (Section 2.2.2.1), as described by the manufacturer. After loading the sample, the column was next washed with 10 column volumes of buffer A and the protein eluted, at half a column volume per minute, in a step-wise gradient of buffer B (Section 2.2.2.2) over 20 column volumes, as detailed in Table 2.6. Typically one peak was observed per gradient step and fractions were collected manually, with the aid of real-time signal measurements of the A_{280} using an ÄKTA prime FPLC system (Section 2.1.4) operating at room temperature. SDS-PAGE analysis (Section 2.3.4) of the peaks obtained showed that the first peak was mostly comprised of impurities and only fractions from the second step (at 20% (v/v) buffer B) were considered for further purification. The third step of the gradient was designed solely to clean the column of the remaining bound proteins.

Table 2.6 – Stepwise gradient for elution of ataxin-3 from the affinity chromatography column.

	Step 1 – 50mM Imidazole	*Step 2 – 100mM Imidazole*	*Step 3 – 500mM Imidazole*
Flow rate	½ column volume/min	½ column volume/min	½ column volume/min
Length of step	5 column volumes	10 column volumes	5 column volumes

2.3.6.3 Preparative size exclusion chromatography

Solutions used were degassed under vaccum for 1h/L, prior to use. Soluble protein purified by affinity chromatography (Section 2.3.6.2) was further purified with a Sephacryl 200 (36/60) HiLoad column (Section 2.1.4), previously flushed with 2 column volumes of RO-purified water (filtered and degassed) and equilibrated with 2 column volumes of storage buffer (Section 2.2.2.3). Loading of the sample, to a limit of 12ml, was performed with the aid of a 50ml Superloop^{TM} (Section 2.1.4). Whenever the volume of protein-containing solution to purify was larger than 12ml, the sample was concentrated (Section 2.3.7.1), prior to loading onto the column. Elution, monitored by the A$_{280}$ signal, was conducted with varying flow rates, controlled by an ÄKTA prime (or prime plus) FPLC (Section 2.1.4) at 4ºC and according to Table 2.7. Fractions of 5ml volume were collected over 333ml (steps 3 through 5 in Table 2.7) at a flow rate of 0.5ml/min.

Table 2.7 – Stepwise elution details of ataxin-3 from the preparative grade gel filtration.

	Step 1	Step 2	Step 3	Step 4	Step 5
Length of step	5ml	12ml	63ml	200ml	70ml
Injection valve setting	Load	Inject	Load	Load	Load
Flow rate	1ml/min	0.5ml/min	0.5ml/min	0.5ml/min	1ml/min
Fraction collector status	Not collecting	Not collecting	Collecting	Collecting	Collecting

2.3.6.4 Analytical size exclusion chromatography

Solutions used were degassed under vaccum for 1h/L, prior to use. Analytical chromatography of the fractions obtained in Section 2.3.6.3 was performed with a Superdex 200 column (Section 2.1.4), flushed with 1.5 column volumes of RO-purified water and equilibrated with 2 column volumes of storage buffer (Section 2.2.2.3). The volume of sample injected per fraction was typically 100µl and the monomeric state of the sample evaluated by elution with 1.5 column volumes of storage buffer. Both elution and injection of sample were performed with an isocratic flow rate of 0.5ml/min, controlled by an ÄKTA prime or ÄKTA purifier FPLC system (Section 2.1.4), at room temperature. Only fractions displaying a single peak at A_{280} were considered for concentration, which typically involved pooling 5 fractions, centred on the corresponding monomeric protein peak.

2.3.7 Protein concentration

2.3.7.1 Centrifugal concentration

Centrifugal Amicon Ultra-15 filter concentrators (Section 2.1.5) were used to concentrate protein samples, with high recovery. The concentrators used were of a 10000Da cut-off and the manufacturer's protocol of use was followed. Briefly, after equilibrating the membrane with 20ml of the respective buffer, up to 10ml sample was applied to the concentrators. These were then centrifugated at 3000x g for 3-15min, depending on volume, after which the permeate was removed, more sample added and the device returned to the centrifuge for another centrifugation cycle, as described. When a suitable volume remained the sample was collected, using a Gilson micropipette (Section 2.1.7), and aliquoted for storage. Typical

protein recovery percentages varied according to the total amount of protein and the final concentration, usually 70-90% recovery for samples concentrated to 45mg/ml-10mg/ml, respectively.

2.3.7.2 Determining protein concentration

2.3.7.2.1 Bradford assay (Bradford et al., 1976)

A commercially available solution, 5X Protein Assay (Section 2.1.5), was purchased and used as described by the manufacturer for 250µL assays. Briefly, an intermediate stock solution of BSA (Section 2.1.5) at 2mg/ml in RO-purified water was prepared, from which four concentrations of BSA were prepared to be used as standard, with a linear range from 1mg/ml to 125µg/ml. Protein sample to determine concentration was next diluted to four unknown concentrations, related to each other by a 10-fold difference in concentration. All samples, standard and protein samples alike, were added 50µl of the commercial solution, diluting it down 5-fold (in 200ul sample). After gentle mixing, the absorbance of the samples was read at 595nm with UVettes® (Section 2.1.5). Measurements of the protein sample concentration that approximated those of the linear range of the standard curve were used to determine concentration by correlation.

2.3.7.2.2 Lowry assay (Lowry et al., 1951)

A commercially available kit for protein concentration determination was purchased from Pierce (Section 2.1.5) and manufacturer's instructions for low-volume assays followed. BSA standards and dilutions of the protein to determine concentration were prepared as described above (Section 2.3.7.2.1), after which the kit's reagent A was added to the samples (both standard and protein samples) and allowed to react for 10min at room temperature. After this, the kit's reagent B was added and allowed to react for 30min at room temperature. Measurements of

sample and standard's absorbance at 750nm were collected and used to determine concentration by correlation.

2.3.8 Protein storage

For purified protein, all storage procedures were identical. Protein aliquots were flash-frozen in liquid nitrogen and immediately stored at -80ºC. No refreezing of protein samples was performed in this thesis.

2.3.9 Protein buffer exchange

To exchange the solution a protein sample was in, with the best optimization of protein recovery (typically above 90%), time consumed (~30min) and efficient exchange of the solution, NAP^{TM} columns were used (Section 2.1.6.2). Two types of columns were used, NAP5 and NAP10 columns, respectively handling sample volumes of 100µl-500µl and 500µl-1000µl. Following the manufacturer's instructions, columns were equilibrated with 4 column volumes of the final buffer, and sample added to the top of the column following 30min centrifugation at 16000x g. After adsorption of the sample solution, buffer was added to make up the maximum volume allowed by the column (500µl for NAP5 and 1000µl for NAP10). Elution was performed by gravity-flow, adding buffer to the top of the column, following manufacturer's instructions. Elution volume was calculated by the equation:

$$V_E = 1.3333V_S + 366.67$$

Equation 2.1

, where V_E is the elution volume and V_S is the sample volume added.

2.3.10 Protein biochemical analysis

2.3.10.1 Mass spectrometry

To determine the mass of the proteins used in this thesis, electrospray ionization mass spectrometry (ESI-MS) measurements were performed, with the assistance of Dr. James Ault at the Mass Spectrometry facility of the Astbury Centre for Structural Molecular Biology, Leeds, UK. Protein sample was thawed and buffer exchanged (Section 2.3.8) to 20mM ammonium acetate buffer, prepared fresh prior to use by dissolving reagent salt (Section 2.1.7) and filter-sterilized (Section 2.1.7). pH of the buffer was measured at 6.6. Protein sample was next diluted to 10µM (estimated based on expected mass), sprayed on the mass spectrometer in positive ionization mode, and the time-of-flight measured. The resulting spectra was used to calculate the mass of the protein, according to the rationale: assuming that any two adjacent charge peaks in the series of multiply charged ions on the spectra differ by one charge but not in the molecule's mass, the molecular weight of a molecule can be calculated with the simultaneous equations:

$$m/z = (MW + n\text{H}^+)/n$$

Equation 2.2

and

$$m/z = [MW + (n+1)\text{H}^+] /(n+1)$$

Equation 2.3

, where m/z is the experimentally measured value (a composite of mass, m, and charge of the molecule, z), MW is the molecular weight of the molecule, n is the integer number of charges on the ions and H^+ is the mass increase of a protonated

ion (equals the mass of a proton, 1.008Da). Solving both equations in function of *n* will allow the value corresponding to MW on the m/z value to be extracted.

2.3.10.2 Circular dichroism (CD) spectroscopy

Circular dichroism spectra were acquired on an Olis spectropolarimeter (On line Instrument Systems, Inc.), equipped with a Peltier thermo-stated single cell holder (Quantum Northwest). Far-UV baseline corrected spectra were collected in triplicate, from 190 to 260nm at a scan rate of $10nm.min^{-1}$, using a bandwidth of 2.4nm with a 5 s/point signal averaging. Quartz cells with path lengths of 0.1 and 0.02cm were used. Protein concentrations ranged from 5 to 10µM, in 50mM phosphate buffer (pH 7.5), supplemented with 1mM EDTA (both components diluted from stock with RO-purified water). Experiments were performed at the University of Coimbra, with the assistance of Dr. Daniela Vaz from the *Structural and Computational Biology* group of the Centre for Neurosciences of Coimbra (CNC).

2.3.10.3 Deubiquitinase activity (Ub-AMC)

To assay the activity of ataxin-3 (and constructs thereof) as a deubiquitinase, thawed protein stock was diluted in storage buffer (Section 2.2.2.3) to an intermediate protein stock at 10-fold the intended test concentration. For each experiment, Ub-AMC stock in 100% DMSO (as purchased) was diluted in the test's cuvette, to a final concentration of 0.5µM in assay buffer (Section 2.2.2.5). A spectrofluorometer (Section 2.1.7), previously equilibrated at 37ºC, was used to measure the fluorescence of the Ub-AMC solution at excitation and emission wavelengths of 380 and 460nm, respectively, until a stable signal output was observed. At this point, the protein sample was added, diluting it 10-fold in the Ub-

AMC solution, gently mixed and the fluorescence of the sample measured at the wavelengths noted above. The total measuring time was 600 seconds.

2.3.11 Protein aggregation

2.3.11.1 *De novo* fibril formation

Four different methods were used to assay fibril formation of ataxin-3 and its constructs from monomeric protein, as detailed below. All protocols derive from the work conducted by LeVineIII (LeVineIII, 1999), and build on the protocol described for ataxin-3 by Gales and co-workers (Gales, 2005).

2.3.11.1.1 Protocol A

Initial aggregation assays were performed as described by Gales and co-workers (Gales, 2005). Frozen monomeric protein in storage buffer (Section 2.3.8) was thawed "in-hand" and diluted to the intended protein concentration with the same buffer, supplemented with Thioflavin T (ThioT) (Section 2.2.2.6) to a final test concentration of 10µM. The protein sample was next aliquoted into replicate wells on a FluoroNunc[TM] 96 well microplate with a clear lid, 100µl per well (Section 2.1.6.2), along with buffer blanks prepared from storage buffer with a final concentration of 10µM ThioT, as control. The microplate was next placed in a SpectraMax Gemini microplate reader (Section 2.1.6.2) and the fluorescence intensity of samples (including buffer blanks) measured with excitation and emission wavelengths of 440 and 480nm, respectively. After measurement by the top down mode, the microplate was covered with its clear lid and placed in an static incubator at 37ºC (Section 2.1.3). At regular intervals the microplate was taken from the incubator and fluorescence measurements taken as before, after

which the lid was replaced and the microplate returned to incubation at 37ºC, as before. For data visualization the single timepoint measurements of fluorescence intensity were then plotted against time. At the end of each experiment the sample was retrieved from the microplate wells and prepared for EM visualisation of the end-products formed, as described in Section 2.3.11.3.1.

2.3.11.1.2 Protocol B

Building on the previous protocol (protocol A, directly above), a new aggregation protocol was designed. Frozen monomeric protein samples (Section 2.3.8) were thawed "in'hand" and centrifugated at 16000x *g* for 30min at room temperature on a bench-top centrifuge (Section 2.1.6.2). Soluble sample collected after centrifugation was buffer exchanged to the required buffer, as described in Section 2.3.9. Buffer exchanged protein sample was recovered, quantified (as described in Section 2.3.7.2), and diluted to the appropriate protein concentration with the required buffer. ThioT was added from a 20mM stock prepared in RO-purified water (Section 2.2.1.15), to a final test concentration of 10μM. After gentle mixing, 100μl aliquots were dispensed into wells on a Corning 96 well microplate (Section 2.1.6.2). Alternatively, 50μl aliquots were dispensed into wells of a Corning 384 well microplate. Buffer blank samples were prepared with the tested buffer, supplemented with ThioT to a final concentration of 10μM, as above. These samples took up the outer wells of each microplate. The microplate was next sealed with a clear film (Section 2.1.6.2), allowing measurements to be collected *in situ* without evaporation of the sample, and placed in the microplate reader, previously equilibrated at 37ºC, for both incubation and measuring of intensity, by the top down mode. Excitation and emission wavelengths of 440nm and 480nm, respectively, were used. Measurements were collected once per hour, on BMG microplate readers (Section 2.1.6.2), each measurement cycle preceded by a short (1 sec) orbital agitation at 100rpm, designed to help resuspend any insoluble material that may have formed and settled by gravity. At the end of each experiment the microplate was allowed to cool to room temperature, and then

briefly spun (500x *g* for 2min at room temperature) to counter condensation. Sample from each test was taken and prepared as described in Section 2.3.11.3.21 for EM visualisation of the end-products formed.

2.3.11.1.3 Protocol C – Automated titration of initial protein concentration

Further modification of the aggregation protocol involved developing a protocol for a Biomek® 2000 series robot to set up the samples for aggregation assays titrating the initial protein concentation.

TXBa buffer solution was prepared as described in Section 2.2.2.7.5 and ThioT solution at 180µM concentration was prepared by diluting a 20mM concentrated ThioT stock (Section 2.2.1.15) in TXBa buffer. The robot was set up with clean reservoir wells, a clean Corning 96 well microplate (sample microplate), a clean Greiner 384 well microplate (mixing microplate) and a clean Corning 384 well microplate (assay microplate) (Section 2.1.6.2) as shown on Fig. 2.2 – Step 0. TXBa buffer and ThioT solution in TXBa buffer were dispensed onto reservoir wells and the program initiated (step 1, below).

While the robot followed the first 3 steps, frozen monomeric protein samples (Section 2.3.8) were thawed "in'hand" and centrifugated at 16000x *g* for 30min at room temperature on a bench-top centrifuge (Section 2.1.6.2). Soluble sample collected after centrifugation was next buffer exchanged to TXBa buffer using NAP columns, as described in Section 2.3.9. Buffer exchanged protein sample was recovered, quantified (as described in Section 2.3.7.2) and diluted with TXBa buffer to a 1.125-fold higher concentration than the highest protein concentration to test. Protein samples were next dispensed onto wells on the sample microplate (Fig. 2.3 – Step 4), 300µl per well, four wells *per* protein tested. The robot program was next unpaused and steps 4 to 7 followed as described below and shown on Fig. 2.3.

- Stepwise clarification of the automated program:

- **Step 0** – All microplates and clean reservoir wells are positioned in the robot. Buffer and ThioT solutions were dispensed manually into reservoir wells and specific wells on the sample microplate (Fig. 2.2).

- **Step 1** – 45µl of TXBa buffer were dispensed into all wells of the mixing microplate, except those reserved for the highest concentration of protein (Fig. 2.2).

- **Step 2** – Pipette tips were changed and 5µl of ThioT solution in TXBa buffer were dispensed into all wells of the assay microplate (Fig. 2.2).

- **Step 3** – Robot pauses automatically. Both 384 well plates were temporarily sealed with clear film, until the robot was unpaused (Fig. 2.2).

- **Step 4** – Protein sample was manually dispensed into the corresponding wells on the sample microplate. Both 384 well plates were unsealed and the program unpaused. Pipette tips were changed and 90µl protein sample were dispensed from the sample microplate into the wells reserved for the highest concentration of protein on the mixing microplate (Fig. 2.3).

- **Step 5** – With the same pipette tips, 45µl sample was retrieved from the highest concentration wells on the mixing microplate, dispensed into wells for the second highest concentration and gently mixed by pipetting up and down (Fig. 2.3).

- **Step 6** – The previous step was repeated, albeit starting from the second highest protein concentration and mixing the third highest concentration. This was then repeated for all other protein concentrations, always by two-fold dilutions of sample. The wells for the last concentration prepared end up with 90µl sample, whereas all others have 45µl (Fig. 2.3).

- **Step 7** – After changing pipette tips, 40ul from every well in the mixing microplate were transferred to matching wells on the assay microplate and gently mixed. Pipette tips were changed after sample from each column was transferred to the assay plate (Fig. 2.3).

Fig. 2.2 – Schematic representation of steps 0 through 3 of the automated mixing and dispensing protocol used to titrate the initial protein concentration. Steps followed are described in the text. Arrows next to the microplates indicate sample transfer. A total of 240 reactions are prepared per plate, establishing 8 protein concentrations for each of 3 monomeric protein samples.
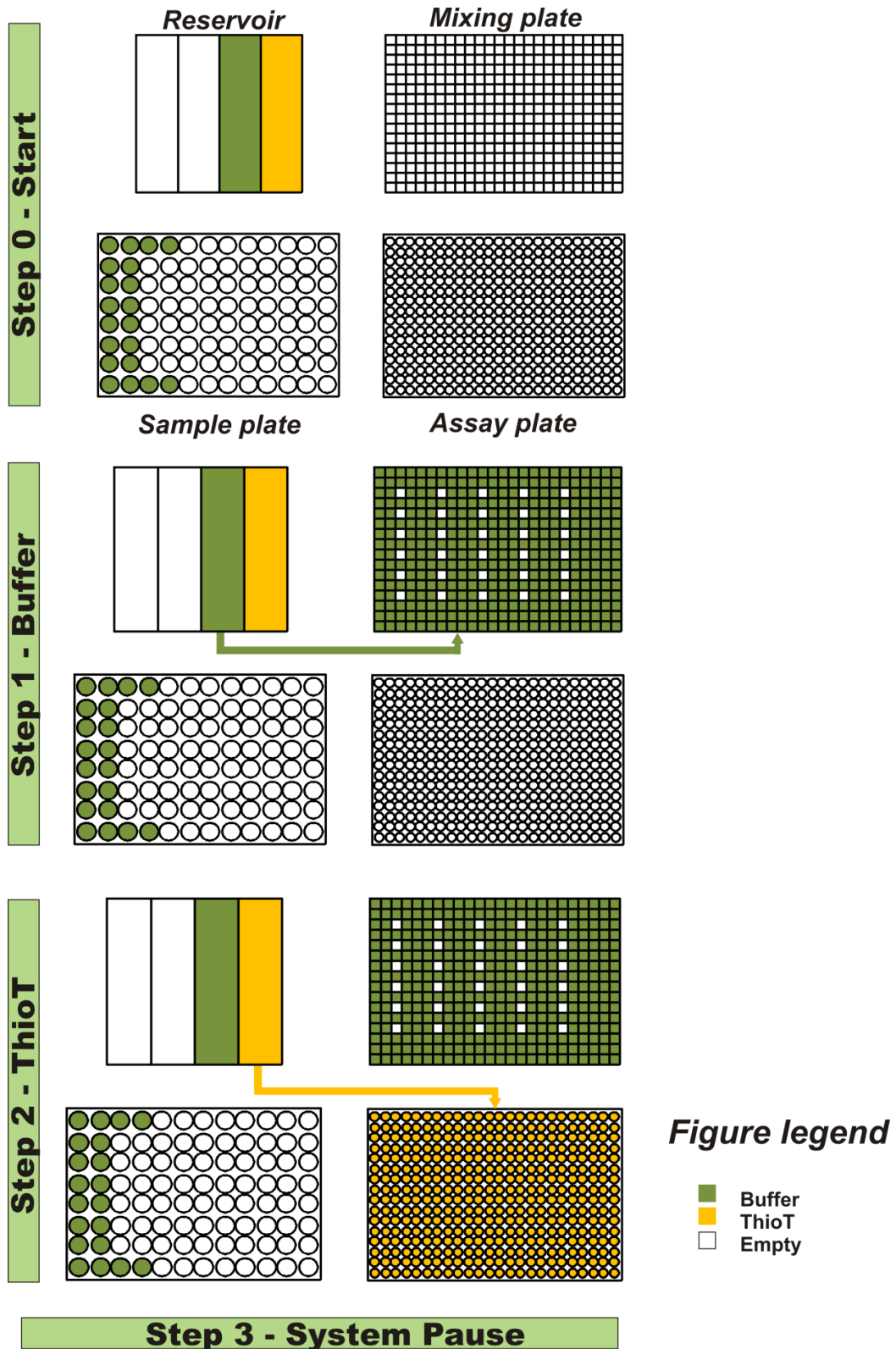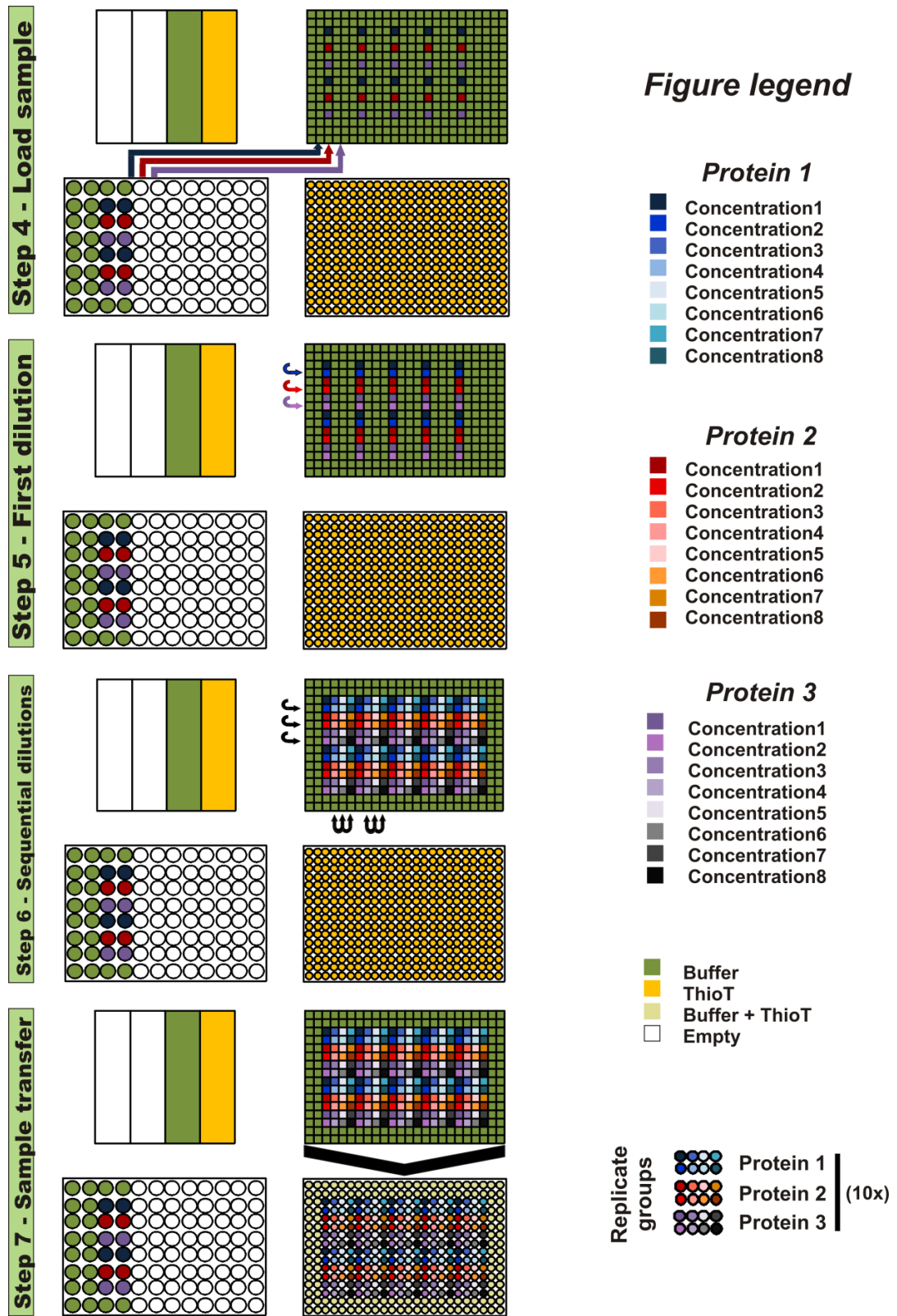
Fig. 2.3 – Schematic representation of steps 4 through 7 of the automated mixing and dispensing protocol used to titrate the initial protein concentration. Steps followed are described in the text. Arrows next to the microplates indicate sample transfer. A total of 240 reactions are prepared per plate, establishing 8 protein concentrations for each of 3 monomeric protein samples.

At the end of step 7, above, the assay microplate was retrieved from the robot, and briefly spun down at 500x *g* for 2min at 4ºC (Section 2.1.6.2) to remove air bubbles that may have formed, after which it was sealed with a clear film (Section 2.1.6.2), as before. The sealed microplate was next placed in a BMG Omega microplate reader (Section 2.1.6.2), previously equilibrated at 37ºC and equiped with excitation and emission filters for 440 and 480nm wavelengths, respectively. Measurements were taken every 6min, in a snaking top-to-bottom-to-top direction, each cycle preceded by a 1 sec orbital agitation at 100rpm. For direct visualization of aggregates formed, sample from microplate wells was used as described in Section 2.3.11.3.2 for EM imaging.

2.3.11.1.4 Protocol D – Automated preparation of ataxin-3 seeds

The robot was set up with clean reservoir wells, a clean Corning 96 well microplate (sample microplate) and a clean Corning 384 well microplate (assay microplate) (Section 2.1.6.2) as shown on Fig. 2.4 – Step 0. TXBa buffer solution was prepared as described in Section 2.2.2.7.5 and ThioT solution at 180µM concentration was prepared by diluting a 20mM concentrated ThioT stock (Section 2.2.1.15) in TXBa buffer. Frozen monomeric protein samples (Section 2.3.8) were thawed "in-hand" and centrifugated at 16000x *g* for 30min at room temperature on a bench-top centrifuge (Section 2.1.6.2). Soluble sample collected after centrifugation was next buffer exchanged to TXBa buffer using NAP columns, as described in Section 2.3.9. Buffer exchanged protein sample was recovered, quantified (as described in Section 2.3.7.2) and diluted with TXBa buffer to a 1.125-fold higher concentration than the intended test concentration. All solutions were next manually dispensed into their allocated positions, according to Fig. 2.4. Samples on the sample microplate (Fig. 2.4) were dispensed as 300µl per well, fourteen wells *per* protein tested. The robot program was next initiated and the steps described below were carried out.

- Stepwise clarification of the automated program:

- **Step 0** – All microplates and clean reservoir wells are positioned in the robot. ThioT solution, buffer solution and protein solution at the set concentration (as above) were dispensed manually into the corresponding wells (Fig. 2.4).

- **Step 1** – 5μl of ThioT solution in TXBa buffer were dispensed into all wells of the assay microplate (Fig. 2.4).

- **Step 2** – Pipette tips were changed and 40μl of sample and buffer were transferred from the sample plate to the assay plate and gently mixed (by pipetting up and down) (Fig. 2.4).

At the end of step 2, above, the assay microplate was retrieved from the robot, and briefly spun down at 500x *g* for 2min at 4ºC (Section 2.1.6.2) to remove air bubbles that may have formed, after which it was sealed with a clear film (Section 2.1.6.2), as before. The sealed microplate was next placed in a BMG Omega microplate reader (Section 2.1.6.2), previously equilibrated at 37ºC and equiped with excitation and emission filters for 440 and 480nm wavelengths, respectively. Measurements were taken every 6min, in a snaking top-to-bottom-to-top direction, each cycle preceeded by a 1 sec orbital agitation at 100rpm. At the end of the experiment, sample from all wells of a given protein was collected and pooled together, prior to storage at room temperature. For direct visualization of aggregates formed, sample from pooled aliquots was used as described in Section 2.3.11.3.2 for EM imaging.
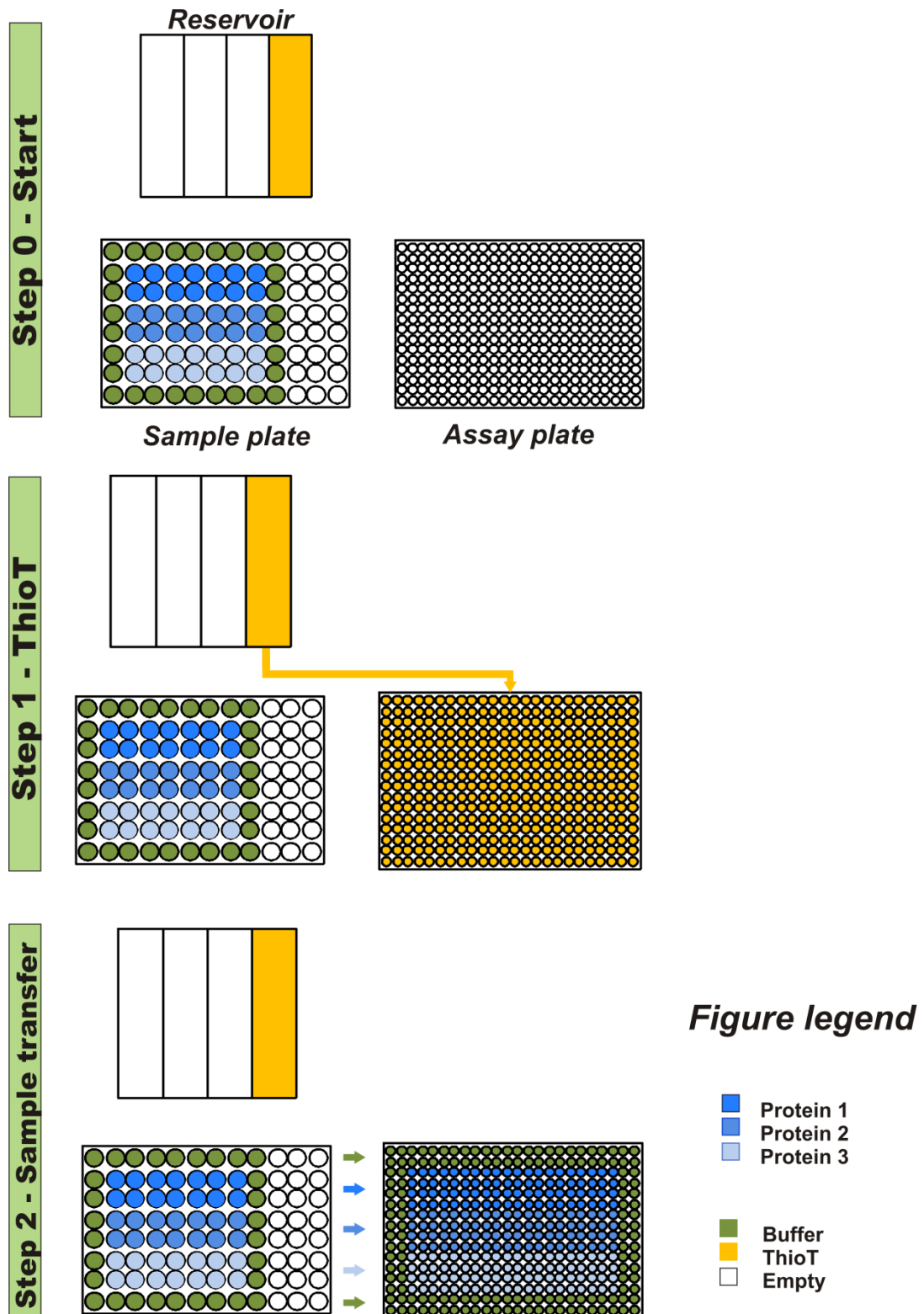
Fig. 2.4 – Schematic representation of the automated mixing and dispensing protocol used to establish growth conditions for fibrillar sample. Steps followed are described in the text. Arrows next to the microplates indicate sample transfer. A total of 240 reactions are prepared per plate, establishing a total of 3500µl of sample from each of 3 monomeric protein samples.

2.3.11.2 Seeded fibril formation

The robot was set up with clean reservoir wells, a clean Corning 96 well microplate (sample microplate), a clean Greiner 384 well microplate (mixing microplate) and a clean Corning 384 well microplate (assay microplate) (Section 2.1.6.2) as shown on Fig. 2.5 – Step 0. TXBa buffer solution was prepared as described in Section 2.2.2.7.5 and ThioT solution at 180µM concentration was prepared by diluting a 20mM concentrated ThioT stock (Section 2.2.1.15) in TXBa buffer. Frozen monomeric protein samples (Section 2.3.8) were thawed "in'hand" and centrifugated at 16000x *g* for 30min at room temperature on a bench-top centrifuge (Section 2.1.6.2). Soluble sample collected after centrifugation was next buffer exchanged to TXBa buffer using NAP columns, as described in Section 2.3.9. Buffer exchanged protein sample was recovered, quantified (as described in Section 2.3.7) and diluted with TXBa buffer to a 1.125-fold higher concentration than the intended test concentration. All solutions were next manually dispensed into their allocated positions, according to Fig. 2.5. Samples on the sample microplate were dispensed as 300µl per well, forty-two wells with protein solution and thirty with buffer solution (as shown on Fig. 2.5 – Step 0). Additionally, seed sample collected as described in Section 2.3.11.1.4 was diluted in the assay buffer to a 2.25-fold higher volumetric percentage than the highest volumetric percentage to be tested. These samples were next added to the sample microplate, 250µl per well, in a total of 1000µl per seed sample tested (as shown on Fig. 2.5). The robot program was next initiated and the steps described below were carried out.

- Stepwise clarification of the automated program:

- **Step 0** – All microplates and clean reservoir wells are positioned in the robot. Buffer, protein sample, seed sample and ThioT solutions were

dispensed manually into allocated wells on the sample microplate or reservoir, as shown on (Fig. 2.5).

- **Step 1** – 5µl of ThioT solution in TXBa buffer were dispensed into all wells of the assay microplate. Pipette tips were changed and 45µl TXBa buffer solutions and monomeric protein solutions were dispensed into corresponding wells of the mixing microplate (Fig. 2.5).

- **Step 2** – 45µl seed sample was next dispensed into the wells corresponding to the highest volumetric percentage of seed sample to be tested and gently mixed (by pipetting up and down), effectively diluting the seed sample 2-fold (Fig. 2.6).

- **Step 3** – With the same pipette tips, 45µl sample from the wells corresponding to the highest volumetric percentage of seed sample were transferred to the wells corresponding to the second highest volumetric percentage and mixed gently by pipetting up and down (Fig. 2.6).

- **Step 4** – The previous step was repeated, albeit starting from the second highest volumetric percentage of seed sample and mixing the third highest concentration. This was then repeated for all other protein concentrations, always by two-fold dilutions of sample. The wells for the last concentration prepared end up with 90µl sample, whereas all others have 45µl. A total of 7 2-fold dilutions of seed sample were prepared (including the initial dilution from seed stock in the sample microplate) (Fig. 2.6).

- **Step 5** – After changing pipette tips, 40ul from every well in the mixing microplate were transferred to matching wells on the assay microplate and gently mixed. Pipette tips were changed after sample from each column was transferred to the assay plate (Fig. 2.6).

At the end of step 5, above, the assay microplate was retrieved from the robot, and briefly spun down at 500x *g* for 2min at 4ºC (Section 2.1.6.2) to remove air bubbles that may have formed, after which it was sealed with a clear film (Section 2.1.6.2), as before. The sealed microplate was next placed in a BMG Omega microplate reader (Section 2.1.6.2), previously equilibrated at 37ºC and equipped with excitation and emission filters for 440 and 480nm wavelengths, respectively.

Measurements were taken every 6min, in a snaking top-to-bottom-to-top direction, each cycle preceded by a 1 sec orbital agitation at 100rpm. For direct visualization of aggregates formed, sample from microplate wells was used as described in Section 2.3.11.3.2 for EM imaging.

Fig. 2.5 – Schematic representation of steps 0 through 1 of the automated mixing and dispensing protocol used to establish a titration of volumetric percentages of seed sample added at the start of the reaction. Steps followed are described in the text. Arrows next to the microplates indicate sample transfer.

Fig. 2.6 – Schematic representation of steps 2 through 5 of the automated mixing and dispensing protocol used to establish a titration of volumetric percentages of seed sample added at the start of the reaction. Steps followed are described in the text. Arrows next to t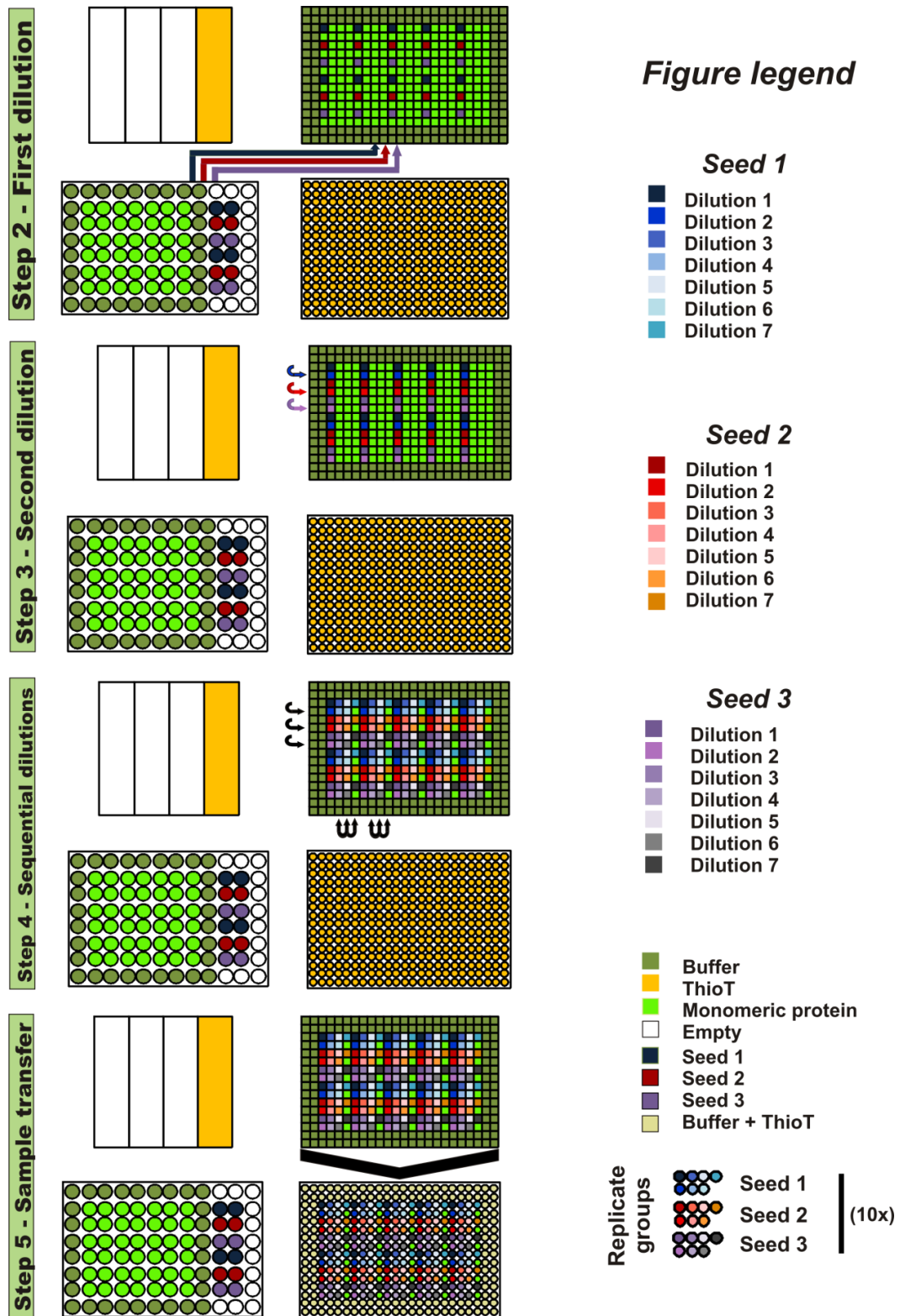he microplates indicate sample transfer. A total of 240 reactions are prepared per plate, establishing 7 volumetric percentages of each of 3 seed samples on one monomeric protein sample.

2.3.11.3 Electron microscopy (EM)

To visualise products formed after or during aggregation of protein samples, negative-stain electron microscopy grids were prepared according to one of two protocols, detailed below. When following the formation of aggregates by EM, sample grids were prepared at regular intervals. Sample grids were visualized on electron microscopes and images collected, as described below.

2.3.11.3.1 Protocol A

Freshly ionized formvar- and carbon-coated copper grids used in this protocol were obtained from the Histology and Electron Microscopy service at IBMC, Porto, Portugal. 10µl sample was placed on the coated side of the grids and left to react for 5min. Grids were then picked up and gently rinsed, coated side down, in 3 RO-purified water drops (placed on a sheet of parafilm) sequentially, after which they were dropped coated side down on a drop of 1% uranyl acetate solution (prepared by dissolution of uranyl acetate in water acidified with HCl, spun down for 1min at 16000x g prior to use) for 1min. Grids were then placed on filter paper, coated side up, to blot the uranyl solution prior to storage in a grid box. Images were retrieved on a Zeiss electron microscope fitted with a digital 10megapixel CCD camera, located at the Histology and Electron Microscopy service at IBMC, Porto, Portugal (Section 2.1.6.2).

2.3.11.3.2 Protocol B

Formvar- and carbon-coated copper grids used in this protocol were obtained from Agar scientific (Section 2.1.6.2) and irradiated for 1h under a 12W UV light (coated side up) immediately prior to use. 2µl sample was diluted 10-fold in RO-purified water and 10µl of the mix placed on the coated side of the grids and

left to react for 1min. Grids were then blotted with filter paper to remove excess sample and 10µl 4% uranyl acetate solution (prepared by dissolution of uranyl acetate in RO-purified water and spun down at 16000x g for 1min prior to use) were placed on the coated side of the grid and allowed to react for 1min, after which the excess uranyl acetate solution was blotted with filter paper. Grids were next placed on filter paper (coated side up) overnight, prior to storage in grid boxes. Images were retrieved on a Philips CM10 electron microscope, fitted with a film photographic camera, and located at the EM service of the Faculty of Biological Sciences, University of Leeds, UK (Section 2.1.6.2).

2.3.12 Data Processing

In order to simplify the comparison of collected data, as well as to further the understanding of the data collected, a set of mathematical approaches and their formulae were used in this thesis.

2.3.12.1 far-UV CD data normalization

2.3.12.1.1 Molar circular dichroism

Experimentally obtained circular dichroism (CD) data refers to the differential absorbance of left and right polarized light. To compare results at different protein concentrations, the results are normalized by the equation

$$\Delta A = (A_{left} - A_{right})/C \times l$$

Equation 2.4

, where ΔA is the molar circular dichroism, A_{left} and A_{right} are the absorbance of left and right polarized, respectively, C is the molar concentration of protein and I the path length (in cm).

2.3.12.1.2 Mean residue molar circular dichroism

From the normalized data, as above (Section 2.3.12.1.1), the results can be further normalized to the amino acid residue number of the protein, allowing a direct comparison of proteins with different lengths. In the above equation (Equation 2.4), the term C becomes C_{MR} (molar residue concentration of protein) by

$$C_{MR} = C \times N$$

Equation 2.5

, where N is the number of amino acid residues in the protein. Substituting on equation 2.4,

$$\Delta A = (A_{left} - A_{right})/C_{MR} \times I$$

Equation 2.6

2.3.12.2 Aggregation data normalization

2.3.12.2.1 Y-axis normalization

In order to directly compare results at different output ThioT intensity, a simple normalization to the highest intensity was applied on the data collected as per equation 2.7, where Y_{norm} corresponds to the normalized value for Y-axis coordinate, Y is the original value and Y_{max} and Y_{min} the highest and lowest original value of the series, respectively.

$$Y_{norm} = (Y - Y_{min})/(Y_{max} - Y_{min})$$

Equation 2.7

2.3.12.2.2 X-axis normalization

Due to occasional spread of results along the time dimension, to directly compare aggregation profiles another normalization method was performed, according to equation 2.8, where X_{norm} corresponds to the normalized value for the X-axis coordinate, X is the original value and X_{T50} the midpoint of the polymerization reaction, determined by a Y_{norm} coordinate of 0.5 on ThioT-monitored aggregation profiles after Y-axis normalization (Section 2.3.12.2.1).

$$X_{norm} = X - X_{T50}$$

Equation 2.8

2.3.12.3 Extracting aggregation-describing parameters

Aggregation data collected by monitoring the ThioT fluorescence at 480nm over time, was processed to extract aggregation-describing parameters, according to the methodology developed by Xue and co-workers (Xue, 2008). After normalizing the data as *per* Section 2.3.12.2.1, algorithms developed by the authors initially detect the upper and lower intensity baselines, corresponding mechanistically to the steady-state and lag phase baselines of a polymerization reaction (Fig. 2.7), and then determine the midpoint of the reaction (t_{50}) as the steepest point of the data between the baselines, corresponding mechanistically to the midpoint of the elongation phase. Tangents are fitted through the t_{50} and the best fit determined by the least square residuals method, (to minimize the errors carried in the approximation). The slope of this linear fit represents the apparent elongation rate of the reaction, *k*, and its intersect with the X axis marks the length of the lag phase (t_{lag}) (Fig. 2.7).
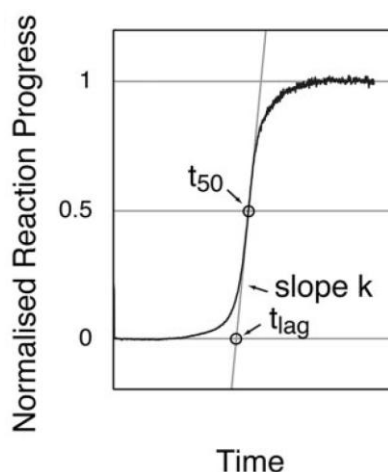
Fig. 2.7 – Schematic representation of parameter extraction by the method described by Xue and co-workers (Xue, 2008). t_{50} – midpoint of polymerization; slope k – elongation rate; t_{lag} – length of lag phase. Normalised reaction progress of 1 and 0 mark the upper and lower baseline of the data, respectively. Taken from Xue, 2008.

2.3.12.4 Estimating the thermodynamical nucleus size

Due to the short-lived nature of this species, its direct observation is currently technically unattainable. It is, however, possible to extrapolate the thermodynamical principles of polymerization and use the extracted aggregation-describing parameters (Section 2.3.12.3) to estimate the size of this important thermodynamical bottleneck of the polymerization reaction.

Two main approaches have been put forth to calculate the thermodynamical nucleus size, the Ferrone method and the Oosawa approach. Both work under the assumption that the polymerization rates do not change during the reaction (and consequently the nucleus size is constant) and that back reactions are negligible.

2.3.12.4.1 the Oosawa approach

The approach described by Oosawa and Asakura (Oosawa and Asakura, 1975), much like the method described above, relies on the direct relationship between the nucleation effects of the initial monomer concentration and the midpoint of reaction progress (t_{50}), both easily calculated parameters experimentally (see Section 2.3.12.3). In practical terms, linear fitting the log-log relationship between the reaction progress (t_{50}) and the initial monomer concentration (Fig. 2.8) yields a slope, *m*, that relates to the thermodynamical nucleus size, *n**, by the equation,
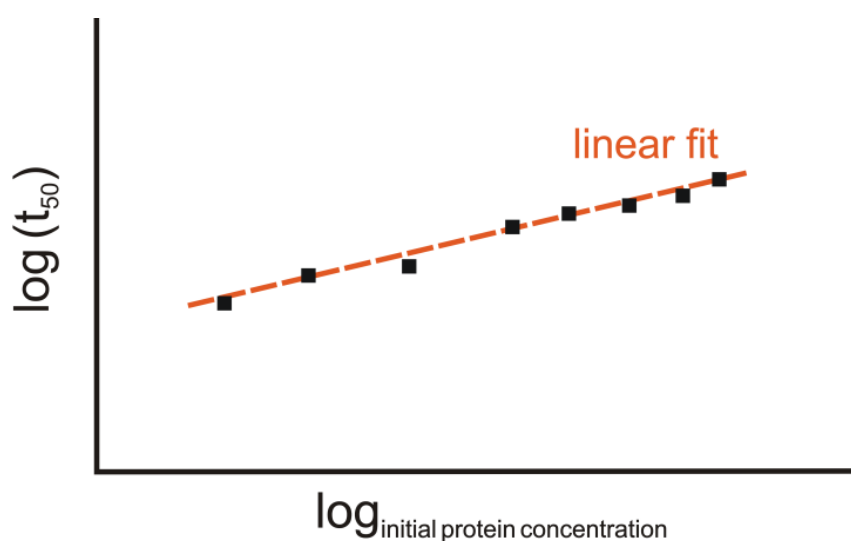
$$m = (n^* + 1)/2$$

Equation 2.9

Fig. 2.8 – Schematic representation of the graphical analysis to estimate the thermodynamical nucleus size according to the Oosawa method. Aggregation data is plotted as a double log plot relating the midpoint of the reaction, t50, and the initial protein concentration. The slope of the resulting linear fit is used in the calculations.

2.3.12.4.2 the Ferrone method

The method published by Ferrone in 1999 (Ferrone *et al.*, 1999) relies on relating the rates of nucleation (which determine the concentration of nuclei present at the start of the elongation phase) with the initial monomer concentration. Higher initial monomer concentration leads to faster nucleation rates, a consideration that is contra posed to the nucleus size, as higher sizes lead to slower nucleation rates. The author then describes a practical approach by which the initial polymerization rate (which should depend solely on the concentration of nuclei) is related to the initial monomer concentration in a log-log plot (Fig. 2.9), and a linear fit of the data shows a slope, *m*, that is directly correlated to the thermodynamical nucleus size, *n**, by the equation,
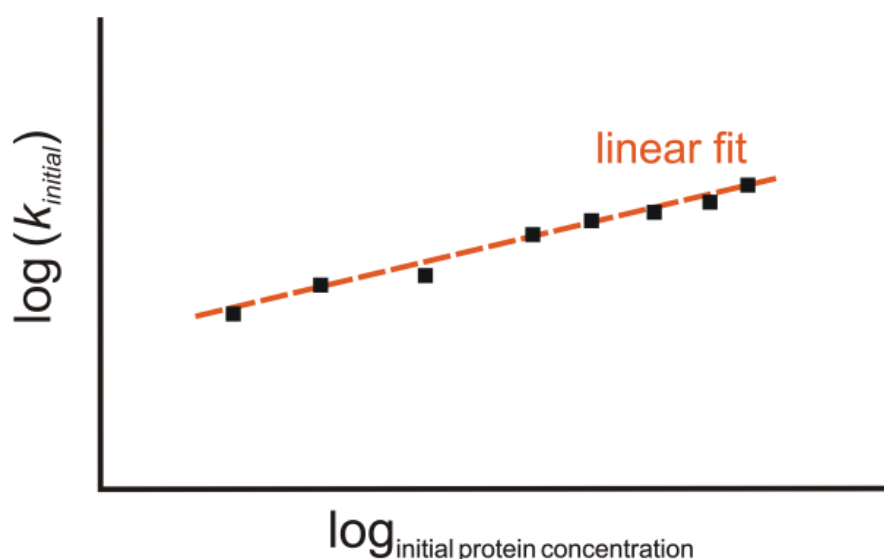
$$-m = n* + 2$$

Equation 2.10

Fig. 2.9 - Schematic representation of the graphical analysis to estimate the thermodynamical nucleus size according to the Ferrone method. Aggregation data is plotted as a double log plot relating the initial rate of polymerization, $k_{initial}$, and the initial protein concentration. The slope of the resulting linear fit is used in the calculations.
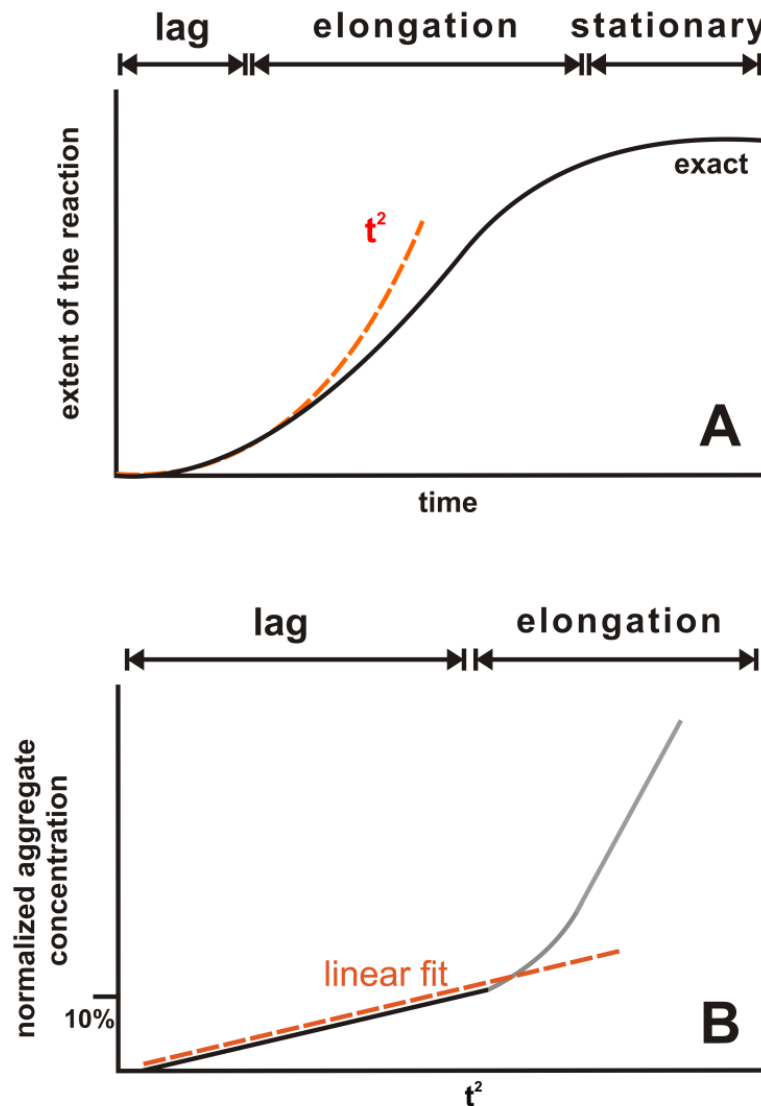
Fig. 2.10 – Schematic representation of the graphical analysis of aggregation data according to the Ferrone method. A) from the Y-axis normalized aggregation data, a second order polynomial is fitted (t2) to describe the initial polymerization rate. B) Alternatively, the aggregation data can be graphically described by a time squared X-axis (along with the normalized Y-axis) so that a linear regression can be fitted to the first 10% of the data. A) was adapted from Ferrone, 1999.

To calculate the initial elongation rate, the authors suggest fitting a first order polinomial function to the first 10% of the aggregation reaction in a plot relating the free monomer concentration (or its inverse, the concentration of monomer in aggregates) to the reaction progress (Fig. 2.10 - A). Alternatively, a linear fit can be applied to the data, if the log-log plot is constructed with time squared X-axis coordinates (Fig. 2.10 - B).

2.3.12.5 X-Ray fibre diffraction

Collection of X-Ray diffraction data of fibrillar samples from the aggregation of ataxin-3 was performed with the assistance of Dr. Sandra Macedo-Ribeiro (IBMC, Porto, Portugal), at the ESRF, Grenoble, France. Sample after ~100h quiescent incubation of 10μM AT3[U3/14Q] in storage buffer at 37ºC (as described in Section 2.3.11.1.2) was collected and pooled together. It was then transferred to a 0.5mm mounting loop, covered with a MicroRT capillary and allowed to air-dry for ca. 3 days. X-ray radiation from the electron beam source was directed to the sample and diffraction patterns collected with a 20 second exposure.

Chapter 3

3 – Characterizing AT3[U3/14Q]

3.1 Introduction

Ataxin-3, the product of the *ATXN3* gene, is a multidomain protein, comprising a globular N-terminal domain, the Josephin domain, a polyQ stretch and 2 or 3 ubiquitin interacting motifs (UIM), depending on the alternative splice isoform (Fig. 3.1). Both transcript variants have been found at the mRNA level in neuronal cells, although the 3UIM isoform seems to be predominantly translated (Harris, 2010). Nonetheless, a recombinant form of the 2UIM isoform has been the most studied form of ataxin-3 *in vitro*, perhaps because it was the first to be described (Kawaguchi, 1994). Importantly, recent studies show that the 3UIM isoform (Goto, 1997) is not only the prevalent expressed form in neurons but is also less aggregation-prone than the 2UIM isoform (Harris, 2010).

MJD1a isoform (2UIM)
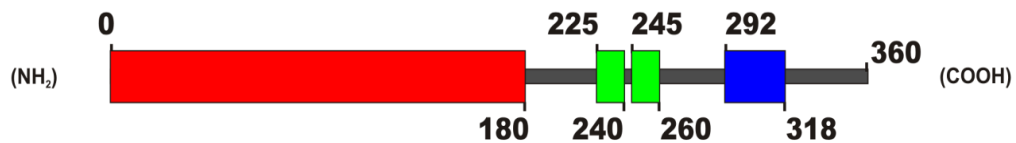
MJD1-1 isoform (3UIM)
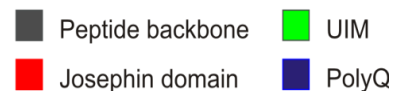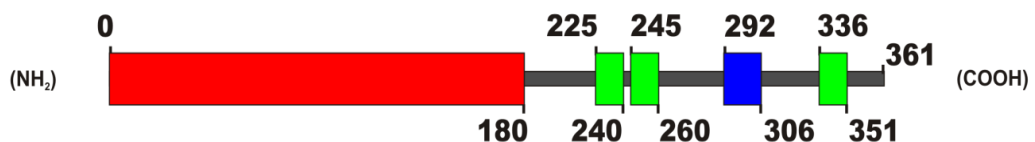
Fig. 3.1 – Schematic representation of the two full-length isoforms of ataxin-3 described. The 2UIM isoform was originally described in Kawaguchi, 1994. It contains two ubiquitin interacting motifs (in green), a polyQ tract (in blue) and a catalytical domain, Josephin (in red). It differs from the 3UIM isoform (Goto, 1997) in that it does not have a third UIM, C-terminal to the polyQ tract.

Notwithstanding the multidomain structure character of ataxin-3, the Josephin domain has been shown to dominate the secondary structure of the protein, giving it a highly α-helical character (Masino 2003; Chow, 2004; Gales, 2005) (Fig. 3.2). In addition, recent studies of the first two UIMs downstream from the Josephin domain (and common to both isoforms) have shown them to be α-helical, adding to the overall α-helical character of ataxin-3 (Song, 2010). On the other hand, the remainder of the 2UIM isoform protein, including the polyQ region, is thought to lack ordered secondary structure (Masino, 2003). In MJD/SCA3 pathology, it is this polyQ tract that is affected, with its length expanded due to a corresponding expansion of CAG repeats in the coding region of the affected gene (*ATXN3*).
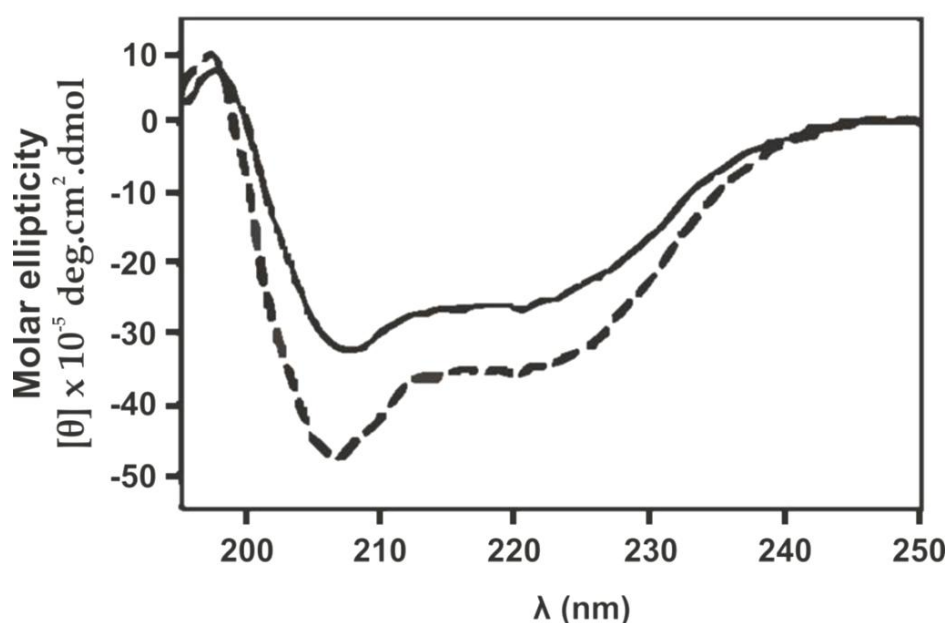
Fig. 3.2 – Far-UV CD measurements of 5µM ataxin-3 in 50mM Tris-HCl pH 8.0 at 25ºC. *Solid line* – full-length ataxin-3 (2UIM isoform) with 28 glutamines in the polyQ tract; *Dashed line* – Truncation mutant of ataxin-3 comprising the Josephin domain alone (residues 1-180 of ataxin-3). Taken from Chow, 2004.

Although the cellular role of ataxin-3 is not yet fully elucidated, the protein is proposed to act in ubiquitin pathways through its deubiquitinase (DUB) activity (Chai, 1999; Burnett, 2003; Doss-Pepe, 2003; Chai, 2004; Winborn, 2008). No major difference in enzymatic properties has been found between the 2UIM and 3UIM isoforms (Harris, 2010), consistent with the assignment of the catalytic residues to

the Josephin domain (Nicastro, 2004 and 2005; Mao, 2005) and the relative importance of each UIM in activity *in vitro* (Donaldson, 2003; Winborn, 2008; Nicastro, 2010) (see Section 1.5.2). Despite preferential binding and cleavage of long ubiquitin homopolymers (Winborn, 2008), *in vitro* studies of ataxin-3 have shown its activity against model DUB substrates with single ubiquitin moieties, such as Ubiquitin-AMC (Burnett, 2003) and GST-Ub$_{52}$ (Chow, 2006). The first of these substrates, Ubiquitin-AMC, comprises one ubiquitin molecule covalently bound to a 7-amido-4-methylcoumarin fluophore (AMC). Cleavage by DUB proteins occurs after the last glycine residue on ubiquitin and releases the fluophore from the protein's quenching effect and fluorescent light is emitted at 460nm wavelength (Dang, 1998). The second model substrate, GST-Ub$_{52}$, is a polyprotein construct comprising a glutathione-S-transferase protein moiety (GST), fused to the gene product of UbA52, a ubiquitin molecule with a C-terminal tail of 52 residues, structurally consistent with ribosomal protein genes (Baker, 1991). In this substrate the ubiquitin moiety sits in the middle of the construct (as opposed to an N-terminal position in Ubiquitin-AMC) and cleavage occurs at the last glycine residue of ubiquitin, releasing a GST-Ubiquitin fusion (which can be detected by SDS-PAGE for example).

As with other polyQ-containing proteins, the expression and purification of recombinant ataxin-3 shows a tendency towards spontaneous self-association, even with non-pathological primary sequences (Chow, 2006). Biochemical and biophysical characterization of the protein, however, requires monomeric protein as a starting point. Stringent methods were developed, therefore, in this thesis (Section 3.3) and in literature (Chow, 2006) to ensure this was the case. Interestingly, the self-association of full-length non-pathologically expanded ataxin-3 resembles, in at least its early stages, the aggregation of pathological protein *in vitro* and under physiologically-relevant conditions (Ellisdon, 2006), displaying nucleated kinetics and amyloid-like characteristics of the final products (Fig. 3.3). In fact, a recently proposed model for the aggregation of ataxin-3 posits a two-step aggregation pathway in which the region upstream of the polyQ tract of ataxin-3 self-associates in the initiation steps of assembly, with pathological expansion being

responsible solely for a secondary step that leads to a higher stability of the products (Ellisdon, 2006). Furthermore, evidence has been found for a co-localization of non-expanded ataxin-3 with intracellular aggregates of amyloid nature (Fujigasaki, 2000; Uchihara, 2001; Fujigasaki, 2001; Kettner, 2002). In this light it is clear that the study of non-expanded protein is crucial for the understanding of MJD/SCA3 pathology, particularly its initiation.
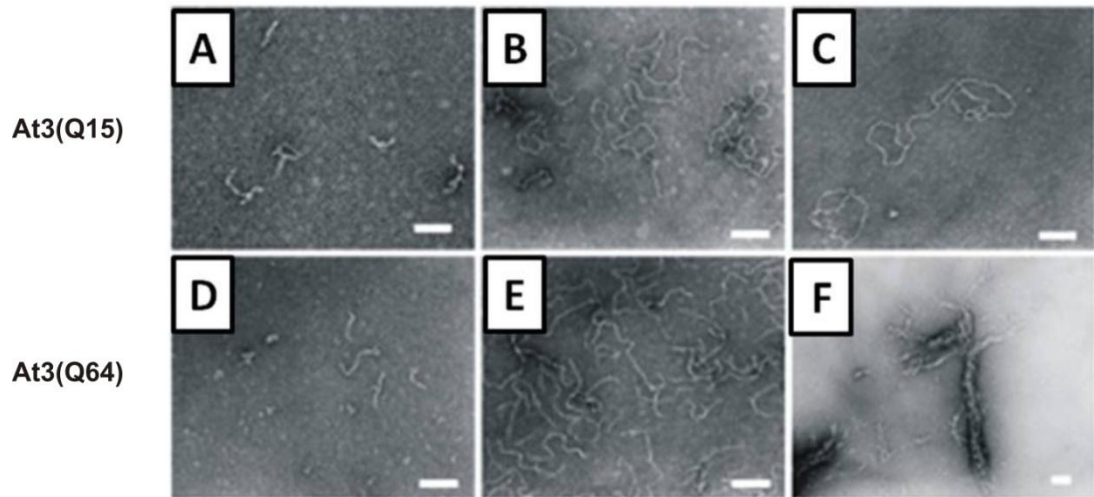
Fig. 3.3 – Negative-stain electron microscopy of aggregates formed during the aggregation of expanded (At3(Q64)) and non-expanded (At3(Q15)) ataxin-3 (2UIM isoform). A, B and C are representative of non-expanded ataxin-3 after aggregation for 11h, 48h and 100h, respectively. C, E and F are representative of aggregation of a pathologically expanded isoform of ataxin-3, with 64 glutamines, after 11h, 48h and 100h incubation, respectively. The aggregates formed by the expanded isoform in the early steps of aggregation (D and E) resemble those formed by the non-expanded isoform (A, B and C). Later steps display aggregates of different morphology, that are thicker and longer (F). Bars represent 100nm. Taken from Ellisdon, 2006.

The present chapter describes *in vitro* methods for the production of recombinant ataxin-3, displaying biochemical and biophysical characteristics consistent with its literature precedents (Burnett, 2003; Chow, 2004; Gales, 2005; Chow, 2006). This recombinant protein is then characterized further in subsequent chapters.

3.2 Protein preparations

3.2.1 Molecular biology

The 3UIM isoform of ataxin-3 was used in the experiments described in this thesis. Originally isolated by Goto and co-workers from patients without phenotypical manifestation of MJD/SCA3, the genetic material for this protein isoform was named pMJD1-1 by the authors, and deposited in the GenBank database (http://www.ncbi.nlm.nih.gov/genbank/) with accession reference U64820.1 (Goto, 1997). As previously mentioned in this thesis, this ataxin-3 isoform comprises an N-terminal Josephin domain, a polyQ stretch and 3UIMs (Fig. 3.4). The length of the polyQ stretch in this particular sequence is 14 residues long, specifically 13 glutamine residues intersected by one lysine. Although the relative importance of ataxin-3 splice isoforms in disease and non-disease states has not been fully elucidated to date, recent work has shown the 3UIM isoform to be the most commonly expressed in cells and brain tissues (Harris, 2010). The genetic sequence for this ataxin-3 isoform was cloned into pDEST17 expression vector (Gales, 2005) (Section 2.1.1.1) under the control of a T7 promoter, which was found to improve solubility and purification yields in previous work with ataxin-3 (Chow, 2006). The parent vector, pDEST17, adds an N-terminal tag of six histidines (6His) and a 12-residue linker to the start of the cloned sequence, in order to aid the purification strategies (Fig. 3.5, highlighted in yellow). The sequence of this linker region was mutated to incorporate a protease-specific cleavage site in the translated protein (Fig. 3.5, in yellow and underlined), allowing for removal of the tag and most of the linker with tobacco etch virus (TEV) protease, if required. This expression vector with ataxin-3's sequence, as described above, was sequenced (Section 2.3.2.2) to confirm its fidelity before being used further in this thesis. The result of sequencing is shown in Fig. 3.5, translated with bio-informatics tools

available on the ExPASy (**Ex**pert **P**rotein **A**nalysis **Sy**stem) proteomics server of the Swiss Institute of Bioinformatics (http://expasy.org).

```
ggggcggactggaggggggtggttcggcgtggggggccgttggctccagacaaataaacatg
 G  A  D  W  R  G  W  F  G  V  G  A  V  G  S  R  Q  I  N  M
gagtccatcttccacgagaaacaagaaggctcactttgtgctcaacattgcctgaataac
 E  S  I  F  H  E  K  Q  E  G  S  L  C  A  Q  H  C  L  N  N
ttattgcaaggagaatattttagccctgtggaattatcctcaattgcacatcagctggat
 L  L  Q  G  E  Y  F  S  P  V  E  L  S  S  I  A  H  Q  L  D
gaggaggagaggatgagaatggcagaaggaggagttactagtgaagattatcgcacgttt
 E  E  E  R  M  R  M  A  E  G  G  V  T  S  E  D  Y  R  T  F
ttacagcagccttctggaaatatggatgacagtggtttttttctctattcaggttataagc
 L  Q  Q  P  S  G  N  M  D  D  S  G  F  F  S  I  Q  V  I  S
aatgccttgaaagtttggggtttagaactaatcctgttcaacagtccagagtatcagagg
 N  A  L  K  V  W  G  L  E  L  I  L  F  N  S  P  E  Y  Q  R
ctcaggatcgatcctataaatgaaagatcatttatatgcaattataaggaacactggttt
 L  R  I  D  P  I  N  E  R  S  F  I  C  N  Y  K  E  H  W  F
acagttagaaaattaggaaaacagtggtttaacttgaattctctcttgacgggtccagaa
 T  V  R  K  L  G  K  Q  W  F  N  L  N  S  L  L  T  G  P  E
ttaatatcagatacatatcttgcacttttcttggctcaattacaacaggaaggttattct
 L  I  S  D  T  Y  L  A  L  F  L  A  Q  L  Q  Q  E  G  Y  S
atatttgtcgttaagggtgatctgccagattgcgaagctgaccaactcctgcagatgatt
 I  F  V  V  K  G  D  L  P  D  C  E  A  D  Q  L  L  Q  M  I
agggtccaacagatgcatcgaccaaaacttattggagaagaattagcacaactaaaagag
 R  V  Q  Q  M  H  R  P  K  L  I  G  E  E  L  A  Q  L  K  E
caaagagtccataaaacagacctggaacgagtgttagaagcaaatgatggctcaggaatg
 Q  R  V  H  K  T  D  L  E  R  V  L  E  A  N  D  G  S  G  M
ttagacgaagatgaggaggatttgcagagggctctggcactaagtcgccaagaaattgac
 L  D  E  D  E  E  D  L  Q  R  A  L  A  L  S  R  Q  E  I  D
atggaagatgaggaagcagatctccgcaggactattcagctaagtatgcaaggtagttcc
 M  E  D  E  E  A  D  L  R  R  T  I  Q  L  S  M  Q  G  S  S
agaaacatatccaagatatgacacagacatcaggtacaaatcttacttcagaagagctt
 R  N  I  S  Q  D  M  T  Q  T  S  G  T  N  L  T  S  E  E  L
cggaagagacgagaagcctactttgaaaaacagcagcaaaagcagcaacagcagcagcag
 R  K  R  R  E  A  Y  F  E  K  Q  Q  Q  K  Q  Q  Q  Q  Q  Q
cagcagcagcaggggggacctatcaggacagagttcacatccatgtgaaaggccagccacc
 Q  Q  Q  Q  G  D  L  S  G  Q  S  S  H  P  C  E  R  P  A  T
agttcaggagcacttgggagtgatctaggtgatgctatgagtgaagaagacatgcttcag
 S  S  G  A  L  G  S  D  L  G  D  A  M  S  E  E  D  M  L  Q
gcagctgtgaccatgtctttagaaactgtcagaaatgatttgaaaacagaaggaaaaaaa
 A  A  V  T  M  S  L  E  T  V  R  N  D  L  K  T  E  G  K  K
taa
 *
```

Fig. 3.4 – Nucleotide sequence (and the translated amino acid residue sequence) deposited in the GenBank database by Goto and co-workers, with accession reference U64820.1 and corresponding to the 3UIM isoform of ataxin-3 (Goto, 1997). The nucleotide sequence deposited was translated using the tools available on the ExPASy proteomics server (http://expasy.org). The protein's first residue is boxed in black, and is also the first residue of the N-terminal Josephin domain (highlighted in red), which is followed downstream by two UIMs, a polyQ tract of 13Gln residues and one Lys (highlighted in blue), in turn followed by a third UIM. All UIMs are highlighted in green and follow the work by Donaldson and co-workers (Donaldson, 2003).

```
atgtcgtactaccatcaccatcaccatcacctcgaaaacctgtattttcagtctaaagca
 M  S  Y  Y  H  H  H  H  H  H  L  E  N  L  Y  F  Q↓ S  K  A
ggctggatggagtccatcttccacgagaaacaagaaaggctcactttgtgctcaacattgc
 G  W  M  E  S  I  F  H  E  K  Q  E  G  S  L  C  A  Q  H  C
ctgaataacttattgcaaggagaatattttagccctgtggaattatcctcaattgcacat
 L  N  N  L  L  Q  G  E  Y  F  S  P  V  E  L  S  S  I  A  H
cagctggatgaggaggagaggatgagaatggcagaaggaggagttactagtgaagattat
 Q  L  D  E  E  R  M  R  M  A  E  G  G  V  T  S  E  D  Y
cgcacgtttttacagcagccttctggaaatatggatgacagtggttttttctctattcag
 R  T  F  L  Q  Q  P  S  G  N  M  D  D  S  G  F  F  S  I  Q
gttataagcaatgccttgaaagtttggggttttagaactaatcctgttcaacagtccagag
 V  I  S  N  A  L  K  V  W  G  L  E  L  I  L  F  N  S  P  E
tatcagaggctcaggatcgatcctataaatgaaagatcatttatatgcaattataaggaa
 Y  Q  R  L  R  I  D  P  I  N  E  R  S  F  I  C  N  Y  K  E
cactggtttacagttagaaaattaggaaaacagtggtttaacttgaattctctcttgacg
 H  W  F  T  V  R  K  L  G  K  Q  W  F  N  L  N  S  L  L  T
ggtccagaattaatatcagatacatatcttgcacttttcttggctcaattacaacaggaa
 G  P  E  L  I  S  D  T  Y  L  A  L  F  L  A  Q  L  Q  Q  E
ggttattctatatttgtcgttaagggtgatctgccagattgcgaagctgaccaactcctg
 G  Y  S  I  F  V  V  K  G  D  L  P  D  C  E  A  D  Q  L  L
cagatgattagggtccaacagatgcatcgaccaaaacttattggagaagaattagcacaa
 Q  M  I  R  V  Q  Q  M  H  R  P  K  L  I  G  E  E  L  A  Q
ctaaaagagcaaagagtccataaaacagacctggaacgagtgttagaagcaaatgatggc
 L  K  E  Q  R  V  H  K  T  D  L  E  R  V  L  E  A  N  D  G
tcaggaatgttagacgaagatgaggaggatttgcagagggctctggcactaagtcgccaa
 S  G  M  L  D  E  D  E  E  D  L  Q  R  A  L  A  L  S  R  Q
gaaattgacatggaagatgaggaagcagatctccgcagggctattcagctaagtatgcaa
 E  I  D  M  E  D  E  E  A  D  L  R  R  A  I  Q  L  S  M  Q
ggtagttccagaaacatatctcaagatatgacacagacatcaggtacaaatcttacttca
 G  S  S  R  N  I  S  Q  D  M  T  Q  T  S  G  T  N  L  T  S
gaagagcttcggaagagacgagaagcctactttgaaaaacagcagcaaaagcagcaacag
 E  E  L  R  K  R  R  E  A  Y  F  E  K  Q  Q  Q  K  Q  Q  Q
cagcagcagcagcagcagcagggggacctatcaggacagagttcacatccatgtgaaagg
 Q  Q  Q  Q  Q  Q  Q  G  D  L  S  G  Q  S  S  H  P  C  E  R
ccagccaccagttcaggagcacttgggagtgatctaggtgatgctatgagtgaagaagac
 P  A  T  S  S  G  A  L  G  S  D  L  G  D  A  M  S  E  E  D
atgcttcaggcagctgtgaccatgtctttagaaactgtcagaaatgatttgaaaacagaa
 M  L  Q  A  A  V  T  M  S  L  E  T  V  R  N  D  L  K  T  E
ggaaaaaaataa
 G  K  K  *
```

Fig. 3.5 – Nucleotide sequence and the translated amino acid residue sequence for the 3UIM isoform of full-length non-expanded ataxin-3 studied in this thesis. The protein contains an N-terminal Josephin domain (highlighted in red), three ubiquitin interacting motifs (highlighted in green) and a polyglutamine tract (highlighted in blue). To aid in purification a hexahistidine tag and a linker were added to the N-terminus (highlighted in yellow), and a tobacco etch virus (TEV) protease recognition site engineered at the end. Cleavage site is noted by a black arrow on the figure. The nucleotide sequence, as obtained from sequencing, was translated using the tools available on the ExPASy proteomics server (http://expasy.org). The UIM sequences highlighted follow the work by Donaldson and co-workers (Donaldson, 2003).

3.2.2 Expression and purification of ataxin-3

The procedures used in this thesis for the expression and purification of ataxin-3 follow previous reports by Gales and co-workers (Gales, 2005), with typical yields ranging from 5 to 8mg of purified monomeric protein *per* liter of cell culture grown.

3.2.2.1 Expression procedures

E. coli BL 21 SI cells were transformed with the pDEST17 – MJD 1N expression vector described above (and in Section 2.1.1.1). This cell line allows the induction of protein expression in bacterial cells by sodium chloride (NaCl), shown to result in increased protein solubility (Bhandari, 1997). The cell line is sensitive to the addition of NaCl and, once the salt is detected, expression of proteins is induced under the control of the T7 promoter. An added bonus, or possibly a contributing factor to the increased solubility reported, is the tight control over the induction of expression using this system. The expression of ataxin-3 before induction (expression leakage) was never seen throughout this thesis. To further enhance ataxin-3 solubility, the expression temperature after induction was lowered to 30ºC and the length of the induction step limited to three hours, consistent with reports by previous authors (Gales, 2005 and Chow, 2006). The repeated addition of 0.2% (w/v) glucose to cell cultures (as described in Section 2.3.5.2) further increased the yield of soluble ataxin-3.

Cells transformed with the expression vector were picked from solid media and starter cultures were initiated. From these, larger expression cultures were prepared and expression induced by the addition of a final concentration of 300mM NaCl, as detailed in Section 2.3.5.2. After 3h induction, the cells were harvested by

centrifugation and ressuspended in imidazole buffer A, supplemented with lysozyme, prior to storage at -20ºC (Section 2.3.5.2).

Soluble expression of ataxin-3 using the procedures described was verified by SDS-PAGE (Section 2.3.4), as shown in Fig. 3.6. Aliquots of cells from three batches expressing recombinant non-expanded ataxin-3 were collected for analysis, before (t_0) or after 3h induction. The latter samples were lysed and separated by centrifugation into supernatant (Sol) and pelleted sample (Ins) and this last one ressuspended with buffer A (Section 2.2.2.1) in equal volume to the soluble fraction. Samples were next prepared for SDS-PAGE analysis, as described in Section 2.3.4). Despite multiple efforts and modifications of the protocol, some of the protein expressed remains insoluble, an observation that is consistent with published data (Chow, 2006). Nonetheless, the soluble expression of ataxin-3 was reproducibly of good yield (Fig. 3.6).
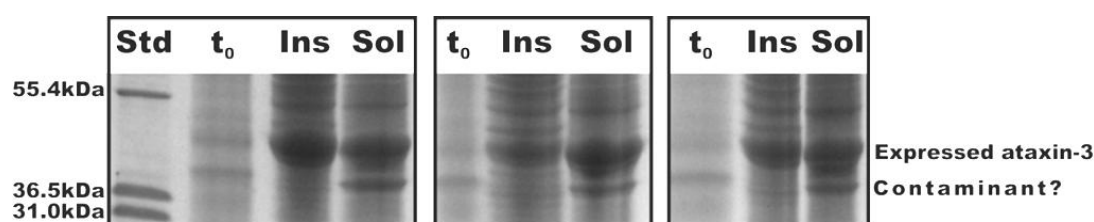
Fig. 3.6 – SDS-PAGE analysis of bacterial expression. Whole cell lysates, prior to induction (t_0) and soluble (Sol) and insoluble (Ins) fractions after 3h induction were analysed. Soluble protein expression was consistently of good level, despite a high percentage of insoluble protein. A soluble contaminant of lower molecular weight is visible in the samples.

3.2.2.2 Purification procedures

Mirroring the solubility issues shown above (Fig. 3.6), spontaneous self-association of recombinant full-length non-expanded ataxin-3 during the purification procedures was commonly encountered, as reported by multiple authors (such as Gales, 2005 and Chow, 2006). To tackle this problem and to minimize protein aggregation, the purification of ataxin-3 was performed at 4ºC,

except for the affinity chromatography steps described below, which were performed at room temperature. All purification steps were performed continuously, minimizing the length of sample preparation to purity. Furthermore, after affinity chromatography, all purification steps include the addition of DTT in the buffers used, to prevent inter or intramolecular disulfide bond formation (as reported by Chow, 2006). EDTA, to scavenge divalent cations that might have co-purified and increase ataxin-3 aggregation (as proposed by Richelli and co-workers, 2007) and glycerol, to increase the stability of ataxin-3, were also used, following the guidelines established both by Gales and co-workers (Gales, 2005), and Chow and co-workers (Chow, 2006). HEPES was the buffering molecule chosen for the later steps of purification, while the earlier steps employed phosphate buffers (Section 2.3.6).

Initial purification steps of ataxin-3 included proteinase inhibitors, though unlike Gales and co-workers (Gales, 2005), a designed cocktail of inhibitors comprising Bestatin (inhibitor of aminopeptidases, a class of metallo-proteases), pepstatin (inhibitor of aspartic proteases), leupeptin and PMSF (both serine protease inhibitors, though the first can also inhibit certain cysteine proteases) (*in* The complete guide for protease inihibition (Roche)) was used, in detriment of a complete inhibitor cocktail (Complete EDTA-free, Roche Applied Science). Consistent with the reported susceptibility of ataxin-3 to serine protease cleavage (Chow, 2006), PMSF was found to be sufficient to avoid significant cleavage during purification, and the remaining components of the cocktail were later discontinued in use.

3.2.2.2.1 Cell lysis and Affinity chromatography

Ressuspended cell pellets stored at -20ºC were thawed and incubated with DNAse, $MgCl_2$ and protease inhibitors for 90min at 4ºC with mild agitation (as detailed in Section 2.3.6.1), after which they were centrifugated and the soluble fraction collected (as described in Section 2.3.6.1). SDS-PAGE analysis was conducted to verify expression levels (as shown in Fig. 3.6).

The soluble fraction was next enriched in ataxin-3 protein by exploiting the affinity of histidine residues, such as those in the protein's tag, for nickel ions. Briefly, the soluble lysate fraction collected was loaded onto an iMAC FF column (Section 2.3.6.2) freshly charged with nickel ions and equilibrated with imidazole buffer A (Section 2.2.2.1), as described by the manufacturer (and detailed in Section 2.3.6.2). Extensive washing of the column with buffer A removed unbound proteins prior to elution. Protein still bound to the column after the washing step was eluted in a stepwise gradient of imidazole concentration, as shown on Fig. 3.7, where the first step removes weakly and unspecificaly-bound proteins and the last step restores the column's binding surfaces by eluting tightly-bound proteins. SDS-PAGE analysis of the eluted fractions (Fig. 3.7 – insets) shows high purity of the sample eluted at 100mM imidazole (the second step, Fig. 3.7).
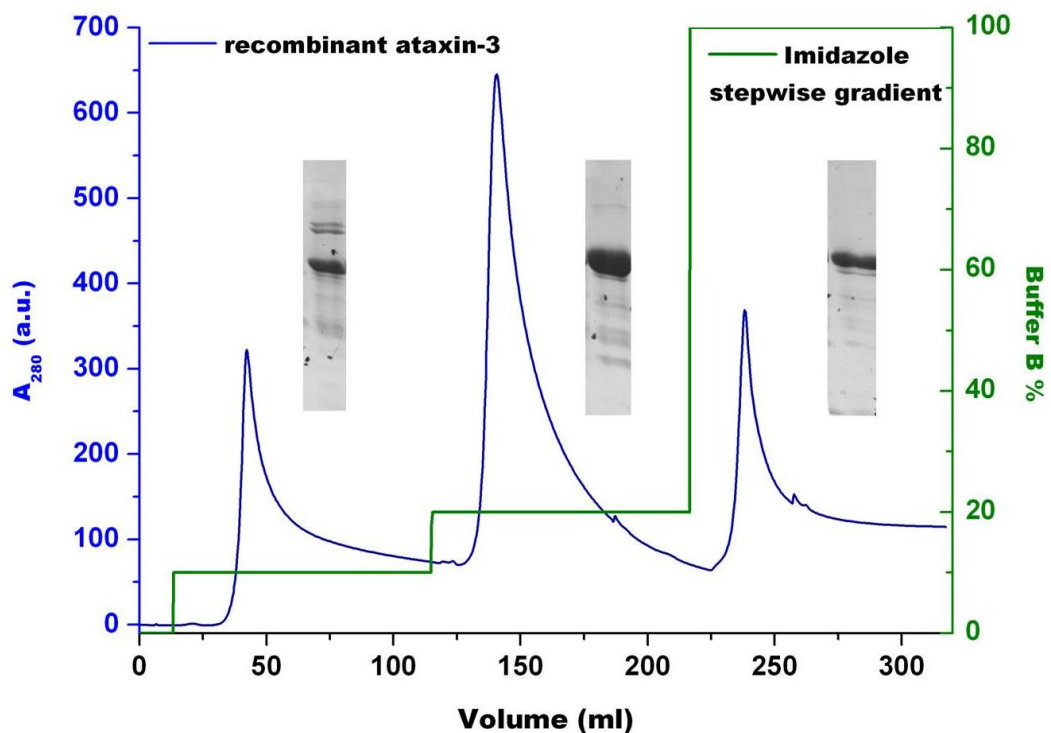
Fig 3.7 – Affinity purification of recombinant full-length non-expanded ataxin-3 (3UIM isoform). Elution was performed in a stepwise gradient to 500mM Imidazole (buffer B, in green) and monitored by absorbance at 280nm (in blue). Fractions were collected manually and run on an SDS-PAGE gel (insets) to judge purity.

3.2.2.2.2 Size-exclusion chromatography

As reported by previous authors, size-exclusion chromatography is key to obtaining soluble monomeric ataxin-3 (Gales, 2005 and later Chow, 2006)(Fig. 3.8).
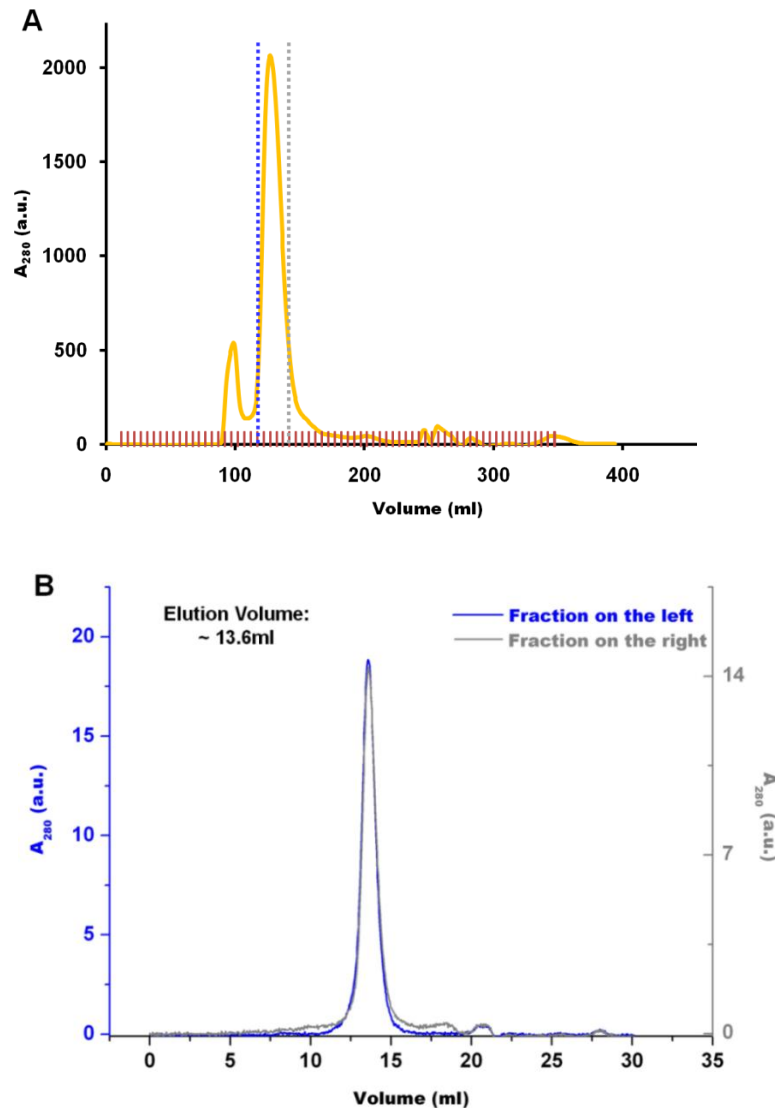
Fig. 3.8 - Size-exclusion chromatography of the purification of monomeric recombinant full-length non-expanded ataxin-3 (3UIM isoform). A) Preparative size-exclusion chromatogram. Shown in orange is the protein eluted (as monitored by the absorbance at 280nm) and in red the start of each 5ml fraction collected. Fractions on the elution front and tail of the peak are highlighted with a blue or a grey dotted line, respectively. B) Analytical size-exclusion chromatograms of fractions highlighted in (A). Monitored absorbance at 280nm is shown in blue and grey for fractions on the elution front and elution tail of the previous peak, respectively. The elution volume of monomeric protein, centred on 13.6ml, is annotated on the top left of the figure.

Recombinant ataxin-3 eluted from the nickel affinity purification step with 100mM imidazole (Fig. 3.7, second step) was next loaded on a preparative-grade size-exclusion chromatography column, previously equilibrated with storage buffer (Section 2.3.6.3), and eluted (Fig. 3.8 - A) under a slow isocratic flow rate, as detailed in Section 2.3.6.3.

Fractions of 5ml were collected. To confirm the purity and integrity of the final product, selected fractions were analysed by analytical-grade size-exclusion chromatography. Briefly, 100µl aliquots from each selected fraction were injected into a Superdex 200 column, previously equilibrated with storage buffer (Section 2.2.2.3), and eluted under an isocratic flow rate of 0.5ml/min (as detailed in Section 2.3.6.4). Fractions displaying an elution profile suggesting a degree of purity above 99% (Fig. 3.8 - B) were pooled and the remaining fractions were discarded. The elution volume of monomeric recombinant full-length non-expanded ataxin-3 was consistently centred at 13.6ml on the analytical size-exclusion step (Fig. 3.8 - B).

3.2.3 Concentration and storage procedures

Pooled fractions of highly pure monomeric recombinant full-length non-expanded ataxin-3 (as described above) were concentrated using Amicon centrifugal filters, previously equilibrated with storage buffer according to the manufacturer's instructions (detailed in Section 2.3.7.1). Sample was concentrated to 15-30mg of protein *per* ml, as calculated using the Lowry or Bradford method (Section 2.3.7.2). Concentrated protein was next aliquoted into microfuge tubes and flash-frozen by submersion in liquid nitrogen, prior to storage at -80ºC.

To ensure the monomeric state and integrity of purified ataxin-3 before and after storage, freshly concentrated or thawed samples were analysed by analytical size-exclusion chromatography, as described above, although only 25µl of concentrated sample were loaded. Elution was performed under isocratic flow rate, as before (detailed in Section 2.3.6.4). No visible change in the elution profiles was

uncovered (Fig. 3.9), with thawed sample displaying high purity, suggesting that no protein aggregation occurs during the concentration and storage procedures.
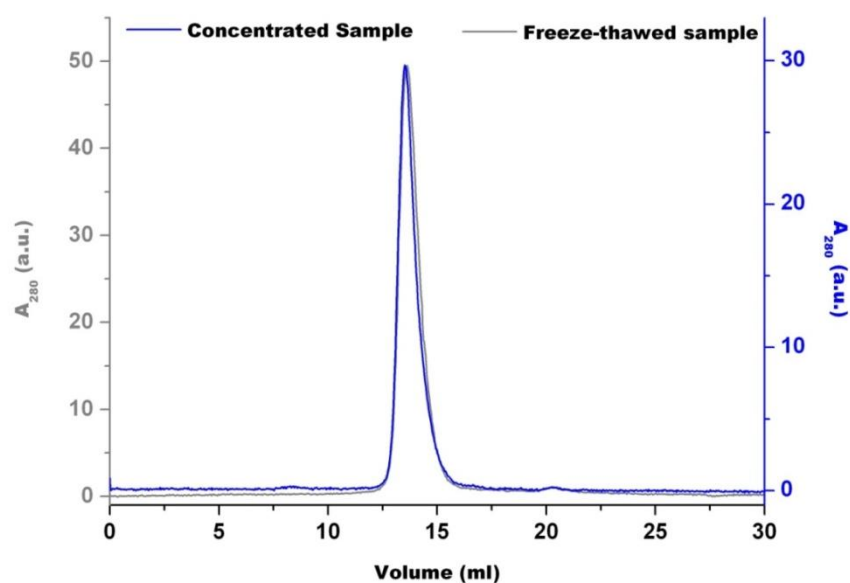
Fig. 3.9 – Analytical size-exclusion profile of freshly concentrated and freeze-thawed recombinant full-length non-expanded ataxin-3 (3UIM isoform). The elution profile of the protein, as monitored by absorbance at 280nm, after concentration and after a freeze-thaw cycle is shown in blue and grey, respectively.

3.3 Quality control

To assess the structural and functional fidelity of the protein purified above, biochemical and biophysical experiments were used to determine the properties of the monomeric protein. These were then compared with the expected parameters determined theoretically or experimentally.

3.3.1 Primary and secondary structure

3.3.1.1 Protein mass as a measure of primary structure

Modulation of the mass of recombinant proteins can occur by aberrant chemical modifications and, although for protein produced in bacterial systems these are not common, it was important to ensure the chemical fidelity of all preparations of recombinant protein used in this thesis, especially given the sensitivity of aggregation rates and mechanisms to covalent modifications, such as suggested by Mueller and co-workers for ataxin-3 (Mueller, 2009), and reported for other proteins such as relaxin or nucleosome core particles (Li, 1995; Yang, 2009). Several tools to aid a theoretical calculation of mass are available online, such as those offered on the ExPASy server (Swiss Institute of Bioinformatics, located at http://expasy.org). Assuming (as an initial approach) that no chemical modifications occur, the theoretically calculated mass of the sequence shown in Fig. 3.5 is 44055.12Da.

To experimentally determine the mass of the ataxin-3 expressed and purified in Section 3.2, electrospray ionisation mass spectrometry (ESI-MS) was chosen for its high accuracy. With the assistance of Dr. James Ault (University of Leeds, Mass Spectrometry Facility), protein sample in volatile buffer was analyzed at 20μM concentration (as described in Section 2.3.10.1) and the mass calculated (as detailed in Section 2.3.10.1) to be 43923.4Da (+/-0.63Da), slightly lower than the theoretically calculated mass of 44055.12Da, specifically 131.72Da lower. While this mass shift is within the expected error for the technique (0.1%), it's possible significance was still considered. Analysing the buffer peaks on the spectra did not indicate adducts of similar mass, suggesting the absence of a salt adduct from the buffer or the preparation steps for the protein. The analysis of the mass change assisted by databases dedicated to mass spectrometry analysis, namely that of the

Association of Biomolecular Resource Facilities (available at http://www.abrf.org), suggested several possibilities, such as a leucine or isoleucine residue missing. Other explanations are also possible, such as the observed mass shift being the result of more than one chemical event, making it harder to characterize based solely on the data collected. In any case, the mass measurement for the purified recombinant ataxin-3 is consistent, within experimental error, with the expected mass.

3.3.1.2 Secondary structure

Several authors have reported a highly α-helical character for ataxin-3, based on far-UV CD measurements (Masino 2003; Chow, 2004; Gales, 2005; Ellisdon, 2006). A high resolution structure of the full-length protein has not been determined, but work with a truncation mutant comprising the Josephin domain has confirmed its globular fold and α-helical character (Fig. 3.2). Nuclear magnetic resonance measurements (Masino, 2003) and analytical ultracentrifugation (Bevivino, 2001) of full-length ataxin-3, together with far-UV CD measurements of the Josephin domain truncation mutant (Masino, 2003; Chow, 2004) have indicated an elongated shape for the full-length protein, with the Josephin domain dominating the secondary structure content of ataxin-3. The remaining C-terminal portion of the protein, in which all differences between splice isoforms and pathological versus non-pathological variants localize, is largely unstructured (Masino, 2003). Accordingly, the secondary structure measurements, by far-UV CD, reports primarily on the fold of the N-terminal Josephin domain that is common to different splice variants. Analysing the purified protein's secondary structure establishes, by comparison with published data, the fidelity of the starting material for experimental assays. For this purpose, frozen protein stock was thawed and buffer exchanged into phosphate buffer at pH 7.5, supplemented with EDTA (Section 2.3.10.2). The sample was next diluted to 10µM final concentration, and

far-UV CD spectra were collected between 190nm and 260nm at 37ºC (as detailed in Section 2.3.10.2).
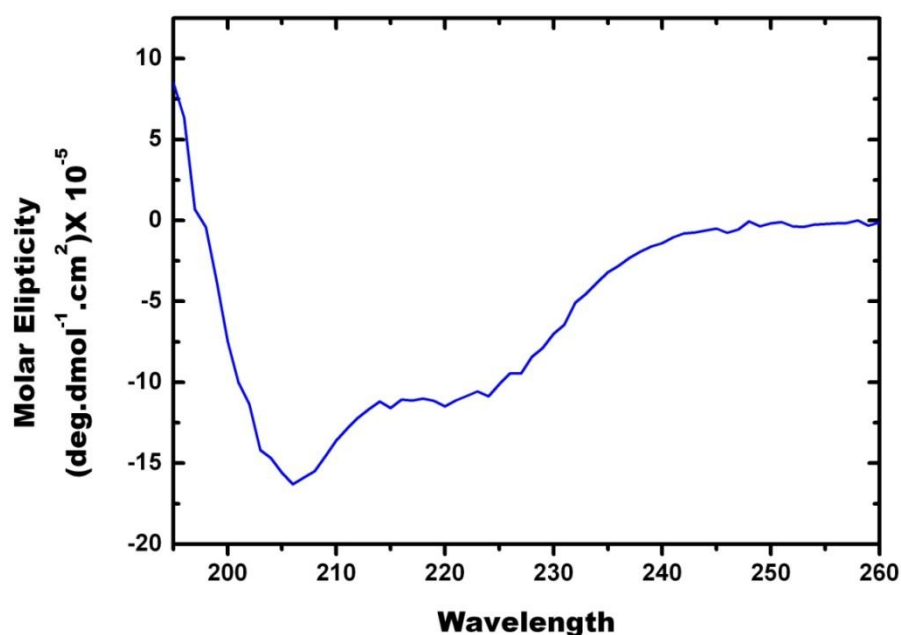
Fig. 3.10 - Far-UV CD spectrum of recombinant ataxin-3 produced and purified as described in the text above. The protein displays the expected high α-helical character (in blue), as seen from the minima at 208 and 222nm.

The results obtained (Fig. 3.10) are consistent with those previously published (compare to Fig. 3.2), with minima at 208nm and 222nm suggesting a high α-helical content and very similar overall spectra to published data (Chow, 2004; Gales, 2005). The recombinant protein purified in Section 3.2 is thus folded into a secondary structure consistent with ataxin-3. In addition, purified ataxin-3 protein displayed activity against Ub-AMC, as previously reported by Burnett and co-workers (Burnett, 2003), further strengthening the conclusion of correctly folded ataxin-3 protein expressed and purified.

3.4 Self-association of ataxin-3

The ability of recombinant ataxin-3, purified above, to assemble into β-enriched fibrillar material *via* characteristic lag-dependent kinetics (Gales, 2005; Ellisdon, 2006; Saunders, 2011), was next evaluated as a final criterion for the fidelity of the recombinant protein produced.

3.4.1 Kinetics of aggregation

In order to verify the self-association properties of recombinant full-length non-expanded ataxin-3, purified as above, monomeric protein stock (Section 2.3.8) was thawed and diluted to 10μM protein concentration in storage buffer (Section 2.2.2.3), supplemented with Thioflavin T (ThioT) (prepared as described in Section 2.2.2.6) to a final concentration of 10μM. The sample was next aliquoted into wells on a 96 well microplate and incubated quiescently at 37ºC (as detailed in Section 2.3.11.1.1). The progress of aggregation was monitored by the fluorescence intensity changes of ThioT at 480nm, read at regular intervals. The results (Fig. 3.11) show three distinct phases: the first, where little to no change in fluorescence intensity is visible, spans the first 30h of incubation at 37ºC. The second is marked by an abrupt increase in the ThioT fluorescence intensity reported and spans ~27h, from 30h to ~57h incubation. The third shows little change in fluorescence intensity and denotes a stationary phase in the reaction.

Classical amyloid theory of aggregation relates these fluorescence intensity-defined phases (and the corresponding aggregate formation being reported) to mechanistic and thermodynamical parameters of nucleated aggregation (Oosawa and Asakura, 1975; Ferrone, 1999). In this light, the initial phase reflects the lag phase, during which time thermodynamically unfavourable structures are formed as fibril nucleation occurs. This phase is followed by an exponential growth of the

aggregates by elongation of the structures formed, the elongation phase. Eventually the system reaches equilibrium, where the amount of aggregates present does not change, accounting for the stable fluorescence reported in the last phase, the stationary phase.
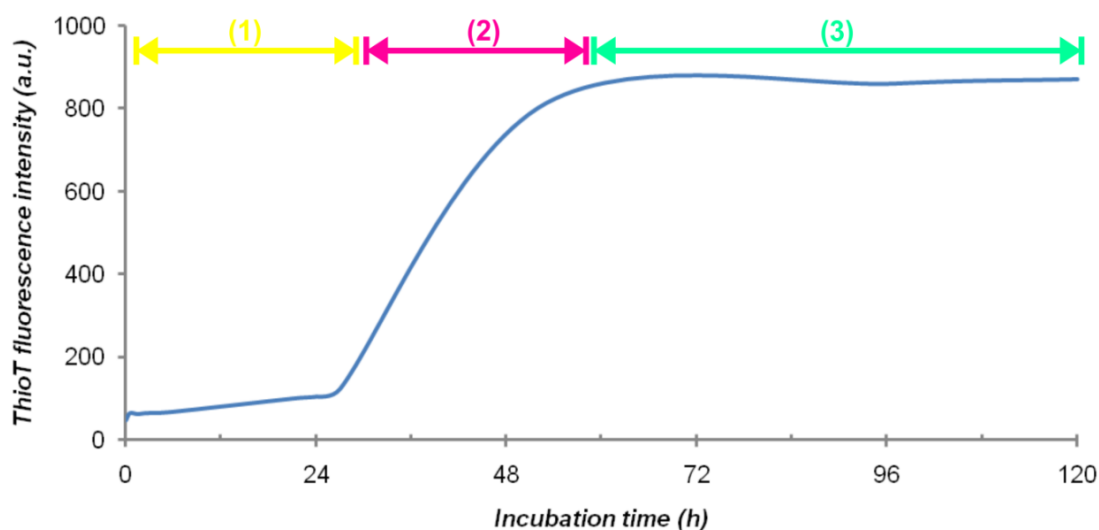
Fig. 3.11 – Self-association kinetics of recombinant full-length non-expanded ataxin-3 (3UIM isoform) followed by Thioflavin T fluorescence. Protein at 10µM concentration in 20mM HEPES pH 7.5, 200mM NaCl, 5% (v/v) glycerol, 1mM EDTA and 1mM DTT was placed at 37ºC, and quiescent aggregation monitored over time by the fluorescence intensity of 10µM ThioT. Three phases are visible in the kinetics, possibly representing the lag phase (1), elongation phase (2) and stationary phase (3) of a nucleated assembly reaction.

3.4.2 Early aggregation events

According to the nucleated assembly theory, during the first phase, where the fluorescence signal of ThioT does not report the formation of aggregated material, monomers or oligomers, but not fibrillar material, should be present in the sample. To verify this, protein sample prepared and incubated as described above (and detailed in Section 2.3.11.1.1) was injected at regular intervals into an analytical-grade size-exclusion column (Fig. 3.12), previously equilibrated with storage buffer (Section 2.2.2.3), and eluted with an isocratic flow rate as previously

described (detailed in Section 2.3.6.4). The aggregation progress of replicate sample aliquots was also monitored concomitantly by ThioT fluorescence, resulting in kinetics identical to those shown on Fig. 3.11.
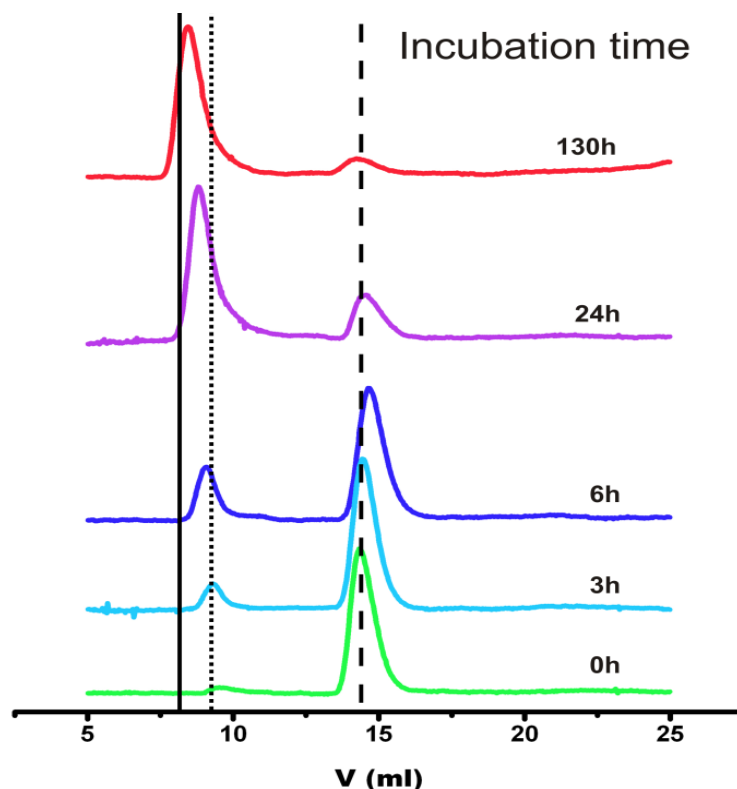
Fig. 3.12 – Analytical size-exclusion profiles of aggregating recombinant full-length non-expanded ataxin-3 (3UIM isoform) in storage buffer, at consecutive time-points of quiescent incubation at 37ºC. The protein concentration used was 10µM. Soluble monomeric protein, eluting at 13.6ml (indicated by a vertical dashed line on the figure) is progressively depleted from solution, with a concomitant increase of soluble oligomers with high molecular mass, eluting earlier. At initial timepoints, these oligomers elute around 9ml (indicated by a dotted black vertical line on the figure), progressively eluting closer to 8ml, the void-volume of the column (indicated by a full black vertical line on the figure).

Over the 30h incubation period corresponding to the lag phase (as suggested from Fig. 3.11), there is a clear and progressive loss of monomer accompanied by a corresponding increase in the concentration of soluble oligomers (Fig. 3.12), of high molecular weight as indicated by an elution volume close to 8ml, the void-volume of the column. Interestingly, the elution volume of these oligomeric species decreases over time, suggesting further growth in size. Low reactivity of the oligomers to ThioT could explain the lack of fluorescence intensity

changes during this phase, in agreement with previous reports (Gales, 2005), suggesting that fibrillar species are not significantly populated at this time. Taken together, these results show a progressive association of soluble monomeric protein into soluble higher molecular weight oligomers (Fig. 3.12) that, over time of incubation, assemble into species with ThioT tinctorial properties (Fig. 3.11), consistent with an amyloid or amyloid-like assembly reaction.

3.4.3 Products of self-association

To characterize the end-products of aggregation, and thus determine whether or not the self-association of the recombinant full-length non-expanded ataxin-3 prepared results in the formation of amyloid (or amyloid-like) fibrils, an aliquot of sample was taken after ~100h incubation (as described in Section 3.4.1) and imaged using negative stain EM as detailed in Section 2.3.11.3.1. Briefly, the sample was adsorbed onto glow-discharged copper grids with formvar coating, washed with RO-purified water and placed in contact with a 1% (w/v) uranyl solution (Section 2.1.6.2) to provide the necessary negative contrast. With the assistance of Rui Fernandes (technician responsible for the Histology and Electron Microscopy service at IBMC, Porto, Portugal), grids prepared with end-products of aggregation were visualized on a Zeiss EM902 microscope and images collected in digital format. The results clearly showed the generation of fibrillar structures (Fig. 3.13), work-like in morphology, with 9-12nm width and variable length, consistent with data published for both full-length non-expanded ataxin-3 and early steps of expanded ataxin-3 aggregation (Ellisdon, 2006).
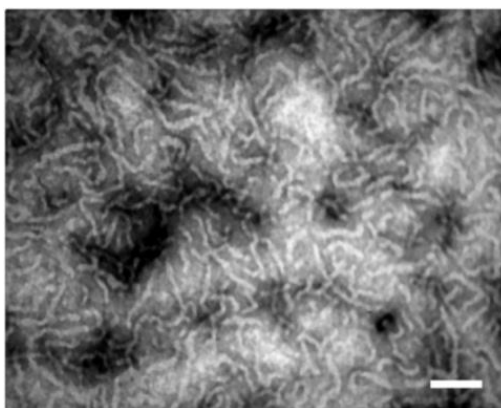
Fig. 3.13 – Characterization of end-products from the quiescent aggregation of recombinant ataxin-3. Negative-stain electron microscopy image of end-products of aggregation by recombinant non-expanded ataxin-3 (3UIM isoform), after ~100h quiescent incubation of 10µM protein in storage buffer at 37ºC. Bar represents 100nm. Width measurements, of the worm-like fibrillar aggregates visible range from 9 to 12nm. Length measurements are not possible to collect accurately due to the overlaying pattern of deposition.

To determine whether the aggregates formed contain ordered β-sheet, identical samples to those imaged (Fig. 3.13) were prepared for X-ray diffraction studies as detailed in Section 2.3.12.5. Briefly, fibrillar sample collected after ~100h quiescent incubation of 10µM ataxin-3 in storage buffer at 37ºC (as before) was allowed to air-dry for 3 days and analyzed under synchrotron radiation. Electron diffraction patterns were obtained with a 20 second exposure (as detailed in Section 2.3.12.5), and show the characteristic pattern of partially aligned amyloid fibrils (Baxa, 2005; Stromer and Serpell, 2005), with reflections at 4.7 and 10.4 Å (Fig. 3.14). Due to the short length and low rigidity common to worm-like fibrils, it is very hard to align them. Thus, the distinction between meridional and equatorial reflections is not trivial, as previously shown for other proteins such as β_2m (Smith, 2003). However, the direction of the strongest intensity for the 4.7Å ring is visible (Fig. 3.14 – dotted white line) and, upon image processing a perpendicular orientation of the strongest intensity for the 10Å ring is hinted, suggesting they represent meridional and equatorial reflections, respectively. Longer acquisition time and/or better oriented sample may provide more detailed insight into the fibrillar architecture, though a high content in beta-sheet is clear from these results.
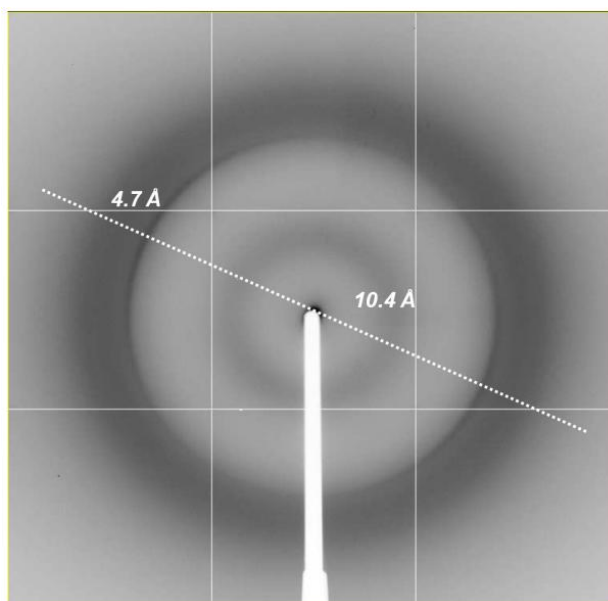
Fig. 3.14 – X-ray diffraction pattern of recombinant non-expanded ataxin-3 (3UIM isoform) partially aligned fibrillar aggregates. The characteristic rings at 4.7 and 10.4 Å are indicated in the figure, as well as the visible direction of intensity on the 4.7Å ring, indicated by a dotted line, representing the meridional direction.

3.5 Conclusions

This present chapter describes the expression and purification strategies followed for the production of recombinant monomeric non-expanded full-length ataxin-3 (3UIM isoform), henceforth referred to in this thesis as "AT3[U3/14Q]" for simplification purposes. The protein produced was shown to share the biochemical and biophysical characteristics described in literature for ataxin-3, namely the expected mass, a high α-helical character and activity as a deubiquitinase (DUB) enzyme. Importantly, under the preparation and storage conditions the protein was shown to be monomeric, and hence the ideal starting point for the study of ataxin-3 aggregation under non-denaturing conditions. Furthermore, the results presented show that the self-association of AT3[U3/14Q] takes place through kinetics that suggest a nucleated assembly process, whereby soluble monomers aggregate to soluble oligomers (that do not bind ThioT), followed by the rapid formation of

ThioT-positive species that eventually display fibrillar morphology and a high β-sheet (possibly cross-β) content. These characteristics are consistent with the formation of ataxin-3 amyloid and amyloid-like assemblies with lag-dependent kinetics, as shown by previous work (Gales, 2005; Ellisdon, 2006).

The following chapters build on these initial observations, developing assays for quantitative analysis of the mechanism of AT3[U3/14Q] aggregation and the role of individual domains within the structure of the protein in directing this process.

Chapter 4

4 – Determining aggregation conditions

4.1 Introduction

In the field of biochemistry and biophysics, physiological conditions for a protein are those displayed by the histological or cellular compartment it is located in, within the organism it refers to. *In vitro* experiments by several research groups over the years have investigated the aggregation of ataxin-3 under physiologically-relevant conditions, i.e., conditions approximating those within human cytoplasm, namely pH and ion concentration (primarily sodium and chloride), along with temperature and pressure (Chow, 2004 and 2006; Gales, 2005; Ellisdon, 2007; Saunders, 2011). There is no absolute formula for *in vitro* conditions to be physiologically-relevant, but a pH around 7.4, ion concentration around 200mM, 37ºC temperature and atmospheric pressure are general guidelines for the cytoplasm. Prof. Steve Bottomley's group (Monash University, Victoria, Australia) has been prolific in their investigations of ataxin-3 aggregation under physiologically-relevant conditions, revealing important insights into ataxin-3 aggregation (among others: Chow, 2004 and 2006; Ellisdon, 2006 and 2007; Saunders, 2011). However, when comparing the results of experiments performed on ataxin-3 by different authors, additional criteria have to be considered, particularly the protein isoform. For example, whilst the group mentioned above has chosen to study the 2UIM isoform of ataxin-3 (originally described in Kawaguchi, 1994), other authors have focused on the 3UIM isoform (originally described in Goto, 1997), differing in the C-terminal part of the protein. Little is known about the *in vivo* significance of these splice variants, for lack of information on the differential expression or post-translational modifications *in vivo*. Recent work by Harris and co-workers has attempted to shed light on this matter, their data indicating that the 3UIM isoform is the predominant isoform expressed in brain tissue (Harris, 2010), with slower *in vivo* aggregation than the 2UIM isoform. A common aggregation pathway, differing only in rate but not in molecular mechanism, was however not ruled out.

Recent reports on the aggregation propensity of proteins structurally comprised of a globular domain followed by a flexible C-terminal tail (Paushkin, 1996; Balguerie, 2003; Baxa, 2005), as is the case of ataxin-3 (Masino, 2003), point to a direct involvement of this C-terminal region on the resulting aggregate core. Together with the findings of Harris and co-workers (Harris, 2010), the study of ataxin-3's C-terminal flexible tail on the protein's aggregation mechanism gains an even greater importance. Interestingly, work by Ellisdon and co-workers on the 2UIM isoform of ataxin-3 has ruled out a role for the polyQ feature, localized to this flexible region of the protein, in the early aggregation steps, validating the study of a non-expanded form of the protein in the *in vitro* study of the pathological aggregation.

Comparing the aggregation of C-terminal truncated forms of ataxin-3 provides further insights into the aggregation mechanism of the protein (see Chapter 6) but, of course, a basis for the comparison must first be established, by determining in kinetic and mechanistic detail the aggregation parameters of the full-length protein (see Chapter 5). Work by several groups (Gosal, 2005, Nichols, 2005; reviewed in Wang, 2010) has shown that the environment conditions has a direct influence on protein aggregation, thus, in order to achieve this thesis' goals, conditions to yield reproducible AT3[U3/14Q] aggregation kinetics were first established. To determine physiologically-relevant conditions for reproducible *in vitro* aggregation of AT3[U3/14Q] is the goal of this chapter.

4.2 Determining the aggregation buffer

To determine the simplest set of physiologically-relevant conditions, able to produce reliable and reproducible AT3[U3/14Q] aggregation kinetics, prepared monomeric protein sample was thawed and buffer exchanged into buffer lacking one of the following additives: reducing agent; ion chelators; polyol. Buffer exchanged sample into complete buffer was also prepared, as control. In addition,

and as a preemptive measure against contamination of the sample during aggregation experiments, the use of biocides in the aggregation buffer was evaluated.

4.2.1 Biocides

Chemicals to attack and destroy living organisms are a common resource in protein biochemistry. Even though much care is taken to keep laboratory equipment and work areas sterile, only specially equipped facilities can actually make the claim of sterility. For the most part, addition of biocides to solutions, when possible, is common practice, avoiding bacterial and fungal contamination, which would lead to degraded sample (as the organisms use the protein sample as a carbon or nitrogen source). Results under contaminated conditions are obviously far from reliable, but at the same time, if a biocide is used it must be shown not to affect the experiment (Fig. 4.1).
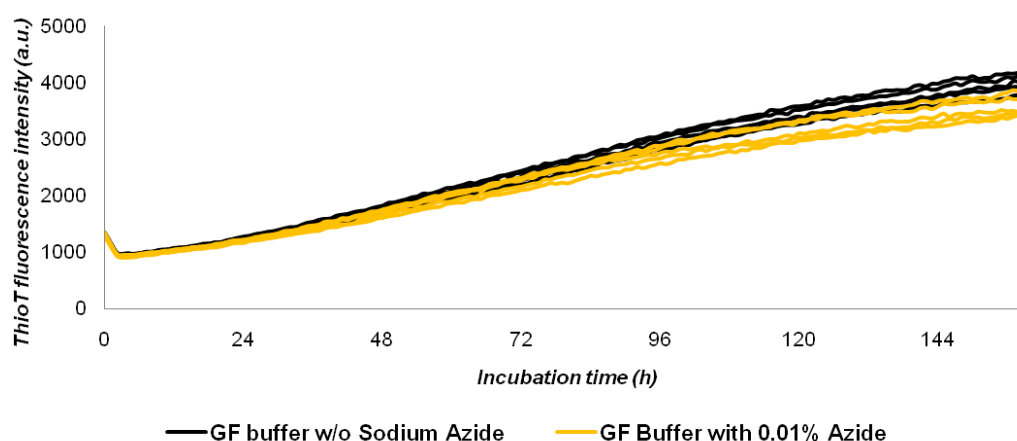
Fig. 4.1 – Effect of NaN3 on the aggregation profiles of 10µM AT3[U3/14Q]. Protein in storage buffer lacking (in black) or containing (in orange) the biocide NaN3, at 0.01% (w/v) concentration, and incubated at 37ºC. Quiescent aggregation of AT3[U3/14Q] was monitored by the fluorescence intensity changes of 10µM ThioT. The experiment was conducted on a BMG Galaxy plate reader, equipped with excitation and emission filters for ThioT: 440nm and 480nm, respectively.

Monomeric protein sample was incubated at 37ºC in storage buffer (Section 2.2.2.3) in the absence or presence of 0.01% (w/v) NaN_3 (as described in Section 2.3.11.1.2). The results (Fig. 4.1) showed no interference by the biocide on the quiescent aggregation of AT3[U3/14Q], and this biocide was thus used in all subsequent experiments in this thesis. For simplification purposes, storage buffer supplemented with 0.01% (w/v) NaN_3 is henceforth referred to as TX buffer (Section 2.2.2.7.1).

4.2.2 Reducing agents

Chemical reducing agents, such as DTT or BME have been used in the study of ataxin-3's aggregation *in vitro* (Gales, 2005 and Ellisdon, 2006, respectively), and help to balance the oxidation of protein chemical groups by the oxygen in solution. In addition, these reagents may also disrupt disulfide bonds that can form intra- or inter-molecularly between cysteine residues. In ataxin-3 no intramolecular disulfide bonds have been reported, although the protein does have five cysteine residues, four of them in the N-terminal Josephin domain (Fig. 4.3), of which two are solvent-accessible, Cys14 (active-site cysteine) and Cys172 (Fig. 4.2).
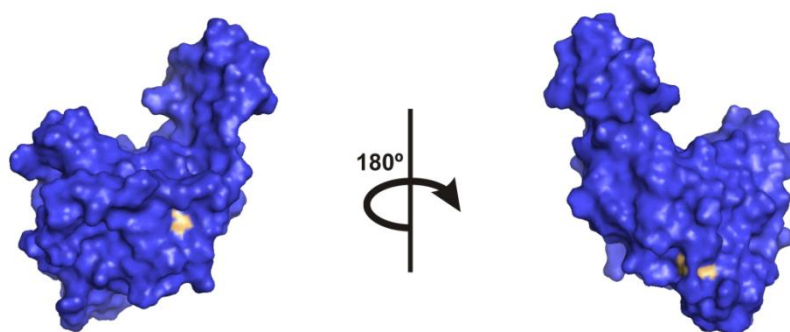
Fig. 4.2 – Surface maps of the Josephin domain of ataxin-3, highlighting the solvent-accessible cysteine residues. The N-terminal domain of ataxin-3, the Josephin domain (PDB access code: 1YZB) is shown in blue and the accessible cysteines are shown in yellow. Cys14 is visible on the image on the left and Cys172 on the image on the right. Surface maps were built with PyMOL (The PyMOL Molecular Graphics System, Schrödinger, LLC).

Despite Bevivino and Loll's work showing that alkylation of cysteines has no effect on ataxin-3 aggregation (Bevivino, 2001), the thiol groups of these amino acid residues may interact intermolecularly, driving aggregation of the protein *in vitro*, in a manner that is highly unlikely to mirror intracellular aggregation.

```
atgtcgtactaccatcaccatcaccatcacctcgaaaacctgtattttcagtctaaagca
 M  S  Y  Y  H  H  H  H  H  H  L  E  N  L  Y  F  Q  S  K  A
ggctggatggagtccatcttccacgagaaacaagaaggctcactttgtgctcaacattgc
 G  W  M  E  S  I  F  H  E  K  Q  E  G  S  L  C  A  Q  H  C
ctgaataacttattgcaaggagaatatttttagccctgtggaattacctcaattgcacat
 L  N  N  L  L  Q  G  E  Y  F  S  P  V  E  L  S  S  I  A  H
cagctggatgaggaggagaggatgagaatggcagaaggaggagttactagtgaagattat
 Q  L  D  E  E  R  M  R  M  A  E  G  G  V  T  S  E  D  Y
cgcacgtttttacagcagccttctggaaatatggatgacagtggtttttttctctattcag
 R  T  F  L  Q  Q  P  S  G  N  M  D  D  S  G  F  F  S  I  Q
gttataagcaatgccttgaaagtttgggtttagaactaatcctgttcaacagtccagag
 V  I  S  N  A  L  K  V  W  G  L  E  L  I  L  F  N  S  P  E
tatcagaggctcaggatcgatcctataaatgaaagatcatttatatgcaattataaggaa
 Y  Q  R  L  R  I  D  P  I  N  E  R  S  F  I  C  N  Y  K  E
cactggtttacagttagaaaattaggaaaacagtggtttaacttgaattctctcttgacg
 H  W  F  T  V  R  K  L  G  K  Q  W  F  N  L  N  S  L  L  T
ggtccagaattaatatcagatacatatcttgcacttttcttggctcaattacaacaggaa
 G  P  E  L  I  S  D  T  Y  L  A  L  F  L  A  Q  L  Q  Q  E
ggttattctatatttgtcgttaagggtgatctgccagattgcgaagctgaccaactcctg
 G  Y  S  I  F  V  V  K  G  D  L  P  D  C  E  A  D  Q  L  L
cagatgattagggtccaacagatgcatcgaccaaaacttattggagaagaattagcacaa
 Q  M  I  R  V  Q  Q  M  H  R  P  K  L  I  G  E  E  L  A  Q
ctaaaagagcaaagagtccataaaacgacctggaacgagtgttagaagcaaatgatggc
 L  K  E  Q  R  V  H  K  T  D  L  E  R  V  L  E  A  N  D  G
tcaggaatgttagacgaagatgaggaggatttgcagagggctctggcactaagtcgccaa
 S  G  M  L  D  E  D  E  E  D  L  Q  R  A  L  A  L  S  R  Q
gaaattgacatggaagatgaggaagcagatctccgcagggctattcagctaagtatgcaa
 E  I  D  M  E  D  E  E  A  D  L  R  R  A  I  Q  L  S  M  Q
ggtagttccagaaacatatctcaagatatgacacagacatcaggtacaaatcttacttca
 G  S  S  R  N  I  S  Q  D  M  T  Q  T  S  G  T  N  L  T  S
gaagagcttcggaagagacgagaagcctactttgaaaaacagcagcaaaagcagcaacag
 E  E  L  R  K  R  R  E  A  Y  F  E  K  Q  Q  Q  K  Q  Q  Q
cagcagcagcagcagcagcagggggacctatcaggacagagttcacatccatgtgaaagg
 Q  Q  Q  Q  Q  Q  Q  G  D  L  S  G  Q  S  S  H  P  C  E  R
ccagccaccagttcaggagcacttgggagtgatctaggtgatgctatgagtgaagaagac
 P  A  T  S  S  G  A  L  G  S  D  L  G  D  A  M  S  E  E  D
atgcttcaggcagctgtgaccatgtctttagaaactgtcagaaatgatttgaaaacagaa
 M  L  Q  A  A  V  T  M  S  L  E  T  V  R  N  D  L  K  T  E
ggaaaaaaataa
 G  K  K  *
```

Fig. 4.3 - Nucleotide sequence and the translated amino acid residue sequence for the 3UIM isoform of full-length non-expanded ataxin-3 studied in this thesis. The protein contains an N-terminal Josephin domain (highlighted in red), three ubiquitin interacting motifs (highlighted in green) and a polyglutamine tract (highlighted in blue). To aid in purification a hexahistidine tag and a linker were added to the N-terminus (highlighted in yellow), and a tobacco etch virus (TEV) protease recognition site engineered at the end. Cleavage site is noted by a black arrow on the figure. The nucleotide sequence, as obtained from sequencing, was translated using the tools available on the ExPASy proteomics server (http://expasy.org). The UIM sequences highlighted follow the work by Donaldson and co-workers (Donaldson, 2003). The cysteines in the sequence are highlighted by a yellow background.

To determine the relative importance of reducing agents on the aggregation details of ataxin-3, 10µM AT3[U3/14Q] was incubated quiescently at 37ºC in TX buffer in the presence or absence of 1mM DTT, and the progress of aggregation monitored by ThioT fluorescence (as described in Section 2.3.11.1.2).

The results show that in the absence of a reducing agent the aggregation profile of AT3[U3/14Q] differs from that obtained with complete TX buffer, with an almost complete lack of a lag phase (Fig. 4.4). The differences are very clear and point to a crucial role of a reducing environment in ataxin-3 aggregation. Given that the environment of a cell typically balances redox potentials to avoid protein oxidation, it is unlikely that this effect reflects an *in vivo* situation and, as such, DTT (at 1mM strength) was kept as part of the aggregation buffer.

Fig. 4.4 – Effect of DTT on the aggregation profiles of 10µM AT3[U3/14Q]. Protein sample was buffer exchanged into TX buffer lacking (in red) or containing (in black) the reducing agent DTT, at 1mM strength, and incubated at 37ºC. Quiescent aggregation of AT3[U3/14Q] was monitored by the fluorescence intensity changes of 10µM ThioT. The experiment was conducted on a BMG Galaxy plate reader, equipped with excitation and emission filters for ThioT: 440nm and 480nm, respectively.

4.2.3 Ion Chelation

Previous work by several authors, has shown increased oligomerization and/or unspecific aggregation in the presence of metallic ions, specifically Cu^{2+}, Cd^{2+}, Zn^{2+}, Ni^{2+}, Mn^{2+} and Al^{3+} (Ricchelli, 2007 and Stawoska, 2009). To determine the influence of divalent ions present in the aggregation buffer (whether co-purifying with the protein or from reagent contamination), purified monomeric AT3[U3/14Q] was buffer exchanged into TX buffer either lacking or containing 1mM EDTA, an ion chelator particularly effective in scavenging divalent cations. Quiescent aggregation kinetics of AT3[U3/14Q] at 37ºC were monitored by ThioT fluorescence, as before (detailed in Section 2.3.11.1.2).

Fig. 4.5 – Effect of EDTA on the aggregation profiles of 10µM AT3[U3/14Q]. Protein sample was buffer exchanged into TX buffer lacking (in green) or containing (in black) the ion chelator EDTA, at 1mM strength, and incubated at 37ºC. Quiescent aggregation of AT3[U3/14Q] was monitored by the fluorescence intensity changes of 10µM ThioT. The experiment was conducted on a BMG Galaxy plate reader, equipped with excitation and emission filters for ThioT: 440nm and 480nm, respectively.

Removal of the ion chelator from the aggregation buffer TX results in slightly more rapid aggregation kinetics when compared to the results with complete TX buffer (Fig. 4.5), consistent with previous reports (Richelli, 2007 and Stawoska, 2009) suggesting a role for divalent cations on ataxin-3 aggregation. EDTA at 1mM concentration was therefore included in the buffer composition used in this thesis' experiments, to obviate discrepancies from the presence of ion contaminants.

4.2.4 Polyols

Glycerol, like other polyols, stabilizes the native structure of proteins by reducing the conformational flexibility of loops (Tsai, 2000). The molecule is strongly excluded from the protein, favoring its hydration. To evaluate the relative effect of this buffer component on the quiescent aggregation of AT3[U3/14Q], monomeric protein sample was buffer exchanged, as before, into TX buffer lacking or containing glycerol, and incubated at 37ºC (as detailed in Section 2.3.11.1.2). The progress of aggregation was determined by monitoring the fluorescence intensity changes of 10μM ThioT.
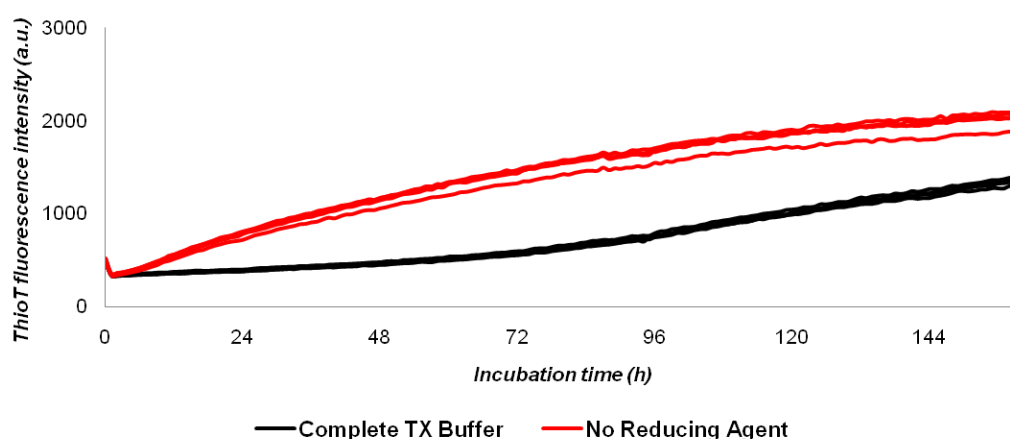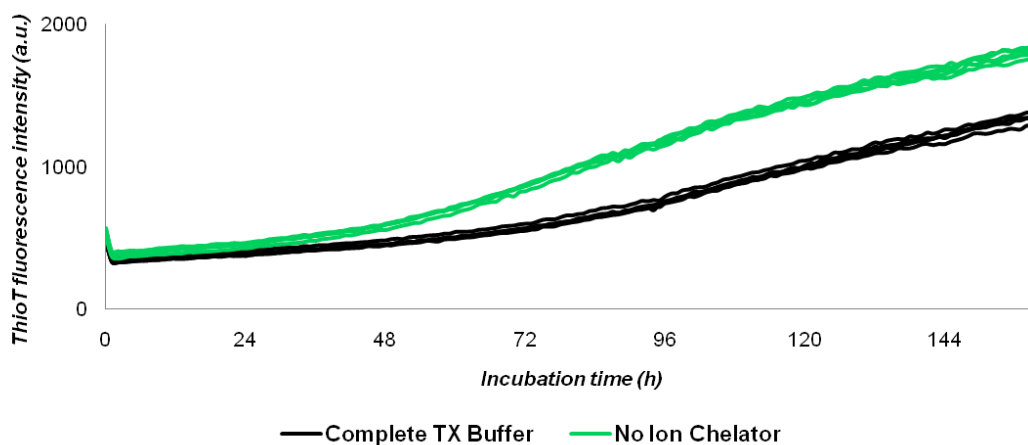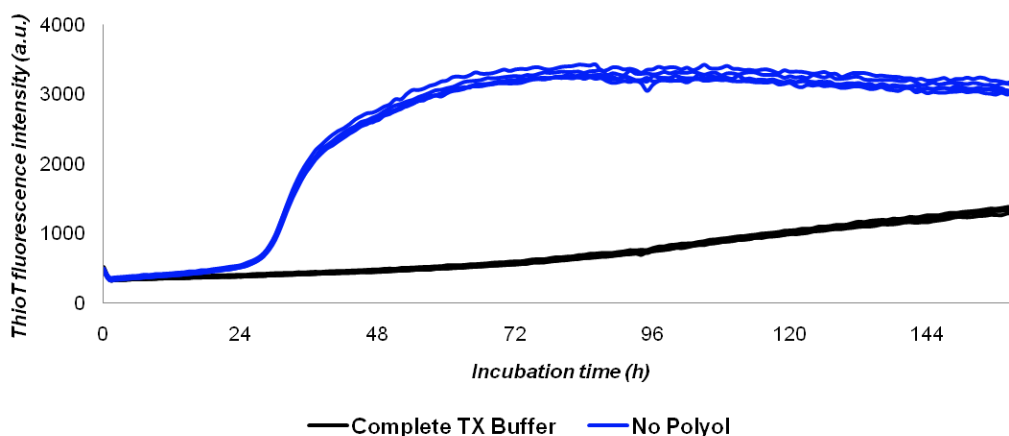

Fig. 4.6 – Effect of glycerol on the aggregation profiles of 10μM AT3[U3/14Q]. Protein sample was buffer exchanged into TX buffer lacking (in blue) or containing (in black) the polyol glycerol, at 5% (v/v) concentration, and incubated at 37ºC. Quiescent aggregation of AT3[U3/14Q] was monitored by the fluorescence intensity changes of 10μM ThioT. The experiment was conducted on a BMG Galaxy plate reader, equipped with excitation and emission filters for ThioT: 440nm and 480nm, respectively.

The effect of the polyol glycerol on the quiescent aggregation of 10μM AT3[U3/14Q] at 37ºC is shown in Fig. 4.6. The results are striking, revealing a dramatic increase in the elongation rate, as well as a well-defined lag phase in the absence of this polyol, when compared to the results obtained with complete TX buffer. After ca. 3 days the elongation phase (as monitored by ThioT fluorescence)

is over and the ThioT signal reaches a plateau, indicating the start of the stationary phase.

Negative stain electron micrographs of the aggregation products formed after approximately 100h hours incubation in TX buffer lacking glycerol, show fibrillar structures (Fig. 4.7 - B) not unlike those previously observed in complete storage buffer (Fig. 4.7 – A). The easily identifiable phases of aggregation, together with the retained characteristic *de novo* assembly into worm-like fibrils, justify the removal of the polyol from the aggregation buffer used to study ataxin-3 aggregation.
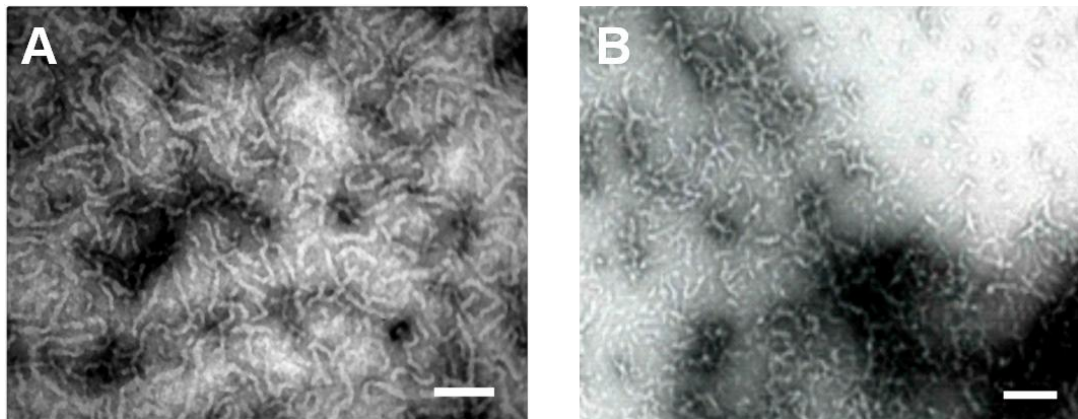
Fig. 4.7 – Negative-stain EM visualization of end-stage products formed by AT3[U3/14Q] aggregation. Sample was collected after approximately 100h of quiescent aggregation at 37ºC, from an initial AT3[U3/14Q] concentration of 10μM, in either storage buffer (A) or TX buffer lacking glycerol (B). Figure A is reprinted from Fig. 3.13. It is important to note that A and B were collected using different negative-stain EM protocols (detailed in Section 2.3.11.3.1 and 2.3.11.3.2 for A and B, respectively). Bars represent 100nm.

4.2.5 Final aggregation buffer

The results discussed above (Sections 4.2.1 through 4.2.4) show clear effects on the quiescent aggregation of AT3[U3/14Q] at 37ºC, when different components of the buffer system are altered. Taken together, they provide a better

understanding of how sensitive the aggregation process must be to the precise conditions in cells. Oxidative conditions are tightly regulated by a cell's biochemistry, potentially affecting protein aggregation in a manner similar to that described above (Section 4.2.2). On the other hand, divalent cations, which are found throughout the cell, have a direct effect on ataxin-3 aggregation, as the results in Section 4.2.3 suggest, though a determination of a specific ion's influence was not conducted as it is outside the scope of this thesis. A significant effect on the quiescent aggregation of AT3[U3/14Q] was found for glycerol, shown in Section 4.2.4, with aggregation profiles displaying clearly identifiable lag, elongation and stationary phases in the absence of the polyol. The addition of NaN$_3$ at 0.01% (m/v) does not affect the aggregation profiles obtained (Section 4.2.1) and, as a safeguard against contamination, this reagent became part of the final aggregation buffer.

Taken together, the results in the previous sections of this chapter dictated a final composition for the buffer used to study quiescent aggregation of ataxin-3 at 37ºC in this thesis: 20mM HEPES pH 7.5, 200mM NaCl, 1mM DTT, 1mM EDTA and 0.01% (w/v) NaN$_3$. Buffer with this composition was named TXB for practicality.

4.3 Influence of the ThioT probe in aggregation

Although not strictly part of the aggregation buffer determined, the probe used to monitor the aggregation progress, Thioflavin T (ThioT), may have a direct influence on the aggregation details. The actual binding mode of the chemical is still a matter of debate (reviewed in Reinke and Gestwicki, 2011), but the interaction with the structures formed is certainly required. For this reason, and good experimental practice, the relative influence of ThioT on the quiescent aggregation of AT3[U3/14Q] was assessed, by titrating the amount of the molecule present in aggregation reactions at 10μM initial protein concentration.

Monomeric protein sample was buffer exchanged (Section 2.3.9) into TXB buffer (Section 2.2.2.7.4) and supplemented with varying amounts of ThioT. Sample

aliquots were concomitantly incubated quiescently at 37ºC, in the presence of 10μM, 20μM, 30μM or 40μM ThioT, and the aggregation progress monitored as before (detailed in Section 2.3.11.1.2). Data for each reaction was normalized in intensity (Y axis) and further normalized along the time dimension (X axis), clustering all traces by the midpoint of the reaction, t_{50}, defined for this purpose as the length of time required for each reaction to reach half of the final intensity measured. The normalization procedures used are detailed in Section 2.3.12.2. Data normalized in this way bypasses both the expected differences in intensity (from the different concentration of the fluorescent probe) and possible differences in aggregation kinetics (from stochastic processes), allowing a direct comparison of the profiles obtained.

The results obtained, Fig. 4.8, suggest that no mechanistic difference in the aggregation of AT3[U3/14Q] arises from varying the concentration of ThioT used, at least for concentrations between 10 and 40μM. Subsequent experiments in this thesis follow this interval in ThioT concentration.

4.4 Data Reproducibility

When establishing experimental conditions to study protein aggregation kinetics, a high level of reproducibility is required. Once the buffer composition for aggregation studies was determined, the next step was to assess data reproducibility under the selected conditions. Experimental variability can arise from intrinsic differences in the protein samples themselves or result from user handling, as well as from the stochastic nature of the aggregation process itself.
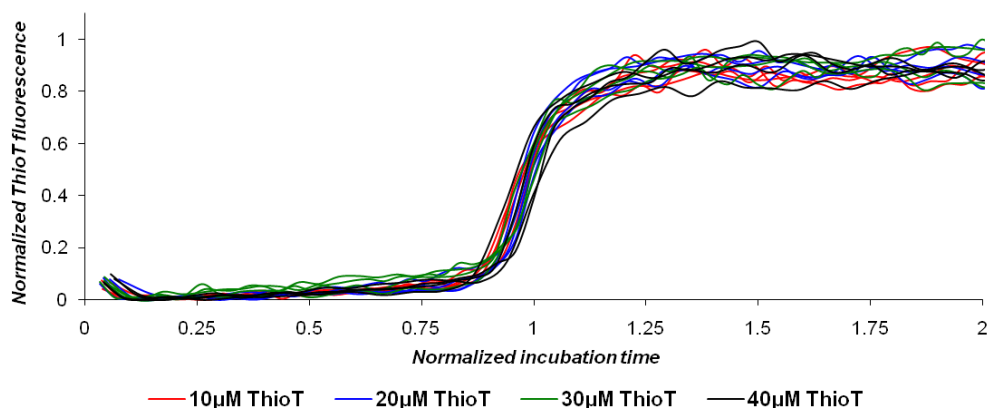
Fig. 4.8 – Effect of ThioT concentration on the aggregation profiles of 10µM AT3[U3/14Q]. Protein sample was buffer exchanged into TXB buffer supplemented with either 10, 20, 30 or 40µM ThioT, and incubated at 37ºC. Quiescent aggregation of AT3[U3/14Q] was monitored by the fluorescence intensity changes of the ThioT probe. The experiment was conducted on a BMG Galaxy plate reader, equipped with excitation and emission filters for ThioT: 440nm and 480nm, respectively.

To assess the reproducibility of sample handling on the aggregation of AT3[U3/14Q], replicate experiments following the aggregation protocol described in Section 2.3.11.1.2 were undertaken separately. For each replicate experiment protein sample was thawed, buffer exchanged into TXB, diluted to the initial protein concentration tested (10µM), and ThioT was added to a final concentration of 10µM (as described in Section 2.3.11.1.2). Sample was next incubated at 37ºC and the aggregation progress was monitored by the fluorescence intensity changes of ThioT. In addition, sample after ca. 100h incubation was collected and EM grids prepared (as described in Section 2.3.11.3.2).

The results shown in Fig. 4.9 reveal that the aggregation kinetics obtained within each experiment cluster very well, but between experiments they do not. However, negative-stain EM images obtained from sample collected after ca. 100h incubation show similar aggregation products, suggesting that replicate experiments are mechanistically consistent. Since the aggregation data collected in each experiment clusters together well (Fig. 4.9), stochastic processes are ruled out as the origin for variability. For each aggregation experiment, thawed sample aliquots are buffer exchanged into freshly prepared buffer, and subsequently supplemented with ThioT solution before being diluted to the tested protein

concentration. Next, the prepared reaction mixture is dispensed into replicate wells on a microplate, prior to sealing and incubation. Given the results obtained in Section 4.2, suggesting the strong influence of buffer composition on the aggregation of AT3[U3/14Q] under quiescent conditions, minute discrepancies in the buffer prepared could be at the root of the irreproducibility. Alternatively, or concomitantly, sample handling during dilution and dispensing steps could originate the variability, either by human error, time taken between steps or vigorousness in mixing, for example.
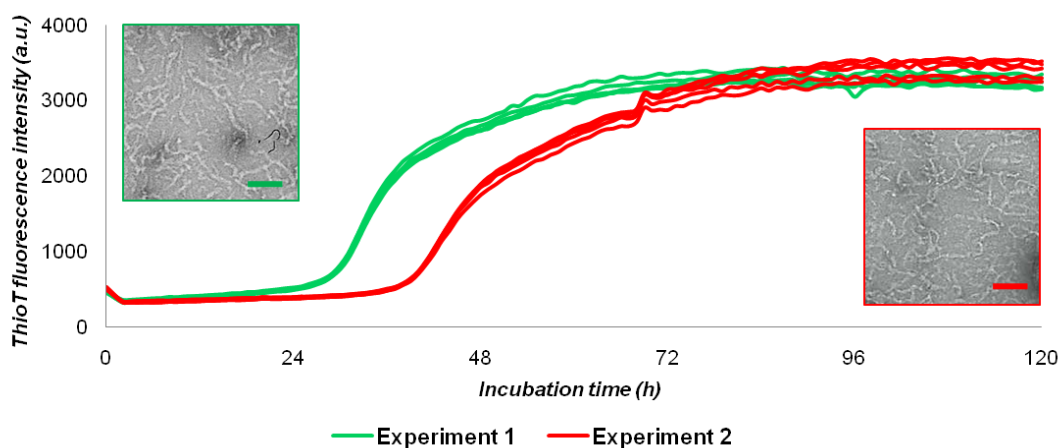
Fig. 4.9 – Reproducibility of quiescent AT3[U3/14Q] aggregation. Replicate experiments with aliquots from the same protein purification batch were prepared for aggregation studies, as before, at 10µM initial protein concentration. Reaction mixtures were incubated at 37ºC and the aggregation progress was monitored by the fluorescence intensity changes of 10µM ThioT. Sample from each replicate experiment was collected after ca. 100h incubation and visualized by negative-stain EM – insets. Experiment 1 is shown in green and experiment 2 is show in red. Bars on the EM images represent 100nm. The ThioT-monitored experiment was conducted on a BMG Galaxy plate reader, equipped with excitation and emission filters for ThioT: 440nm and 480nm, respectively. The EM grids were visualized on a Philips CM-10 electron microscope.

To tackle the variability issue, and following the reasoning discussed above, the use of a concentrated buffer stock instead of freshly prepared buffer was evaluated (Section 4.4.1). In addition, to bypass possible experimental discrepancies in dilution and dispensing of sample, the use of a robotic strategy in the setting-up of an aggregation trial was investigated (Section 4.4.2).

4.4.1 Buffer preparation

As mentioned above, one of the possible sources for the variability observed in Fig. 4.9 is the freshly prepared buffer for each experiment. To assess the influence of buffer preparation on the quiescent aggregation of AT3[U3/14Q], monomeric sample aliquots were buffer exchanged into either freshly prepared TXB buffer (as before), or buffer of identical composition prepared from the dilution of a 5-fold concentrated stock (5TXB, described in Section 2.2.1.13). Note that the buffer stock prepared does not include DTT, as this reagent decays over time at room temperature. For this reason, DTT was added immediately prior to use, to a final concentration of 1mM. For practicality, buffer originating from this concentrated stock (5TXB) is from here on named "TXBa".

Quiescent aggregation of 30µM initial concentration of AT3[U3/14Q] at 37ºC was next monitored by ThioT fluorescence changes, as described above (detailed in Section 2.3.11.1.2). The resulting aggregation kinetics show good reproducibility between buffers (Fig. 4.10), suggesting that quiescent aggregation of AT3[U3/14Q] at 37ºC is unaffected by buffer preparation. Given the practical advantages of a buffer stock, together with the bypassing of any possible influence from buffer variability (helping the quality of collected data and aiding in data interpretation), the use of this stock was thus adopted for all subsequent aggregation experiments in this thesis.
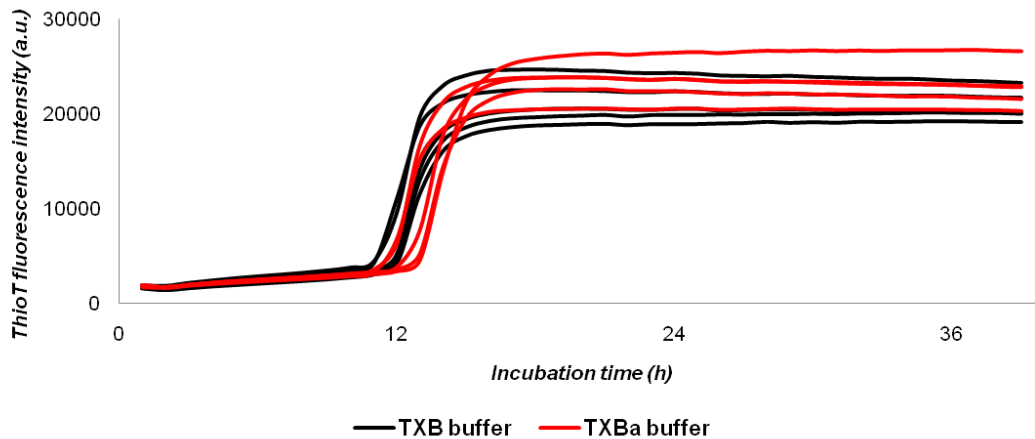
Fig. 4.10 – Influence of buffer preparation on the quiescent aggregation of AT3[U3/14Q]. Sample aliquots from the same protein purification batch were buffer exchanged into either freshly prepared buffer (TXB) or buffer of identical composition, prepared from the dilution of a concentrated stock (TXBa). Quiescent aggregation of 10µM AT3[U3/14Q] at 37ºC was monitored by the fluorescence changes of 10µM ThioT over time of incubation. The experiment was conducted on a BMG Omega plate reader, equipped with excitation and emission filters for ThioT: 440nm and 480nm, respectively.

4.4.2 Robotic strategy

In order to bypass experimental discrepancies in the dilution and dispensing steps of an aggregation trial (mentioned above), an automated strategy was developed. In addition, the expected speed increase in setting-up a microplate for an aggregation trial allows earlier timepoints of aggregation to be collected.

4.4.2.1 384 well microplates

In order to take full advantage of a robot's greater accuracy, 384 well Corning 3821BC, black polystyrene microplates were considered to offer advantages over 96 well Corning 3915 microplates of the same material and well

shape, used previously in this thesis. These 384 well microplates not only have a lower volume per reservoir (50µl), with its obvious maximization of sample use, but also allow more tests to be performed on the same microplate, in turn allowing for a more accurately faithful comparison between tests. Nonetheless, the smaller volume and dimension of each reservoir represent a change in the aggregation conditions and the consistency of aggregation data needed to be evaluated. Monomeric AT3[U3/14Q] sample was buffer exchanged into TXBa buffer and manually diluted and dispensed into replicate wells, as before (detailed in Section 2.3.11.1.2), this time on a 384 well Corning microplate. After sealing the plate, the sample was incubated quiescently at 37ºC, as before, and the aggregation progress of 30µM initial AT3[U3/14Q] concentration was monitored by the fluorescence changes of 10µM ThioT (as detailed in Section 2.3.11.1.2) to evaluate the consistency of kinetic parameters. Sample after ca. 100h incubation was collected and used to prepare negative-stain EM grids, as before (detailed in Section 2.3.11.3.2), in order to evaluate the consistency of mechanistic parameters.

The results obtained from the quiescent aggregation of AT3[U3/14Q] on 384 well microplates (Fig. 4.11) resemble those previously obtained with 96 well microplates (Section 4.2.4), albeit a greater spread of results along the time dimension is seen on 384 well microplates, possibly resulting from lower volume of sample utilized and an inherent stochasticity of the aggregation. Nonetheless, aggregation profiles display clear lag, elongation and stationary phases, and negative-stain EM images show fibrillar aggregates, worm-like in appearance (Fig. 4.11 – inset) and indistinguishable from those previously observed (Fig. 4.9 - insets).
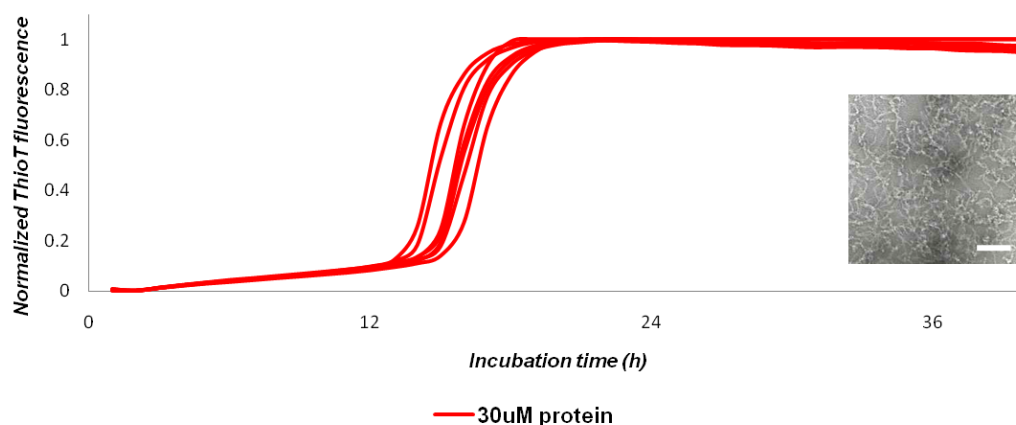
Fig. 4.11 – Judging the influence of reservoir dimensions and sample volumes on the quiescent aggregation of AT3[U3/14Q]. Monomeric sample in TXBa at 30µM initial concentration of AT3[U3/14Q] was incubated quiescently at 37ºC and the progress of aggregation monitored by the fluorescence changes of 10µM ThioT. Sample after ca. 100h incubation was visualized by negative-stain EM (inset). Bar represents 100nm. The ThioT-monitored experiment was conducted on a BMG Omega plate reader, equipped with excitation and emission filters for ThioT: 440nm and 480nm, respectively. The EM grids were visualized on a Philips CM-10 electron microscope.

In conclusion, 384 well microplates are perfectly suited to further study the aggregation of non-expanded ataxin-3, both in terms of aggregation kinetics, and aggregation products formed. To increase the reliability of data collected, a higher number of replicates will be needed in future experiments (specifically ten replicates, as described in Section 2.3.11.1.3), to reduce the effects of the inherent variability.

4.4.2.2 Programming the robot

With the help of Dr. David Bunka (Astbury Centre for Structural Molecular Biology, University of Leeds, UK), mixing and dispensing protocols for a Biomek® 2000 series robot from Beckman-Coulter (Brea, CA, USA) were designed using the manufacturer's provided program, Bioworks™. The intellectual goal was to plan for a final layout of the microplate that simultaneously made the best out of the robot's capabilities, both in speed and accuracy, and allowed simple modifications

to fit future test needs. In addition, the method of signal measurement on the plate reader was taken into account when designing the layout of the microplate.

The first premise followed pertains to the thermostability of incubation. To ensure stable temperature of the sample throughout the incubation, the two outermost reservoirs on all sides of the microplate are not taken up by protein samples, as they are more likely to suffer from temperature variations. Instead, they are reserved for buffer controls, which typically show little to no effect from temperature (Dr. W.F. Xue, personal communication).

The second premise equates the time required to read all the wells on the 384 well microplate and the speed of the events monitored. To ensure an even distribution of measurements and adequate sampling of fast events (since aggregation occurs also during the measurements), the data collected follows the guidelines:

- **Higher concentration reservoirs are measured before lower concentration reservoirs**
- **Each set of replicates is measured only after all the tests of the previous set are themselves measured**

To ensure these guidelines are followed, and keeping the measurement time as short as possible, a snaking top-to-bottom-to-top movement of the optical components of the plate reader is required (avoiding excessive repositioning of the components), and dictates the sequence of tests for each column of reservoirs. This sequence alternates between tests on the microplate, as much as possible since the robot operates an 8-lane multitip pipette that moves only in X and Y vectors. The resulting layout has the potential for larger spreads of results along the time dimension (due to the time between measurements of replicates), compensated by ensuring a high number of replicates per test on each layout (specifically, ten replicates) (Fig. 4.12).
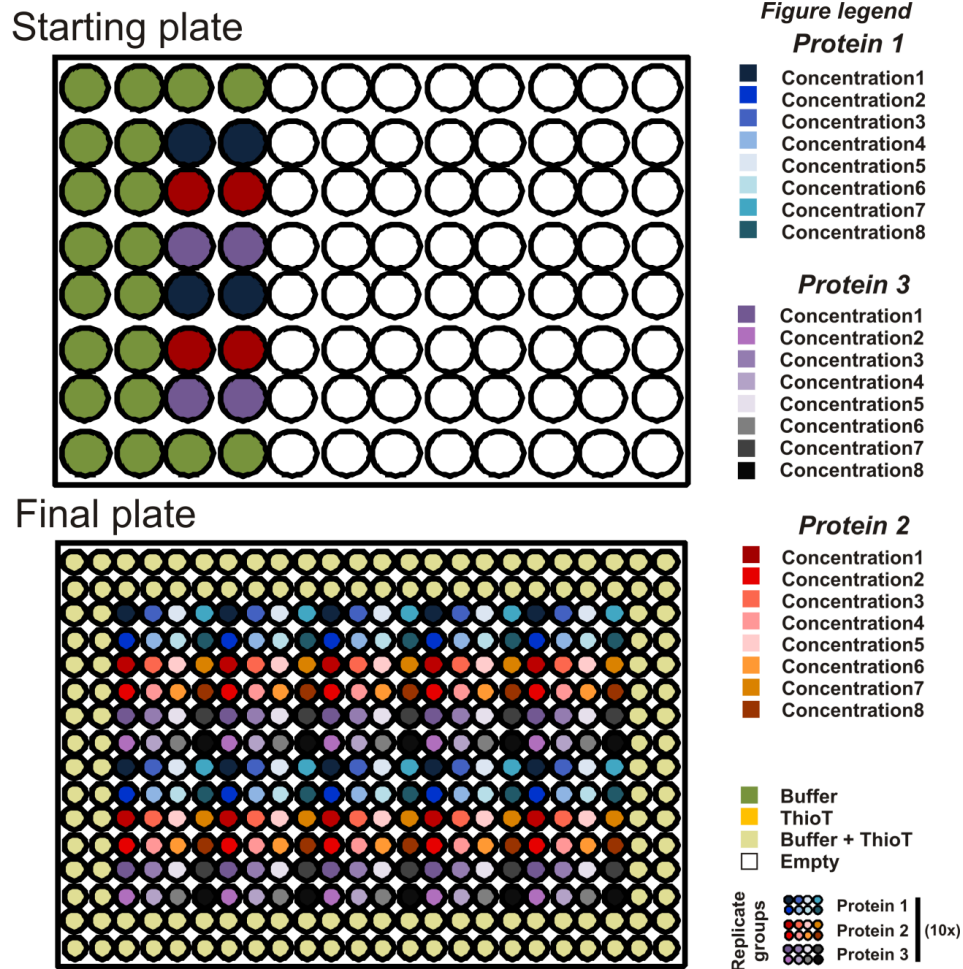
Fig. 4.12 – Schematic representation of the starting and final microplate layout used in the protein titration robot program. Protein sample is dispensed onto specific wells on the starting plate (top), at the highest protein concentration of the test, and diluted and dispensed by the Biomek® robot, originating the layout of sample on the final 384 well microplate (bottom). Three proteins can be tested per plate, with 8 concentrations of each prepared and clustered together in the layout (see bottom of figure legend). Each 4 columns with sample represent a replicate set of tests. Buffer blanks are placed in the outer wells, so as to further stabilize the temperature on the inner wells.

A third premise followed relates to the length of time required to set-up a 384 well microplate and the importance of collecting timepoints as early as possible so that extremely fast processes can be monitored. To reduce sample handling time, a pause point was added to the program. As much as possible, the solutions without protein are dispensed and mixed before this pause, with protein sample added to the system after it. This allows for a tighter control of the sample handling time, as the time between placing the protein in aggregation buffer and starting the monitoring of the reaction is reduced.

Finally, a fourth premise was followed to simplify the adapting of the program to future needs. All components of each reaction are separate at the start of the program and arranged in a modular manner (Fig. 4.12), so that more tests or more replicas can be included without changes to the program. For example, to measure more replicates of protein 1 (and consequently less tests), the wells for one of the other proteins, shown in Fig. 4.12, can be used to include more sample of protein 1, knowing that the well on the left for each protein contains the sample for three sets of ten replicas, and the one on the right for the other two sets of ten replicas. The same principle can be used to perform more tests on one plate (and consequently fewer replicas).

Three programs were designed for automated strategies, as required for the experiments conducted in this thesis: protein concentration titration; preparation of fibrillar products; seeding experiments. The details for each program are discussed in Sections 2.3.11.1 and 2.3.11.2.

4.4.2.3 Testing the robot accuracy

To assess the accuracy of dilution and dispensing steps performed on the Biomek® robot, the program designed for protein concentration titration (detailed in Section 2.3.11.1.3) was run, with a low molarity solution of Alexa Fluor® 488 dye, a very sensitive fluorophore, as test sample. The resulting microplate was centrifugated briefly at 500x g (2min at 4ºC) and a single measurement of all reservoirs was collected on a BMG Omega microplate reader, at excitation and emission wavelengths of 480nm and 520nm, respectively. As *per* the program's design, eight concentrations of sample with ten replicate wells were prepared by sequential dilutions. Of these, five representative replicas at each concentration are plotted in Fig. 4.15 as single bars.
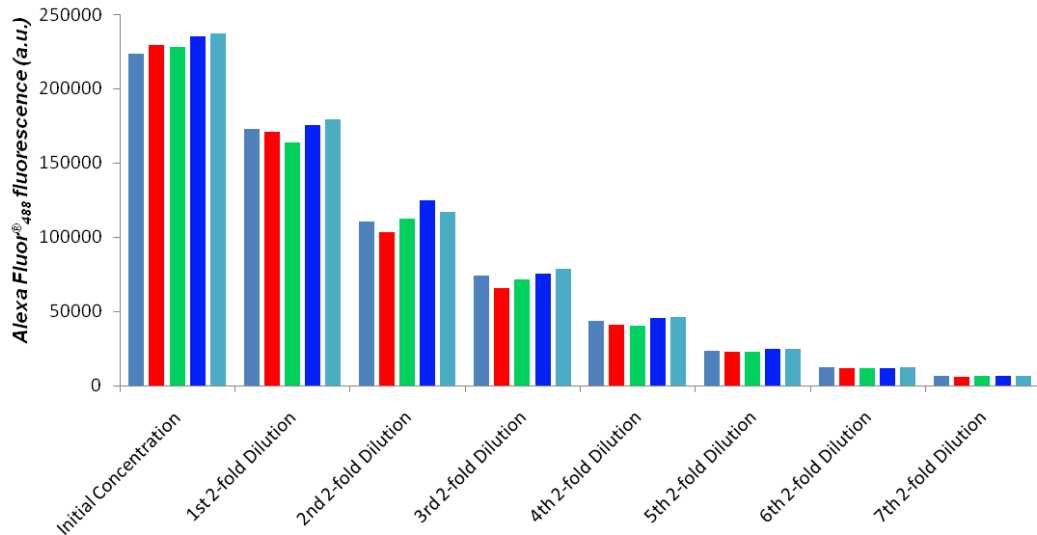
Fig. 4.13 – Accuracy of the dilutions and dispensing steps performed by the Biomek® robot running the protein concentration titration program designed. Five representative replicates for each of the eight concentrations tested are shown as single bars. Each concentration is related to the previous by a 2-fold dilution.

The measurements collected show comparable values between replicate reservoirs, as well as a visible linear titration of signal for the five highest sample concentrations (Fig. 4.13), indicating accurate dilution and dispensing by the robot. At the three lowest concentrations, however, when signal intensity recorded drops below 50000 fluorescence units (a.u.), the linearity of signal titration is lost, indicating a lower limit for the sensitivity of the plate reader. Due to this result, in all subsequent experiments a fluorescence count above this limit was maintained (by adjusting the gain settings on the plate reader), to ensure sensitivity and reproducibility.

4.4.2.4 Preliminary protein concentration test

After designing and programming a protein concentration titration method for the Biomek® robot, and validating its accuracy in diluting and dispensing sample, the next step was to prepare a preliminary aggregation trial with AT3[U3/14Q] in TXBa. Monomeric protein stock was buffer exchanged into TXBa buffer and used in

an aggregation trial set-up using the automated strategy described (detailed in Section 2.3.11.1.3). Protein concentrations in the microplate ranged from 50µM to 0.39µM, in sequential 2-fold dilutions, each with 10 replicate wells. The microplate set-up by the Biomek® robot was sealed and centrifugated at 500x *g* for 2min (at 4ºC), after which it was placed in the BMG Omega fluorimeter, previously equilibrated at 37ºC. Quiescent aggregation of AT3[U3/14Q] at 37ºC was monitored by the fluorescence change of 20µM ThioT over time (as detailed in Section 2.3.11.1.3).

To enable a simple direct comparison of results across initial protein concentration, the data was Y-axis normalized (as detailed in Section 2.3.12.2.1). Only six of the eight concentrations prepared are shown in Fig. 4.14, as the other two had not, after ~42h incubation, reached signal intensity above the previously determined threshold (50000 arbitrary fluorescence units). The traces for the lowest concentration shown (Fig. 4.14 – 1.56µM protein) reveal profiles that, at the end of measurements, had only just begun to hint the end of the elongation phase.
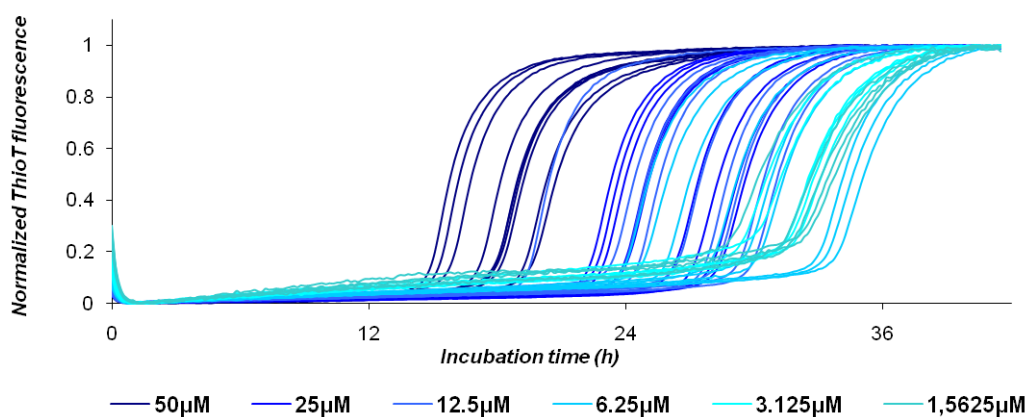
Fig. 4.14 – **Preliminary protein concentration titration of AT3[U3/14Q], performed using the designed program for the Biomek® robot. Monomeric sample in TXBa was titrated from 50µM to 1.56µM initial protein concentration, and incubated quiescently at 37ºC. The progress of aggregation was monitored by the fluorescence changes of 20µM ThioT. The experiments were measured on a BMG Omega plate reader with excitation and emission filters for ThioT: 440nm and 480nm, respectively.**

Nonetheless, it is clear that traces at a given concentration cluster together, and that the aggregation kinetics of aggregation titrate with protein concentration,

namely the lag time length. As the number of protein molecules in solution increases, the lag phase is shortened. The overall quality of the data is excellent and allows detailed analysis of the kinetics of aggregation to be undertaken, as discussed in the following chapters (Chapter 5 and Chapter 6).

4.4.3 Retesting data reproducibility

After bypassing the sample handling variability (as above), the inherent reproducibility of AT3[U3/14Q]'s quiescent aggregation was re-evaluated. Three purified protein batches were prepared (as detailed in Section 2.3.6) and processed for aggregation trials (as detailed in Section 2.3.11.1.4) to compare the quiescent aggregation properties of AT3[U3/14Q] at 37ºC as described above. The results (Fig. 4.15) showed significantly reduced variation in the aggregation kinetics (compare with Fig. 4.9), despite the use of different purification batches, falling now in an acceptable range for further detailed analysis.
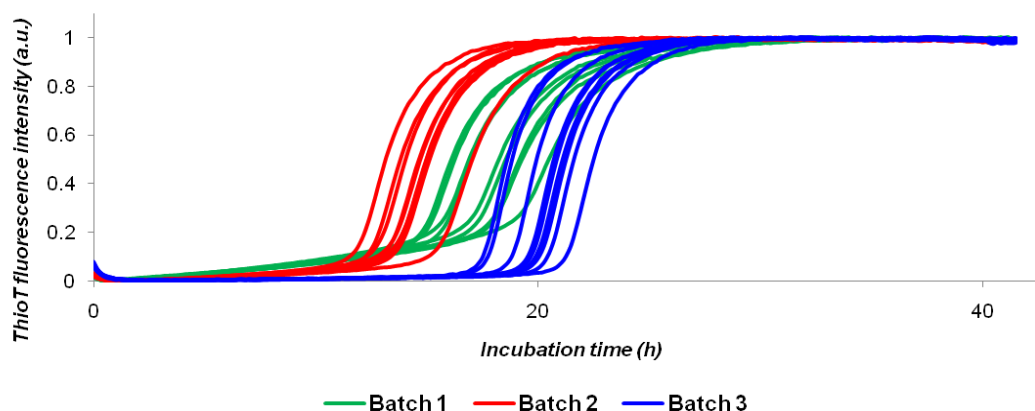
Fig. 4.15 – Reproducibility of quiescent AT3[U3/14Q] aggregation. Replicate experiments with aliquots from the three protein purification batches were prepared for aggregation studies, at 30μM initial protein concentration. Reaction mixtures were incubated at 37ºC and the aggregation progress was monitored by the fluorescence intensity changes of 10μM ThioT. Batches are coloured in green (1), red (2) and blue (3). The experiment was conducted on a BMG Omega plate reader, equipped with excitation and emission filters for ThioT: 440nm and 480nm, respectively.

4.5 Conclusion

Environmental factors are known to have a marked effect on the aggregation of proteins (Gosal, 2005; Nichols, 2005; reviewed in Wang, 2010). For this reason, and in order to find the simplest buffer system that yields reproducible aggregation kinetics, much care was taken into the design of a quiescent aggregation buffer for ataxin-3 studies. Following on work by Gales and co-workers (Gales, 2005), the influence of buffer components on the kinetics of non-expanded ataxin-3 (MJD 1N) aggregation was investigated, bearing in mind that physiologically-relevant conditions were aimed for.

The work presented here details insights into the effect of different variables on AT3[U3/14Q] aggregation under physiologically-relevant conditions. A reducing environment, as the cellular environment typically is, shows a clear influence on aggregation. The full details of this influence were not studied, as they were not the goal of this section, but they certainly dictated the preservation of the reducing agent (DTT) in the aggregation buffer, as described initially in Gales, 2005. For the same reason, the influence of divalent cations, viewed indirectly by their scavenging with EDTA, was not studied further, even though results pointed to an effect. This would not be a surprising find, as other groups have published work on the influence of some divalent cations on ataxin-3 aggregation (Richelli, 2007 and Stawoska, 2009). It would certainly be interesting to study the influence of divalent cations such as Mg^{2+} and Ca^{2+}, for their common occurrence in cells but, for now at least, EDTA was kept in the final aggregation buffer, at the same concentration as published by Gales and co-workers (Gales, 2005), to minimize variability in aggregation kinetics by the presence of such ions. The most surprising finding on the influence of buffer components in ataxin-3 aggregation was found for glycerol. This polyol has a marked influence on the aggregation kinetics of AT3[U3/14Q], keeping clearly defined lag and elongation phases. This reagent was therefore removed from the aggregation buffer. An addition to the published buffer was made, with the biocide sodium azide added at 0.01% (w/v) for its known ability to

avoid microbial contamination. This reagent had little, if any, effect on the aggregation of non-expanded ataxin-3 (MJD 1N) under the conditions employed. The influence of the chemical probe to monitor aggregation, ThioT, was also evaluated and shown to be low or non-existent for ThioT concentrations between 10 and 40µM.

The aggregation protocol initially followed in this thesis, as published in Gales, 2005 to study ataxin-3 aggregation, was found to yield poorly reproducible results, crucial for a detailed analysis of the aggregation mechanism. Two key points were addressed to tackle the reproducibility problem: variability in buffer preparation; variability in sample handling. Of these, only sample handling was shown to have an effect on the quiescent aggregation of ataxin-3 monitored, bypassed by the development of an automated strategy for the dilution and dispensing steps of an aggregation trial. Additionally, the use of this automated strategy allows more samples to be assayed simultaneously, allowing in turn for a more accurate comparison of results. Furthermore, the lower volume of sample used, through the use of 384 well microplates of reduced volume, maximizes the potential for aggregation trials. The aggregation kinetics resulting from the use of a robotic strategy are of good quality and reproducibility, although the latter was not fully achieved. Nevertheless, an increase in the number of replicates assayed compensates for the variability. The automated aggregation protocol detailed in this chapter (and in Section 2.3.11.1.3) was used as a tool in describing and characterizing the quiescent aggregation of ataxin-3 in the chapters that follow.

Chapter **5**

5 – Characterizing AT3[U3/14Q] aggregation

5.1 Introduction

As previously mentioned in Chapter 1, many human pathologies have been shown to develop *via* an ordered protein aggregation mechanism, into amyloid or amyloid-like aggregates, whether intra or extracellular in localization. For some of these pathologies, *ex vivo* fibrils have been isolated from patients and shown to be similar to those produced *in vitro* (Hirschfield, 2003 and Alexandrescu, 2005). Initially, the discovery of fibrillar deposits in patients led to the idea that they were the causative agents of disease (reviewed in Hardy and Selkoe, 2002). More recently, soluble precursors of fibrillar structures have been proposed to be the culprits of toxicity, sharing an epitope that is recognized by the A11 antibody across several disease-associated protein aggregates (Kayed, 2003). However, the cells affected in pathological conditions are often specific, even with ubiquitously expressed proteins, stressing the need to address each pathological condition and related protein individually (reviewed in Havel, 2009 and Takahashi, 2010).

In the MJD/SCA3 pathology the identity of the toxic species is unknown, primarily due to an incomplete description of the aggregation pathway. The disruption of homeostasis, however, is certainly related to an expansion of CAG repeats in the coding region of the *ATXN3* gene, translated to an expansion of polyglutamines in the protein ataxin-3 (Stevanin, 1994). Proteins with polyQ expansions such as ataxin-3 are of particular interest in aggregation because of the possible additional array of side-chain hydrogen bonds, which may lead to different morphologies and stability (Esposito, 2008; Berryman, 2009; Natallelo, 2011). Interestingly, work by Ellisdon and co-workers has shown that the polyQ region of ataxin-3 is not directly involved in the early steps of aggregation, pointing to the remaining domains and regions of the protein as responsible for early interactions. Accumulating evidence has indicated the importance of the flanking regions to the polyQ stretch in modulating the aggregation of polyQ-containing proteins (Ellisdon, 2006; Thakur, 2009; Williamson, 2010) and engineered proteins constructs (Duenvald, 2006; Robertson, 2008; Dougan, 2009), strengthening the notion that

canonical aggregation-prone sequences are not the sole responsible regions in directing protein self-association. Simultaneously, these results underline the importance of studying non-pathological protein aggregation for the understanding of the molecular mechanisms of disease progression and pathology.

The molecular events in early protein assembly steps are poorly understood, due largely to the heterogeneity and complex nature of association events. Protein polymerization is thought to occur in one of two ways: a *downhill reaction*, in which each structure formed has a lower free energy than its predecessors (without involving energy barriers), or a *nucleated reaction* where energetically unfavourable steps bottleneck the formation of larger structures (see Section 1.2.1). Amyloid formation typically shows many of the characteristics of a nucleated mechanism, with an initial "quiescent" step (lag phase), followed by an exponential growth (elongation phase), until equilibrium is reached (stationary phase). The lag phase is thought of as representing the overcoming of unfavourable steps, leading to the formation of a critical nucleus, a transiently populated species whose formation tips the thermodynamical equilibrium towards polymerization and dictates a sharp increase in the rates of polymerization. Bypassing the nucleation step, by addition of preformed seeds or altering the environmental conditions (see below), allows a thermodynamical equilibrium to be reached more rapidly.

Several factors have been reported to influence the rate of aggregation of amyloidogenic peptides and proteins and can be divided into intrinsic and extrinsic (reviewed in Wang, 2010). The interaction of a given protein with cellular components (chaperones, proteases and other species, not necessarily of proteinaceous nature), protein clearance pathways through quality control mechanisms (autophagy or the ubiquitin proteasome system) and physicochemical parameters defining the environment, constitute **extrinsic** factors. Of the latter, pH, temperature, ionic strength and peptide or protein concentration are the most studied *in vitro*. On the other hand, parameters such as charge and hydrophobicity (net or in localized patterns), along with the propensity of the primary structure to adopt specific secondary structure motifs, are known **intrinsic** factors. Cross-talk between both sets of factors is expected, and even under the same environmental

conditions, a multitude of aggregate morphologies can be formed simultaneously, highlighting the complexity and multiplicity of aggregation pathways (Walsh, 1997 and Chen, 2001; Gosal, 2005; Shtainfeld, 2010). A simplified aggregation system to study, such as non-pathological ataxin-3 (bypassing the effect of the glutamine expansion), provides valuable insights not just for MJD/SCA3 pathology, but also other protein aggregation pathologies, namely other polyglutaminopathies.

Understanding the factors influencing aggregation rates will increase the effectiveness of therapeutic strategies and/or the search for them. However, to fully characterize the aggregation process of a protein, all conformational states and oligomeric structures formed need to be identified, as well as the kinetic and thermodynamic rates determined, a daunting task. In addition, residues or regions of the polypeptide sequence involved in promoting the various aggregation steps need to be described and characterized, in order to understand the assembly routes and ultimate aggregate morphology and properties. While the identification of ataxin-3's regions involved in early assembly events will be discussed in the next chapter (Chapter 6), the goal for the present chapter was to establish the characteristics of AT3[U3/14Q] aggregation, providing a clearer understanding of the mechanism and establishing a basis of comparison for subsequent studies. The influence of environmental parameters such as pH, ionic strength and protein concentration on AT3[U3/14Q] aggregation kinetics was evaluated, and mechanistic parameters determined, such as the sequence of structures formed on-pathway to fibril formation and the identity of the thermodynamical nucleus estimated.

5.2 Describing the assembly process

The aggregation progress of AT3[U3/14Q] shows lag-dependent kinetics, where an initial monomer population progressively aggregates into soluble higher molecular weight oligomers, later assembled further into fibrillar species (Gales, 2005 and Section 3.4). Characterization of the structures formed during the

aggregation progress of a protein is a technical challenge. The end-products of aggregation are typically too large to be analyzed using solution NMR spectroscopy and they are not crystalline, voiding the option of beam diffraction for a molecular characterization. On a morphological level, however, direct visualization of the assembly products has proven helpful in distinguishing reaction pathways and describing the progression of structures formed by aggregating protein, such as described for β$_2$m (Gosal, 2005), Htt (Dahlgren, 2005), α-Syn (Giehm, 2011) and ataxin-3 (Ellisdon, 2006).

The aggregation progress of AT3[U3/14Q] under physiologically-relevant conditions was followed by negative-stain electron microscopy (as described in Section 2.3.11.3.2). Two initial protein concentrations were studied, 100μM and 25μM, both to ensure visualization of scarcely populated aggregates and an uncluttered view of the sample (respectively). Briefly, at consecutive time-points, 10μl of sample incubated at the initial protein concentrations noted were taken from reaction mixtures and diluted 10-fold with filter-sterilized RO-purified water, in order to prepare sample grids (as *per* Section 2.3.11.3.2). Sample grids prepared were visualized on a Philips CM-10 electron microscope fitted with a photographic camera, allowing the images below to be collected (Fig. 5.1 – **c** to **j**). Concomitantly, the aggregation progress of replicate aggregating samples was followed by ThioT fluorescence, as before (Section 2.3.11.1.4).

Morphologically, the end-stage products at both protein concentrations tested are fibrillar, with a worm-like appearance and heterogeneous length distribution. The aggregate width, however, varies much less, from 7 to 10nm (Fig. 5.1 – **f** and **j**). Analyzing the negative-stain images collected from samples at the lowest initial protein concentration reveals a sequence of structures formed over time, where an initial population comprised mainly of small spherical aggregates (Fig. 5.1 – **g**) precedes a population of elongated structures (Fig. 5.1 – **h** and **i**), referred to in literature as "protofibrils" (Ross & Poirier, 2004). These, in turn, grow further in length with the onset of the elongation phase, and by the stationary phase only a small percentage of structures present are protofibrillar, the majority being worm-like fibrils (Fig. 5.1 – **j**).
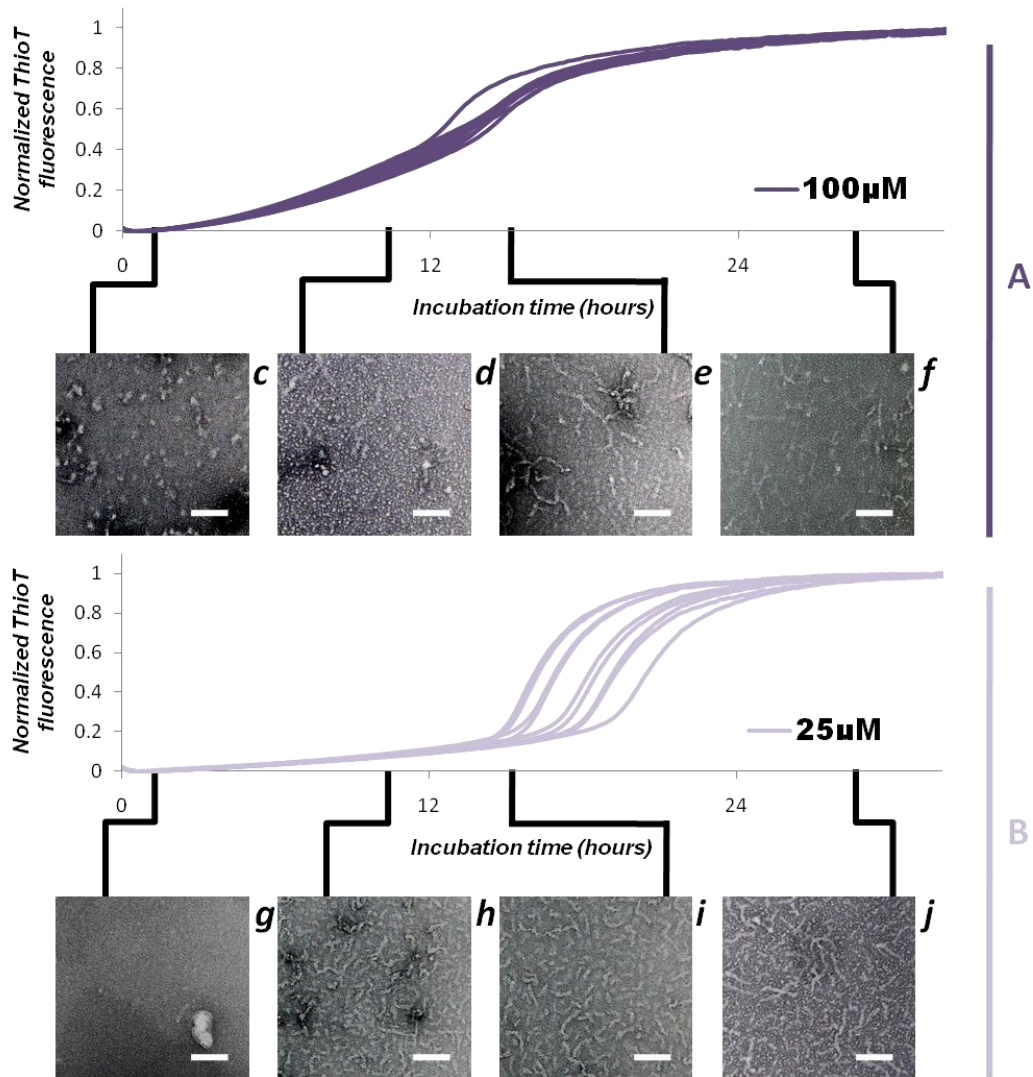
Fig. 5.1 – AT3[U3/14Q] aggregate formation over time. Monomeric protein sample at two initial concentrations, 100µM (A) and 25µM (B), was incubated quiescently at 37ºC in TXBa buffer. The progress of the polymerization reaction was followed by ThioT fluorescence and direct visualization of the structures formed by negative-stain electron microscopy (c to f and g to j). The bar on the microscopy images represents 100nm.

Results at the highest initial protein concentration (Fig. 5.1 – A), although with faster kinetics, follow an identical trend, with two major deviations: the end-stage products (Fig. 5.1 – **f**), although of similar widths, are longer than those formed at lower protein concentration (Fig. 5.1 – **j**); the first timepoint visualized, immediately after initiating fluorescence measurements of replicate samples, shows structures, approximately spherical (Fig. 5.1 – **c**), with a diameter (~25nm) that is larger than the width of protofibrillar (Fig. 5.1 – **d**) and fibrillar structures

formed (Fig. 5.1 – **e** and **f**), suggesting they are not direct precursors to the latter. These are visible at all timepoints earlier than the stationary phase.

5.3 Nucleated reaction

Worm-like fibrils formed by the assembly of proteins such as β_2m, TTR and β-lactoglobulin, result from a *downhill*, non-nucleated assembly (Modler, 2003, Gosal, 2005 and Smith, 2006). However, Ellisdon and co-workers have shown that the assembly of ataxin-3 (2UIM isoform with a 15Q tract) under physiologically-relevant conditions, results in worm-like fibrils through a nucleated process (Ellisdon, 2007). Mechanistically, the addition of preformed aggregates to a nucleated polymerization reaction reduces or removes the lag phase (known as "seeding" the reaction), whereas a non-nucleated reaction would remain unaltered (see Fig. 1.4).

To determine whether AT3[U3/14Q], under the conditions determined in Chapter 4, follows a nucleated or non-nucleated process, seeded reactions at 20μM protein were prepared. Briefly, end-products from a quiescent aggregation reaction of 20μM AT3[U3/14Q] in TXBa buffer (as described in Section 2.3.11.1.4), were collected after 36h incubation at 37ºC (Fig. 5.2) and identified with negative-stain electron microscopy, as above (detailed in Section 2.3.11.3.2). These products, seeds, were then added at varying volume percentages (as detailed in Section 2.3.11.2), to freshly prepared reaction mixtures of monomeric AT3[U3/14Q] at the same initial protein concentration and incubation conditions (Fig. 5.3), keeping the total protein mass constant across all reactions.
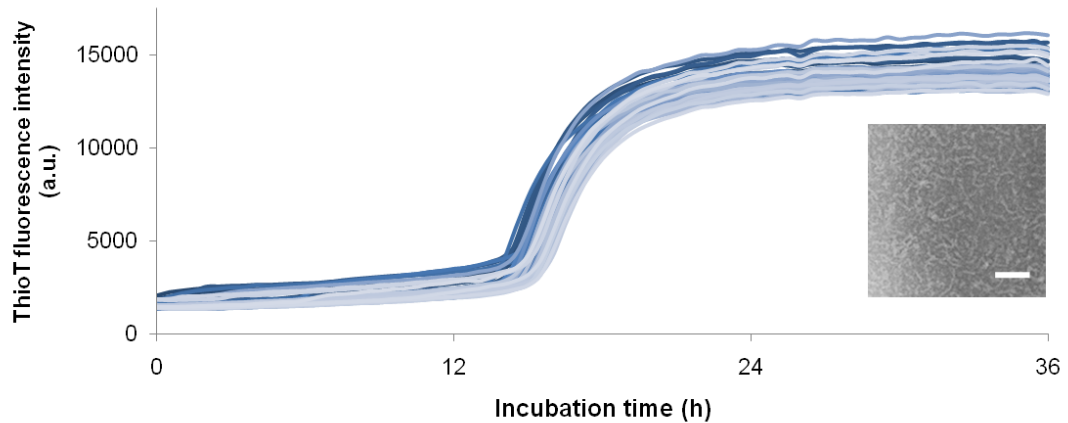
Fig. 5.2 – Production of end-stage ataxin-3 aggregates. 20µM monomeric AT3[U3/14Q] in TXBa buffer was incubated at 37ºC. Aggregation progress was monitored by ThioT fluorescence. Sample after 36h incubation was collected and visualised by negative-stain electron microscopy (inset), showing the presence of worm-like fibrils. The bar on the microscopy image represents 100nm.

The results in Fig. 5.3 conclusively show that the assembly of AT3[U3/14Q] under physiologically-relevant conditions follows a nucleated process, consistent with results for the 2UIM isoform of the protein (Ellisdon, 2007). A clear effect of seeding in reducing the lag phase length is visible, and titrates with the amount of preformed aggregates added at the start of the reaction. However, the reduction of the lag phase isn't complete even when supplementing the reaction with as much as 30% (v/v) preformed aggregates. Interestingly, not only are the lag times reduced by the addition of preformed aggregates, but also the lag phases progressively gain an upward slope as the amount of preformed aggregates added increases. The final products formed from seeded and unseeded reactions are identical (Fig. 5.2 and 5.3 – insets), as expected for a nucleated assembly process.
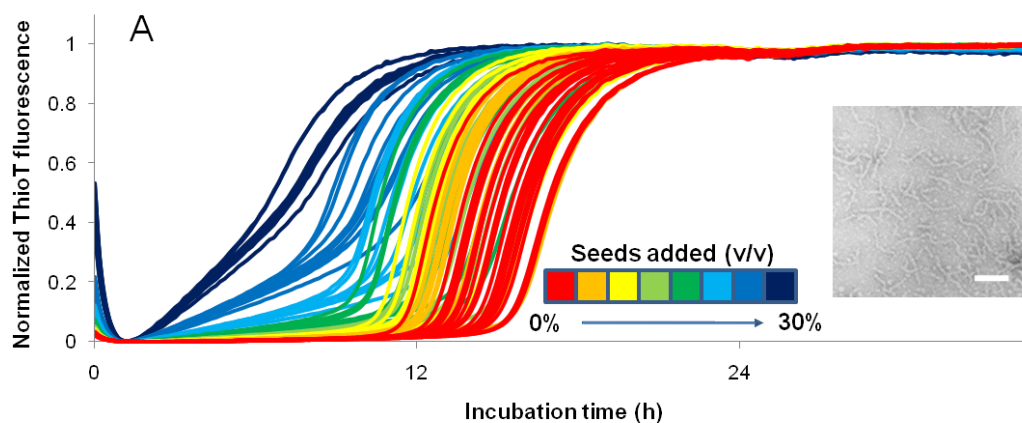
Fig. 5.3 – Seeding AT3[U3/14Q] aggregation. Reaction mixtures of 20µM monomeric AT3[U3/14Q] were prepared in TXBa buffer and supplemented with varying volumetric percentages of AT3[U3/14Q] seed sample, formed under identical conditions, keeping the total mass of protein per reaction mixture identical. Quiescent aggregation at 37ºC was monitored by ThioT fluorescence, as before. Unseeded reaction mixtures were also prepared as control (shown in red). Aggregates from seeded reactions with 30% (v/v) end-stage sample were visualised by negative-stain electron microscopy – inset. The bar on the microscopy image represents 100nm.

5.4 Environmental parameters

As previously mentioned (Sections 1.2.3 and 5.1), environmental parameters affect the rates of polymerization and a study of the effects of pH, ionic strength and protein concentration can therefore provide important clues on the mechanism of ataxin-3 aggregation and how it could be influenced by the cellular environment. In addition, from an *in vitro* study standpoint, it is important to determine the conditions boundaries for phase change in protein assembly, in order to properly analyse the results obtained in light of the possible assembly pathways (see Sections 1.2.2 and 1.2.3).

5.4.1 pH

To assess the influence of pH on the assembly process of AT3[U3/14Q], varying pH values were tested. The cytoplasm typically has a pH around 7.4 but, depending on the cellular compartment the pH can be as low as 4.0 (in lysossomes). In addition, unpublished data from the Macedo-Ribeiro group (Sandra Macedo-Ribeiro, University of Porto, personal communication) indicates that AT3[U3/14Q] solubility decreases substantially as the pH is lowered, particularly below pH 6.0 as the pH nears the isoelectric point of the protein (at pH 4.9). As such, monomeric AT3[U3/14Q] assembly was studied at pH values between 7.5 and 6.0, under comparable conditions. Frozen monomeric sample was thawed, centrifugated for 30min at 16000x *g* and the supernatant divided into 100ul aliquots. Each was next buffer exchanged into one of the buffers listed on Table 5.1, the final protein concentration set by dilution with the corresponding buffer and a quiescent microplate aggregation assay at 37ºC monitored by the fluorescence intensity of 10μM ThioT (as detailed in Section 2.3.11.1.1). It is important to note that for pH values lower than 7.0, HEPES was replaced by MES as the buffering molecule, since below pH 7.0 the former loses its buffering capacity. Both HEPES and MES are part of Good's list of biologically inert buffers (Good, 1969).
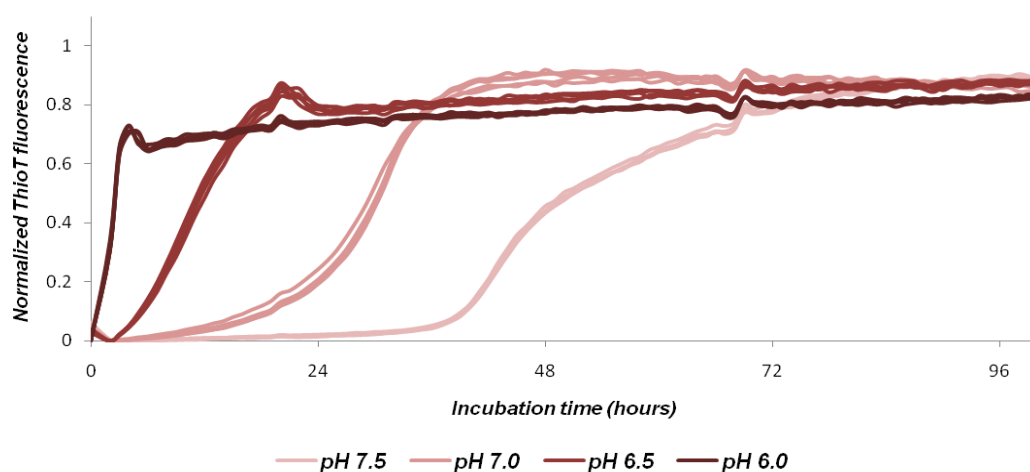
Fig. 5.4 – Aggregation kinetics of AT3[U3/14Q] under varying pH conditions. Monomeric protein sample at 10μM concentration was placed in buffers with varying pH between 7.5 and 6.0 (detailed in Table 5.1) and a quiescent aggregation experiment at 37ºC initiated, monitored by ThioT fluorescence.

Table 5.1 – Composition of the varying pH buffers used

pH set	Buffer composition
7.5	20mM HEPES pH 7.5, 200mM NaCl, 1mM EDTA, 1mM DTT, 0.01% (w/v) NaN$_3$
7.0	20mM HEPES pH 7.0, 200mM NaCl, 1mM EDTA, 1mM DTT, 0.01% (w/v) NaN$_3$
6.5	20mM MES pH 6.5, 200mM NaCl, 1mM EDTA, 1mM DTT, 0.01% (w/v) NaN$_3$
6.0	20mM MES pH 6.0, 200mM NaCl, 1mM EDTA, 1mM DTT, 0.01% (w/v) NaN$_3$

The aggregation kinetics of AT3[U3/14Q] are markedly influenced by the solution's pH, as Fig. 5.4 shows. The lag phase is severely reduced as the pH is lowered, down from ~36h length at pH 7.5, to an undetectable length at pH 6.0, accompanied by a tendency for an upward slope during this phase. In addition, the aggregation profiles obtained at pH 6.5 or lower show a curious, though apparently characteristic, overshoot in the transition between the elongation phase and the stationary phase. The molecular origins of this characteristic growth feature remain unsolved. Interestingly, replicate samples are tightly distributed at all pH values tested, indicating that the effect seen is not, at least entirely, ruled by stochastic processes.

5.4.2 Ionic strength

Another of the physicochemical parameters known to have a major influence on protein aggregation kinetics is ionic strength, shown clearly for proteins such as β$_2$m and β-lactoglobulin (Gosal, 2005 and Majhi, 2006, respectively). Its effect, although well documented, is poorly understood at a molecular level due to its complexity, but disruption of electrostatic forces by ions in solution can affect protein stability and solubility, as well as protein-protein interactions (reviewed in Wang, 2010).

To assess the effect of ionic strength on the aggregation kinetics of AT3[U3/14Q], frozen monomeric protein stock was thawed and centrifuged as

before, and buffer exchanged to 0M NaCl buffer, listed on Table 5.2. To establish a titration of ionic strength, protein sample in 0M NaCl buffer was aliquoted and diluted to 20µM final protein concentration, with appropriate combinations of both buffers listed on Table 5.2, preparing reaction mixtures at final NaCl concentrations ranging from 400 to 0M NaCl. These sample aliquots were next prepared for a microplate aggregation assay as before (detailed in Section 2.3.11.1.2) and incubated quiescently at 37ºC. Aggregation progress was monitored by ThioT fluorescence.

Table 5.2 – Composition of the varying NaCl concentration buffers used

NaCl Concentration	Buffer composition
1M	20mM HEPES pH 7.5, 1M NaCl, 1mM EDTA, 1mM DTT, 0.01% (w/v) NaN$_3$
0M	20mM HEPES pH 7.5, 1mM EDTA, 1mM DTT, 0.01% (w/v) NaN$_3$
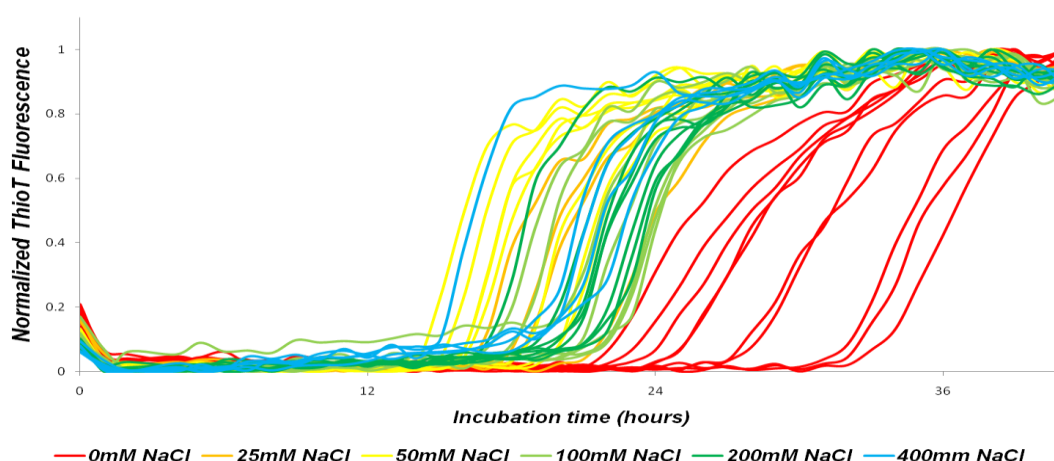
Fig. 5.5 – Aggregation kinetics of AT3[U3/14Q] under varying NaCl concentration conditions. Monomeric protein sample at 10µM concentration in 20mM HEPES pH 7.5, 1mM EDTA, 1mM DTT and 0.01% (w/v) NaN$_3$, supplemented with varying NaCl concentrations, from 0mM to 400mM, was incubated quiescently at 37ºC and aggregation progress monitored by ThioT fluorescence.

In the absence of NaCl, the aggregation kinetics are slower than in the presence of the salt (Fig. 5.5), with reactions showing longer lag phases and lower apparent elongation rates. However, the effect does not titrate with NaCl concentration, with all concentrations tested between 25mM and 400mM of the salt displaying comparable aggregation kinetics, i.e., lag phases length vary within

the same interval, corresponding to the spread of results, and apparent elongation rates are similar. The variability of lag phase length between replicate samples is higher than previously seen (see Fig. 5.4), for all NaCl concentrations tested, suggesting a greater influence of stochasticity in ionic strength's effect. Interestingly, the spread of results along the time dimension lowers with increasing concentration of NaCl tested, suggesting a tighter control over the nucleation mechanism in the presence of the salt. In summary, there is a clear effect on the aggregation kinetics of AT3[U3/14Q] by the ionic strength of the environment, whereby concentrations between 25mM and 400mM NaCl display faster reaction rates than those seen in the absence of the salt. A more complete characterization of the effect was not attempted, as it was outside the scope of this thesis. For practical purposes, there is no transition of phase on or around 200mM NaCl.

5.4.3 Protein concentration

A well studied and markedly influential environmental parameter of protein aggregation is the initial protein concentration. As an assembly reaction progresses, the loss of entropy from polymerization (monomers immobilized as aggregates) slows the aggregation rates, unless the effect is balanced by an increase of enthalpy (from contacts within the aggregates). When the initial protein concentration is high, the loss of entropy is not felt as strongly, which then allows for faster aggregation rates and reduced lag time length. Given the previously shown tendency for variability in aggregation kinetics of AT3[U3/14Q] (discussed in Section 4.4), and even though the observed variability was reduced, sufficient data had to be collected to allow accurate analysis. For this reason, three sets of experiments in which protein concentration was titrated were performed as described in Section 4.4.2.4 (detailed in Section 2.3.11.1.3), spanning two orders of magnitude of protein concentration each. The quiescent aggregation progress of AT3[U3/14Q] at 37ºC in TXBa buffer was followed monitoring the fluorescence changes of 20μM ThioT.

As the comparison between experimental datasets shown in Fig. 5.6 indicates (and as discussed in section 4.4), the aggregation kinetics of AT3[U3/14Q] display some variability. Nevertheless, the accumulated data from three datasets, with a total of 240 individual reactions, enables the analysis of relevant aggregation parameters, despite this variability. A qualitative look at the results in Fig. 5.6 reveals a clear tendency for lag-dependent kinetics, as previously reported in Chapters 3 and 4 of this thesis. A lag phase is apparent in all reactions, followed by a sharp increase in fluorescence as the elongation phase is initiated (Fig. 5.6), eventually levelling off as the system reaches equilibrium in the stationary phase. At higher protein concentration some reactions show a tendency towards an upward slope where a flat lag phase would be expected, an effect that is dependent on the purification batch and is attenuated as the concentration of protein is lowered. This observation suggests the existence of prenuclear aggregates, on- or off-pathway to fibril formation, in the starting sample of these purification batches and undetected during the purification procedures. These aggregates are ThioT-positive, resulting in a progressively increasing signal over the lag phase length at higher initial protein concentration. They do not, however, dramatically interfere with the aggregation mechanism, since the kinetics observed at lower concentrations are matched across protein batches and resemble previous experimental datasets (compare Fig. 4.14 and Fig. 5.6).

Further analysis of the results was performed, following the methodology developed by Xue and co-workers, to extract aggregation parameters from ThioT-monitored experimental kinetics (Xue, 2008). With the normalised progress curves shown in Fig. 5.6 as a starting point, both the midpoint of the reaction (t_{50}) and the elongation rate for each reaction were extracted, along with the length of the lag phase (as detailed in Section 2.3.12.3).
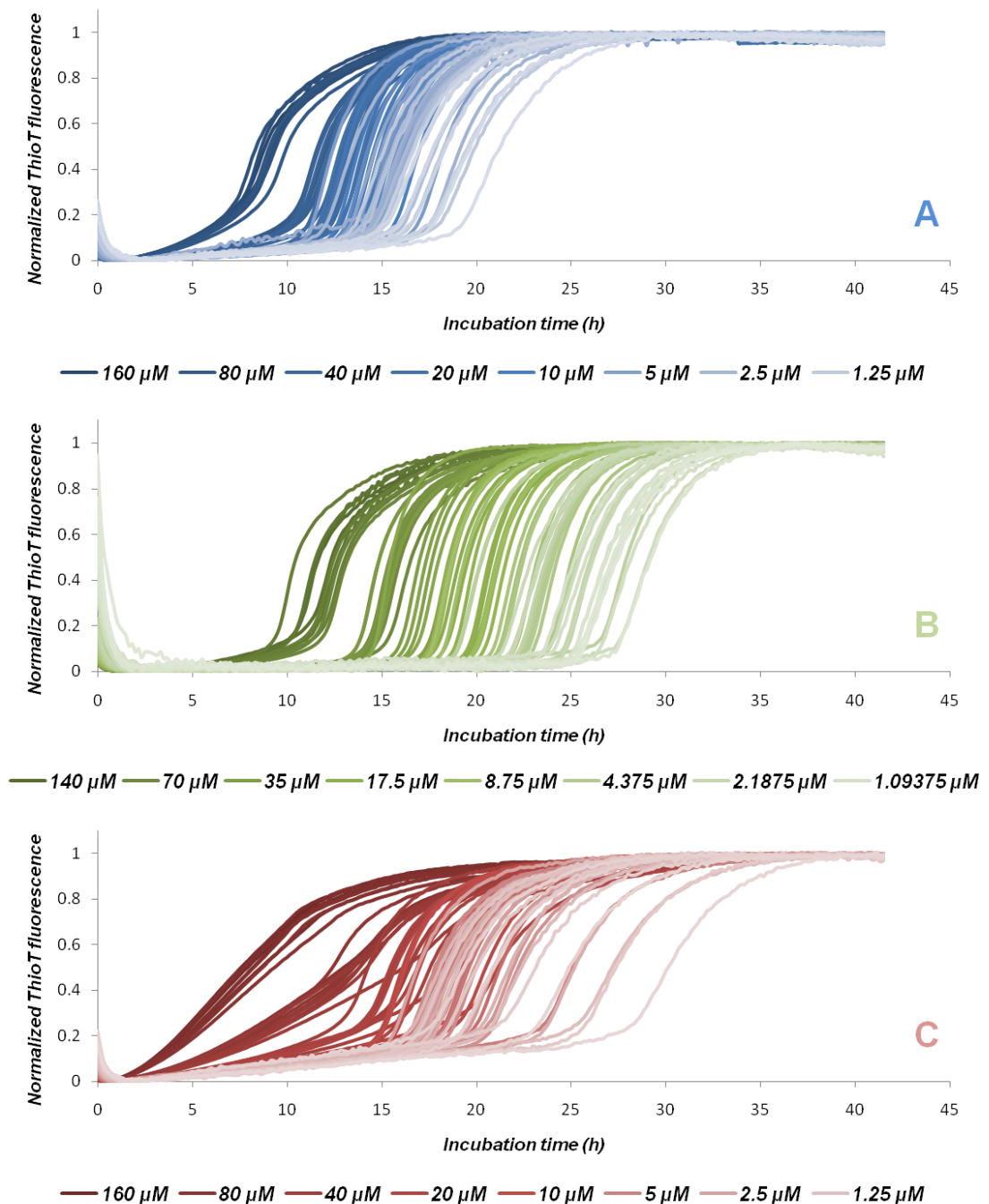
Fig. 5.6 – Influence of protein concentration on the quiescent aggregation kinetics of AT3[U3/14Q] at 37ºC in TXBa buffer. Monomeric sample from three separate purification batches (A, B and C) was dispensed and diluted using the automated strategy developed in Section 4.4.2. Eight initial monomer concentrations were tested for each purification batch, spanning two orders of magnitude, with ten replicates each, in a total of 240 individual aggregation reactions. Results shown are Y-axis normalized.

Relating extracted parameters with protein concentration individually, in log-log plots, allows a better understanding of the aggregation progress and its

phases at varying initial protein concentrations. Comparing the changes in lag phase length and elongation rate at varying initial protein concentrations reveals a different response, with a stronger effect by the initial protein concentration felt on the lag phase length, suggesting a greater influence on the nucleation part of the reaction. Fitting the data to a linear function (Fig. 5.7), shows that the effect of protein concentration on the length of lag phase is one order of magnitude higher than the effect on the elongation phase.

$$y = -0.1878x + 1.3916$$

$$y = 0.1006x - 0.6261$$

Fig. 5.7 – Relationship between the initial protein concentration and physicochemical parameters describing the aggregation of AT3[U3/14Q]. Log-log plots are shown relating the initial protein concentration and the length of the lag phase (t_{lag}) (above), or the rate of elongation (k) (below). Equations shown are linear functions fitted to the data at initial protein concentration of 100μM or less, graphically represented as a black line on each plot. Data points on the plot are coloured according to the protein purification batch they were extracted from, blue for Batch A, green for batch B and red for batch C (see Fig. 5.6).

Note that, in analysing the data collected, the higher variability at higher protein concentration poses a problem that was bypassed by fitting only the data for concentrations lower or equal to 100µM. However, from the plots in Fig. 5.7, it is clear that the majority of the reactions follow this trend, the exceptions possibly due to the presence of oligomeric species at the start of the reactions, as noted above.

The dependence on the initial protein concentration for both parameters is, however, relatively shallow, at least when compared to data obtained from the aggregation kinetics of other proteins, namely β_2m (Xue, 2008) (Fig. 5.8). This observation suggests fundamentally different aggregation mechanics, where the effect of protein concentration is either masked by the production of off-pathway aggregates on ataxin-3, or enhanced by the fragmentation secondary reaction of β_2m aggregation, as shown by Xue and co-workers (Xue, 2008). Strategies to assess the influence of secondary processes to the reaction progress have been developed by several authors (Xue, 2008; Ferrone, 2002; Padrick and Miranker, 2002; Ruschak and Miranker, 2005; Ruschak and Miranker, 2009), but were not possible to apply reliably in this thesis due to the variability, albeit reduced, seen for ataxin-3 aggregation kinetics. Unfortunately, the shallow dependence on the initial protein concentration observed also excludes the application of mathematical models to determine the identity of the structural nucleus, an as-of-yet poorly defined theoretical structure that fosters further growth of the aggregates after thermodynamical nucleation (Xue, 2008). These impairments do not, however, exclude the estimation of the identity of the thermodynamical nucleus, as previously performed by Ellisdon and co-workers for the 2UIM isoform of ataxin-3 (Ellisdon, 2007), as well as other authors for proteins such as β_2m (Xue, 2008) and the polyQ-containing Htt (Chen, 2002).
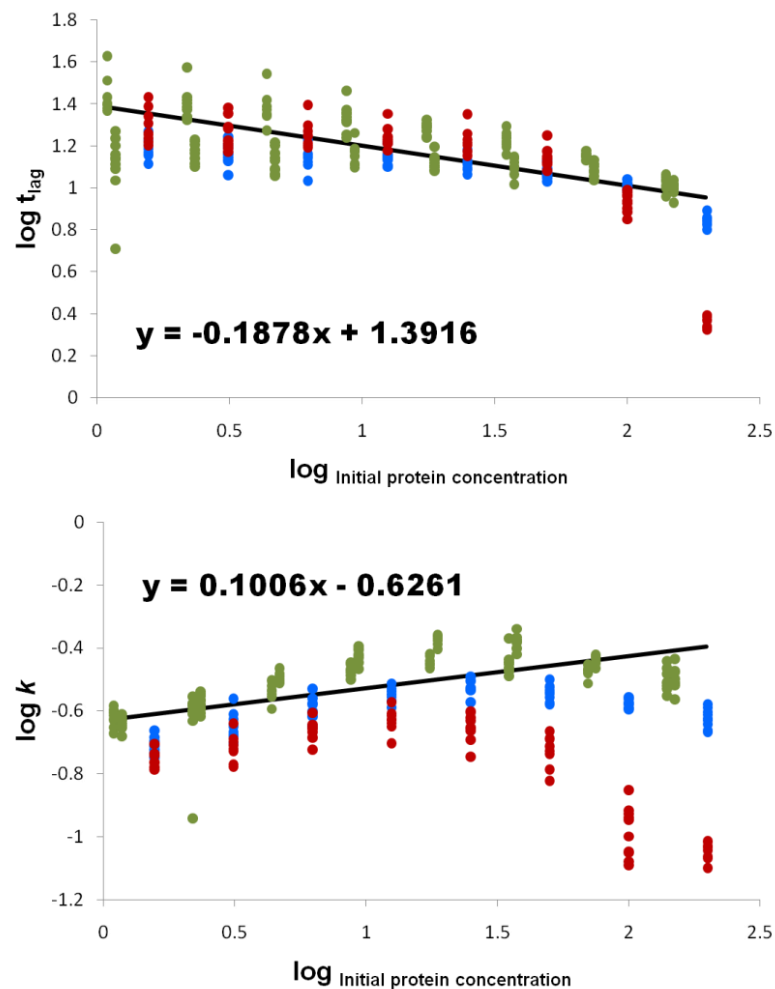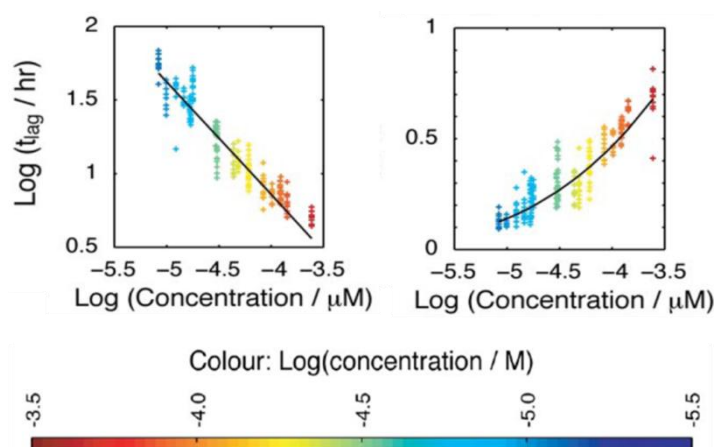
Fig. 5.8 – Relationship between the initial protein concentration and physicochemical parameters describing the aggregation of β$_2$m. Plots are shown relating the initial protein concentration and the length of the lag phase (t$_{lag}$) (left), or the rate of elongation (right). Initial protein concentration was colour coded as shown below. Taken from Xue, 2008.

5.5 Thermodynamical nucleus size

The size of the thermodynamical nucleus can be estimated from the parameters calculated for the previous section (Section 5.4.3). This ephemeral species is known to stipulate the aggregation rate of a protein, as it is simultaneously a requirement for a nucleated polymerization reaction and requires the highest energy barriers in the polymerization reaction to be overcome for its formation. Mathematical approaches to estimate the thermodynamical nucleus size were developed independently by Oosawa and Asakura (the Oosawa method) and Ferrone (the Ferrone method). For the purpose of this thesis, the size of the thermodynamical nucleus in the aggregation of AT3[U3/14Q] was calculated by both approaches from the same dataset (Fig. 5.6 – A), chosen for its flat lag phase baselines and consistent results. For the Oosawa approach however, being that it is less sensitive to the initial timepoints of the monitored aggregation profiles, all reactions of the three datasets shown in Fig. 5.6 were used in the calculations.

5.5.1 the Oosawa method

A practical mathematical approach to calculate the thermodynamical nucleus size of a nucleated polymerization reaction was developed by Oosawa and Asakura, discussed in "***Thermodynamics of the Polymerization of Protein",*** Chap. 4, Academic Press, New York (1975). Briefly, this method correlates the half-time of the reaction (t_{50}) with the initial monomer concentration, yielding a transition describable by a linear function, whose slope, *m*, relates directly to the thermodynamical nucleus size (*n**) by the equation,

$$m = (n^* + 1)/2$$

Equation 5.1

Using t_{50} values calculated from the reactions shown on Fig. 5.6, as mentioned above (detailed in Section 2.3.12.4.1), a log-log plot was constructed as per Oosawa's method (Fig. 5.9).
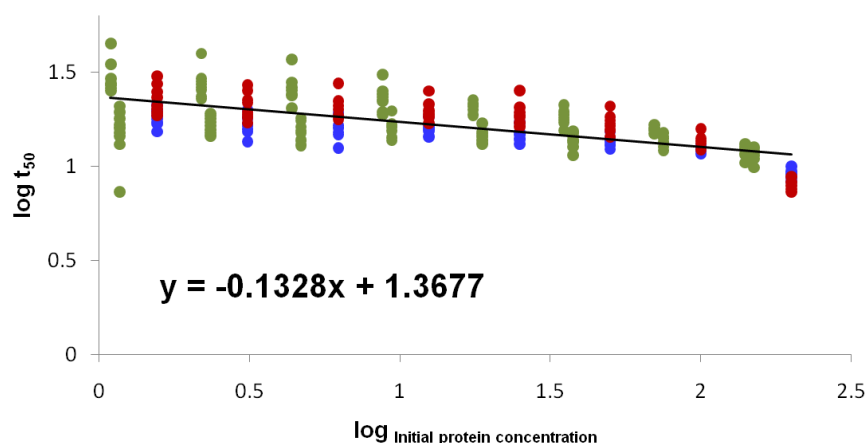
Fig. 5.9 – log-log relationship between the initial protein concentration and the midpoint of the aggregation reaction (t_{50}) of AT3[U3/14Q]. A linear function was fitted to the data and is shown graphically as a black line on the plot, with its corresponding equation also in black (bottom left). The slope of this linear function allows the calculation of the thermodynamical nucleus size by solving equation 5.1.

Using the slope of the linear function fitted to the data (Fig. 5.9 – equation), and solving equation 5.1, $n*$ equals 0.73. Given the mechanistic necessity for a positive integer as the nucleus size, the calculated thermodynamical nucleus size for the aggregation kinetics of AT3[U3/14Q] is rounded up to a monomer.

5.5.2 the Ferrone method

Described by Ferrone in Methods in Enzymology (volume 309, 1999), this mathematical approach has been used to determine the thermodynamical nucleus size for several proteins (Chen, 2002 - huntingtin; Ignatova, 2005 - cellular retinoic acid-binding protein I; Rotter, 2005 – sickle cell heamoglobin), including non-expanded ataxin-3 (2UIM isoform), for which a thermodynamical nucleus size of one monomer was calculated (Ellisdon *et al.*, 2007). Briefly, and unlike the previous approach, the author's interpretation of the polymerization mechanism states that the initial portion of the polymerization reaction is dependent on the amount of nuclei formed, in turn dependent on the size of the nucleus. In practical terms, a polynomial function is fitted to the aggregation data up to 10% completion of the reaction or, alternatively, a linear function to the data plotted with time-squared dependence. It is the slope of this fit that relates to the size of the thermodynamical nucleus, estimated from its log-log relationship with the initial slope against the initial monomer concentration. A linear fit to the data on this double log plot yields a slope, *m,* that relates to the thermodynamical nucleus size (n^*) by the equation,

$$-m = n^* + 2$$

Equation 5.2

According to the approach described (detailed in Section 2.3.12.4.2), the double log plot in Fig. 5.10 was constructed from data extracted as described above, from the ThioT-monitored aggregation kinetics (as detailed in Section 2.3.11.1.3). By solving equation 5.2, a value of n^* equal to -2.268 is calculated. As

before, the mechanistic necessity for a positive integer as the nucleus size dictates a rounded up value of 1, suggesting a monomeric thermodynamic nucleus size for AT[U3/14Q], consistent with results published for non-expanded ataxin-3 (2UIM isoform) by Ellisdon and co-workers (Ellisdon, 2007), using this same approach.

$$y = -0.268x - 1.9313$$
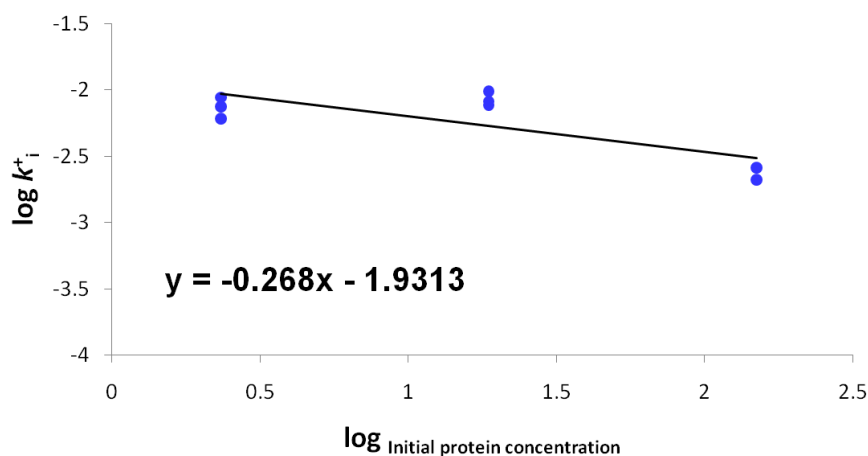
Fig. 5.10 – Estimation of the thermodynamical nucleus size according to the Ferrone method. Calculated initial aggregation rate values (for the first 10% of the reaction completion), are plotted against the initial protein concentration in a log-log relationship. A linear function was fitted to the data, shown graphically as a black line, with the corresponding equation (also in black) shown on the bottom left. The slope of this linear function was used to calculate the size of the thermodynamical nucleus by solving equation 5.2.

5.5 Conclusion

Consistent with reports by Ellisdon and co-workers for the 2UIM isoform of full-length ataxin-3 (Ellisdon, 2007), the recombinant full-length ataxin-3 used in this thesis (AT3[U3/14Q]) aggregates under physiologically-relevant conditions, through a nucleated assembly reaction, seedable with preformed aggregates. A complete ablation of the lag phase was not achieved, even at 30% (v/v) seeds added, suggesting either inefficient seeding or a complex mechanism to produce aggregation-competent species. The thermodynamical nucleus that bottlenecks the

de novo fibril formation process under these conditions was calculated to be a monomer, both by Ferrone's and Oosawa's approaches, once again consistent with the literature (Ellisdon, 2007). It is important to note the deviation of the estimated thermodynamical nucleus size from a positive integer, yielding a positive fractional value with Oosawa's approach and a negative fractional value with Ferrone's. While mechanistically these values need to be rounded to the nearest positive integer (nucleus size cannot be smaller than a monomer), the calculated values may encode additional information, such as described by Powers and Powers (Powers and Powers, 2008), specifically information on secondary nucleation mechanisms, concomitantly present in the system. To this date no reliable processing of results has been proposed to distinguish these effects, reason for which no more conclusions were drawn on this matter.

The formation of the AT3[U3/14Q]'s nucleus is strongly affected by pH, and at pH as low as 6.0 the lag phase for the aggregation kinetics of 10µM AT3[U3/14Q] is completely abolished, suggesting a much faster nucleation process. The effect of pH on a protein is varied and dependent on the protein itself, but partial denaturation and rupture of electrostatic forces, affecting the net charge of a protein or hydrophobicity of residues, have been published (reviewed in DuBay, 2004). More likely than a single effect, a combination of factors such as those listed, accounts for the effect of the environmental parameters on protein aggregation. On the other hand, the effect of ionic strength on the aggregation kinetics of AT3[U3/14Q] is much less pronounced and restricted to low salt concentrations: absence of NaCl in the buffer results in slower aggregation kinetics, though this effect does not titrate with concentrations of NaCl in the buffer higher than 25mM.

Increasing protein concentration has an expected acceleration of aggregation kinetics, though a closer look reveals a greater effect on the nucleation step of the reaction, when compared to the changes in elongation rate observed. Together with the persistence of a lag phase on the aggregation kinetics at high protein concentration (200µM), it seems likely that a considerable energy barrier exists for nucleation under physiologically-relevant conditions. At the same time, the concentration of protein does not affect the morphological aspect of

aggregation, with fibrillar aggregates resulting from low and high concentrations (25µM and 100µM, respectively), of similar appearance and dimensions. Polymerization follows an expected route where small rod-like aggregates (protofibrils) are initially formed and later extended to form fibrils, worm-like in appearance. Spherical structures of ~25nm diameter are visible at earlier timepoints of the aggregation reaction at high protein concentration, yet are too wide to be direct precursors of the worm-like end-stage products. More likely, they constitute another (or part of another) pathway, available to the protein at higher protein concentrations, and responsible for the loss of linearity in the changes of elongation rate with concentration. Mechanistically, they may still be precursors of worm-like fibrils, though not directly but rather through dissociation and reassociation of aggregating units or reorganization (as discussed in Section 1.2.2).

With the results detailed in this chapter, a deeper understanding of the parameters influencing the aggregation kinetics of AT3[U3/14Q] was achieved, at the same time a strong base was built to compare the determinants of aggregation in the protein's structure, which will be addressed in the next chapter.

Chapter 6

6 – Structural determinants in AT3[U3/14Q] aggregation

6.1 Introduction

As previously mentioned in this thesis, the proteins associated with polyglutaminopathies share no structural similarities other than the expansion of their polyQ tracts (reviewed in Williams and Paulson, 2008). In all cases, this expansion is related to the formation of *in vivo* protein deposits, typically intraneuronal, able to recruit other polyQ-containing proteins (Fujigasaki, 2000; Williams and Paulson, 2008), suggesting a direct involvement of this structural trait in aggregation. Nonetheless, several authors have reported a modulation of polyQ-related aggregation by the flanking regions to these sequences (as mentioned in Section 5.1) and, more recently, by the primary sequence of polyQ tracts themselves (Sen, 2003; Dougan, 2009; Jayaraman, 2009) which, in over a third of the disease-related polyQ-containing proteins, do not comprise solely of glutamine residues (Fig. 6.1). Consistently, when working with highly insoluble and/or aggregation-prone proteins or peptides, it is common practice to produce recombinant forms for expression that contain an added protein domain (or even a whole other protein), that is highly soluble and able to keep the protein of interest soluble. Glutathione-*S*-transferase (GST), Maltose-binding protein (MBP), N-utilization substance (NusA), and many others are examples of peptide sequences recombinantly added to expressed proteins *in vitro* (reviewed in Esposito and Chatterjee, 2006).

Taken together, such strategies and observations indicate the importance of the protein context in determining the solubility and self-association properties of a polypeptide. For ataxin-3, this claim is supported by work with polyQ expanded and non-expanded forms of the protein, suggesting that the initial steps of aggregation for the expanded variant are common to the aggregation of non-expanded variants, clearly de-localizing the initial steps of aggregation from the polyQ tract (Ellisdon, 2006). Data from limited proteolysis and protein sequencing of monomeric ataxin-3 (Sandra Macedo-Ribeiro, University of Porto, personal communication) allowed the

identification of two possible ataxin-3 C-terminal truncations with logical interest for the study of the structural determinants of ataxin-3 aggregation (Fig. 6.2).


```
AR|48-99            - - - - - APPGASLLLLQQQQQQQQQQQQQQQQQQQQQQQQETSPRQQQQQQGEDGSPQAHR - - - - - - - - - - - - - - - - - - 
ATN1|474-512        - - - - - - - - - - - - - PFVSTHHHHHQQQQQQQQQQQQQQQQQHHGNSGPPPP - - - - - - - - - - - - - - - - - - - - - - - - 
ATXN1|187-235       LSQTPGHKAEQQQQQQQQQQQQQHQHQQQQQQQQQQQQQQHLSRAPGLIT - - - - - - - - - - - - - - - - - - - - - 
ATXN2|156-197       - - - - - - - YGPLTMSLKPQQQQQQQQQQQQQQQQQQQQQPPPAAANVRK - - - - - - - - - - - - - - - - - - - - - - - 
ATXN3|282-315       - - - - - - - - - - - - - - - RKRREAYFEKQQQKQQQQQQGDLSGQSSHP - - - - - - - - - - - - - - - - - - - - - - - - - 
ATX7|20-59          - - - - - - - - - PRRAAAAAGGAAAAAARQQQQQQQQQQQPPPPQPQRQQHPPPPPRRTRP - - - - - - - - - - - - - - - 
CACNA1A|2304-2334   - - - - - - - - - - - - - - - - - - IRKAGGSGPPQQQQQQQQQQQQQAVARPGRAAT - - - - - - - - - - - - - - - - - - - - 
HTT|08-48           - - - MKAFESLKSFQQQQQQQQQQQQQQQQQQQQQPPPPPPPPPPPPPQLPQPPPQAQPLLPQPQPPPPPPPPPPGPAVA
TBP|45-105          - - - NTNSLSILEEQQRQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQAVAAAAVQQS - - - - - - - - - - 
```

Fig. 6.1 – Schematic alignment of the polyQ-rich stretch and immediate flanking sequences for disease-related polyQ-containing proteins. For many of these proteins the polyQ region is not comprised solely of glutamine residues, but rather interrupted by other residues such as lysine, histidine and arginine (in the ATXN3, ATXN1 and TBP genes, respectively). On the left are the affected genes for each polyglutaminopathy and the residue number, on the corresponding protein, for the sequences shown. Sequence alignment and colour-coding according to β-sheet propensity was performed with the Jalview tool (Clamp, 2004). Colouring scheme used was part of the Jalview program, the "strand propensity" colour scheme.

The first of the truncation variants identified corresponds to the Josephin domain alone (Fig. 6.2 - bottom), comprising the first 182 residues of ataxin-3. This is not a novel truncation mutant, as several groups have used similar or identical truncations of the full-length protein in their study of ataxin-3 (Masino, 2003 and 2004; Nicastro, 2004, 2005, 2006, 2009 and 2010; Chow, 2004; Gales, 2005). Studying the aggregation details of this variant, and comparing its properties with those of full-length protein (described in Chapter 5), reports on the influence of the Josephin domain in the aggregation of AT3[U3/14Q], proposed in literature to be responsible for early association events (Chow, 2004; Gales, 2005; Masino, 2010; Robertson, 2010).

The second of the truncation variants identified corresponds to the first 262 residues of ataxin-3, comprising the Josephin domain and the first two tandem UIMs (Fig. 6.2 – Middle). Unlike the previous variant, this is a novel truncation mutant, at least in that such a construct had not been studied hitherto, although recently Robertson and co-workers experimented with a similar truncation of ataxin-3 expressed in cells, with results that suggested a direct involvement of this N-terminal portion of ataxin-3 in determining the aggregation and toxic effects of ataxin-3 (Robertson, 2010). Since this variant does not include the polyQ tract or

the third UIM present in the primary structure of AT3[U3/14Q], comparing its aggregation details to those of full-length protein (Chapter 5) provides information about their influence in aggregation events.
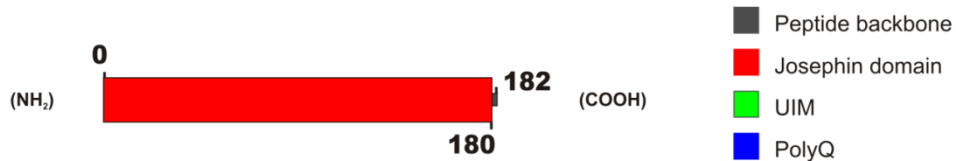

Fig. 6.2 – Schematic representation of the predicted structural details for ataxin-3 (3UIM isoform) and the C-terminal truncation variants designed and utilised in this study. The smallest truncation variant comprises solely the Josephin domain (bottom), while the longest truncation variant, (middle) comprises the Josephin domain and the two tandem UIMs. Prediction of structured regions was performed with the Prosite bioinformatics tool available on the ExPASy server (http://expasy.org/tools/).

While the polyQ tract length may not have a direct influence on the initial steps of ataxin-3 aggregation (Ellisdon, 2006), the third UIM, along with the remainder primary structure of the C-terminus, may modulate the aggregation of ataxin-3 in cells. Harris and co-workers found that the naturally occurring alternative splicing of the last exon, replacing the C-terminal region of 3UIM ataxin-3 with a hydrophobic sequence (forming the 2UIM ataxin-3 isoform described in Kawaguchi, 1997) leads to increased aggregation in cells (Harris, 2010).

Furthermore, comparison of the aggregation properties of the two truncation variants described provides data on the influence of the region between

the polyQ tract and the Josephin domain in the full-length protein (residues 182 to 262). This region contains two tandem UIMs (Donaldson, 2003; Song, 2010) whose presence may, as shown for huntingtin (Miller, 2007), have an influence on protein aggregation. Indeed, co-expression of huntingtin with ataxin-3's UIMs in cells leads to decreased aggregation of the former protein (Miller, 2007).

In this chapter, the comparison of the aggregation properties of the variants introduced above is described, and, thereby, the role of ataxin-3's regions and domains in self-association is determined. The results paint a more detailed picture of the early association events in the pathological and non-pathological states of ataxin-3.

6.2 Protein preparations

6.2.1 Molecular Biology

To prepare truncation variants of ataxin-3, the expression vector described in Section 3.2.1 was used as a starting point. A mutagenic approach was adopted to create the intended sequences, by mutating to a stop codon the nucleotide triplet corresponding to the residue immediately downstream from the C-terminal residue for each intended protein variant (in AT3[U3/14Q]'s sequence, Fig. 6.3).

```
atgtcgtactaccatcaccatcaccatcacctcgaaaacctgtattttcagtctaaagca
 M  S  Y  Y  H  H  H  H  H  L  E  N  L  Y  F  Q  S  K  A
ggctggatggagtccatcttccacgagaaacaagaaggctcactttgtgctcaacattgc
 G  W  M  E  S  I  F  H  E  K  Q  E  G  S  L  C  A  Q  H  C
ctgaataacttattgcaaggagaatattttagccctgtggaattatcctcaattgcacat
 L  N  N  L  L  Q  G  E  Y  F  S  P  V  E  L  S  S  I  A  H
cagctggatgaggaggagagatgagaatggcagaaggaggagttactagtgaagattat
 Q  L  D  E  E  R  M  R  A  E  G  G  V  T  S  E  D  Y
cgcacgttttttacagcagccttctggaaatatggatgacagtggttttttctctattcag
 R  T  F  L  Q  Q  P  S  G  N  M  D  D  S  G  F  F  S  I  Q
gttataagcaatgccttgaaagtttggggtttagaactaatcctgttcaacagtccagag
 V  I  S  N  A  L  K  V  W  G  L  E  L  I  L  F  N  S  P  E
tatcagaggctcaggatcgatcctataaatgaaagatcatttatatgcaattataaggaa
 Y  Q  R  L  R  I  D  P  I  N  E  R  S  F  I  C  N  Y  K  E
cactggtttacagttagaaaaattaggaaaacagtggtttaacttgaatctctcttgacg
 H  W  F  T  V  R  K  L  G  K  Q  W  F  N  L  N  S  L  L  T
ggtccagaattaatatcagatacatatcttgcacttttcttggctcaattacaacaggaa
 G  P  E  L  I  S  D  T  Y  L  A  L  F  L  A  Q  L  Q  Q  E
ggttattctatatttgtcgttaagggtgatctgccagattgcgaagctgaccaactcctg
 G  Y  S  I  F  V  V  K  G  D  L  P  D  C  E  A  D  Q  L  L
cagatgattagggtccaacagatgcatcgaccaaaacttattggagaagaattagcacaa
 Q  M  I  R  V  Q  Q  M  H  R  P  K  L  I  G  E  E  L  A  Q
ctaaaagagcaaagagagtccataaaacagacctggaacgagtgttagaagcaaatgatggc
 L  K  E  Q  R  V  H  K  T  D  L  E  R  V  L  E  A  N  D  G
tcaggaatgttagacgaagatgaggaggatttgcagagggctctggcactaagtcgccaa
 S  G  M  L  D  E  D  E  E  D  L  Q  R  A  L  A  L  S  R  Q
gaaattgacatggaagatgaggaagcagatctccgcagggctattcagctaagtatgcaa
 E  I  D  M  E  D  E  E  A  D  L  R  R  A  I  Q  L  S  M  Q
ggtagttccagaaacatatctcaagatatgacacagacatcaggtacaaatcttacttca
 G  S  S  R  N  I  S  Q  D  M  T  Q  T  S  G  T  N  L  T  S
gaagagcttcggaagagacgagaagcctactttgaaaaacagcagcaaaagcagcaacag
 E  E  L  R  K  R  R  E  A  Y  F  E  K  Q  Q  Q  K  Q  Q  Q
cagcagcagcagcagcagcagggggacctatcaggacagagttcacatccatgtgaaagg
 Q  Q  Q  Q  Q  Q  Q  G  D  L  S  G  Q  S  S  H  P  C  E  R
ccagccaccagttcaggagcacttgggagtgatctaggtgatgctatgagtgaagaagac
 P  A  T  S  S  G  A  L  G  S  D  L  G  D  A  M  S  E  E  D
atgcttcaggcagctgtgaccatgtctttagaaactgtcagaaatgatttgaaaacagaa
 M  L  Q  A  A  V  T  M  S  L  E  T  V  R  N  D  L  K  T  E
ggaaaaaaataa
 G  K  K  *
```

Fig. 6.3 – Nucleotide sequence and the translated amino acid residue sequence for ataxin-3 proteins studied in this thesis. The nucleotide at the C-terminus of each truncation variant is boxed in red or green, for the shortest or the longest truncation variant prepared, respectively. The stop codon for the full-length protein is marked with an asterisk (*). The smallest truncation variant comprises the N-terminal Josephin domain (highlighted in red), while the longest truncation variant contains two ubiquitin interacting motifs (highlighted in green) in addition to the Josephin domain. Of the full-length protein, the polyglutamine tract (highlighted in blue) and a third UIM (highlighted in green), C-terminal to the polyQ, are not included in any of the truncation variants prepared. The N-terminal tag and linker to aid purification (highlighted in yellow) was kept as part of the construct and the engineered recognition site for TEV protease is underlined. The expected cleavage site by this enzyme is labelled with a black arrow. The nucleotide sequence, as obtained and confirmed by sequencing, was translated using the tools available on the Expasy proteomics server (http://expasy.org). The UIM sequences highlighted follow the work by Donaldson and co-workers (Donaldson, 2003).

The nucleotide and translated amino acid sequence used as basis for the mutagenic experiments described is shown on Fig. 6.3, highlighting the C-terminus of each truncation variant with an arrow next to the C-terminal arginine residue (boxed), red for the smallest mutant and green for the longest. Expression plasmids constructed for the truncation variants are listed in Table 2.1 in Section 2.1.1.1. For simplification purposes, the ataxin-3 truncation proteins generated will be referred to as "AT3[Jos]" (Josephin domain of ataxin-3) and "AT3[U2/Jos]" (Josephin domain and the two tandem UIMs) for the smallest and longest truncation variant, respectively (Fig. 6.3).

6.2.2 Expression and Purification

To obtain monomeric ataxin-3 truncation protein for both variants, the protocols for expression and purification of ataxin-3 described in Chapter 3 were followed. No major differences in the behaviour of the truncated proteins were encountered, apart from the expected change in elution volume from size-exclusion chromatography (Fig. 6.4). Interestingly, the final yield of purified protein for both truncated proteins ranged from 10 to 15mg *per* litre of cell culture, higher than the yield of AT3[U3/14Q], usually 5-8mg/l of cell culture (Section 3.2.2).

Analytical size-exclusion chromatography elution profiles and elution volumes of the purified truncated ataxin-3 proteins are shown in Fig. 6.4. Both proteins were purified to high monomeric protein purity (as determined by size-exclusion chromatography and SDS-PAGE) and concentrated to 15-30mg *per* ml, as described in Section 3.2.3. The concentration and storage procedures used, as for AT3[U3/14Q], did not visibly affect the structural integrity of the purified truncated ataxin-3 variants, as judged by size-exclusion chromatography and circular dichroism spectroscopy (Fig. 6.4 and 6.5, respectively).
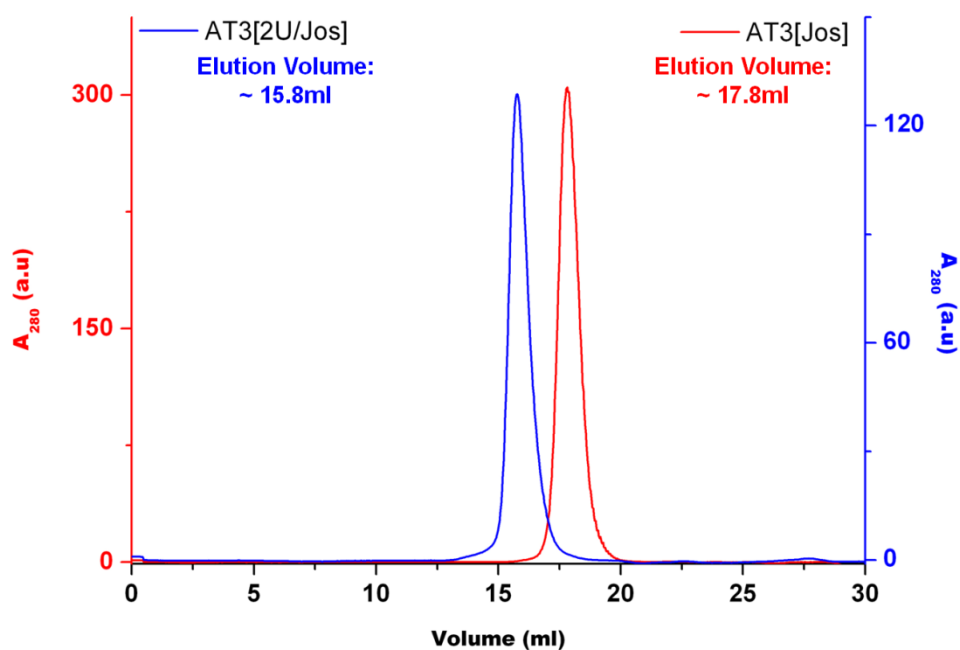
Fig. 6.4 – Typical analytical size-exclusion elution profiles of purified monomeric protein batches of both ataxin-3 truncation mutants studied in this thesis. Absorbance of AT3[U2/Jos] and AT3[Jos], at 280nm, is shown in blue and red, respectively. The elution volume, as judged by the highest intensity point of the monomeric peak, is shown on the top left and right, colour coded as before (blue for AT3[U2/Jos] and red for AT3[Jos]).

6.3 Quality control

Following the same reasoning that led to the experiments described in Section 3.3 for AT3[U3/14Q], the biochemical and biophysical characteristics of the truncated ataxin-3 proteins produced (Section 6.2.2) were studied.

6.3.1 Mass measurements

Similarly to the analysis described in Section 3.3.1, the theoretical mass of truncated ataxin-3 mutants was calculated from the amino acid sequence (Fig. 6.3) with the help of bioinformatics tools available online on the ExPASy server (Swiss Institute of Bioinformatics, available at http://expasy.org). According to this analysis, the expected mass of the variants is 32809.90Da and 23597.66Da for AT3[U2/Jos] and AT3[Jos], respectively. ESI-MS was used to experimentally determine the mass of the truncated ataxin-3 proteins produced, as described in Section 2.3.10.1. Protein sample at 20μM concentration in ammonium acetate buffer at pH 6.6, as determined by the Bradford method (Section 2.3.7.2.1), was sprayed into the mass spectrometer and spectra collected to calculate the mass of each truncated ataxin-3 protein experimentally (as described in Section 2.3.10.1). The result of the calculations indicates a mass of 32848.03Da (+/- 0.27) for AT3[U2/Jos] and 23635.99Da (+/- 1.27) for AT3[Jos], consistent, within error (+/- 0.1%), with the expected mass.

6.3.2 Structural character

The secondary structure of ataxin-3 has been shown to be dominated by the globular-fold Josephin domain (Chow, 2003; Masino, 2004 and Gales, 2005) and therefore is expected to be similar between both the truncated ataxin-3 mutants, prepared as above, and the full-length protein, the latter of which was analyzed in Section 3.3.1. A high α-helical character should thus be apparent for all proteins produced in this thesis. To investigate the persistence of secondary structure in the truncated variants, frozen monomeric protein of both truncated ataxin-3 proteins was thawed and buffer exchanged into phosphate buffer at pH 7.5, and far-UV CD

measurements between 190 and 260nm were taken at 37ºC with 10µM protein concentration, as previously described for AT3[U3/14Q] (Section 3.3.1 and detailed in Section 2.3.10.2).

The results (Fig. 6.5) confirm the expected highly α-helical secondary structure character, with minima at 208nm and 222nm, consistent with previous results with AT3[U3/14Q] (Chapter 3), shown in Fig. 6.5 for comparison. Furthermore, these results suggest that no interaction, or at least not one that significantly affects the protein's secondary structure, exists between the N- and C-terminal portions of ataxin-3. In consonance with correctly folded protein, all truncation variants displayed deubiquitinase activity against Ub-AMC, as shown by Burnett and co-workers for the full-length protein (Burnett, 2003).
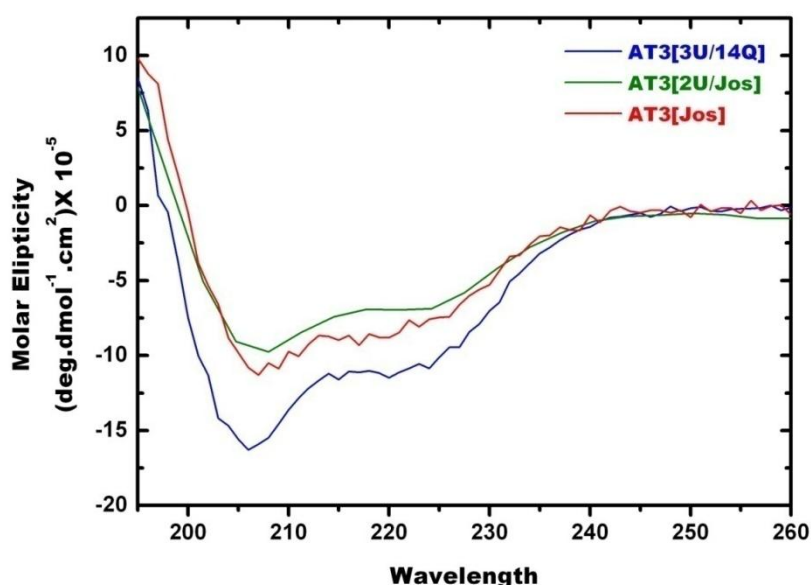

Fig. 6.5 - Far-UV CD spectra of recombinant ataxin-3 proteins prepared in this thesis. Collected spectra of AT3[U3/14Q], AT3[U2/Jos] and AT3[Jos] is shown in blue, green and red traces, respectively. The intensity signal was corrected to the protein concentration used. All proteins display the expected high α-helical character, as seen from the minima at 208 and 222nm.

The results presented in this section show that the monomeric proteins produced display the expected biochemical and biophysical characteristics of ataxin-3, validating their use in the study of the determinants of ataxin-3 aggregation, the goal of this chapter and final objective of this thesis. The comparison of the aggregation characteristics of the three proteins studied in this

thesis, detailed in Chapter 5 and subsequent sections of the present chapter, aimed to provide a better understanding of the molecular events leading to ataxin-3 self-association and, potentially, provide targets for therapeutical strategies designed against ataxin-3-related pathologies.

6.4 Self-association of ataxin-3

With the characterisation of AT3[U3/14Q] aggregation as a basis for comparison (described in Chapter 5), the relative influence of ataxin-3's domains and regions in aggregation was evaluated, by determining the kinetic and thermodynamic details of aggregation of both ataxin-3 truncated variants described above.

6.4.1 Aggregation kinetics

To determine the aggregation kinetics of both truncation variants, frozen monomeric protein sample was thawed and buffer exchanged into TXBa buffer, and an automated aggregation trial, at a series of protein concentrations, initiated as described for AT3[U3/14Q] (detailed in Section 2.3.11.1.3). The resulting assay plate was next sealed and placed inside a BMG Omega microplate fluorimeter, previously equilibrated at 37ºC, and the fluorescence of 20µM ThioT at 480nm monitored over time, reporting the aggregation of the proteins. Motivated by the variability previously seen for AT3[U3/14Q], three replicate experiments, with different purification batches, were conducted as described above and the results are plotted below (Fig. 6.6 for AT3[2U/Jos] and Fig. 6.7 for AT3[Jos]). The intensity measured was normalized for direct comparison across protein concentration, according to Section 2.3.12.2.1.

Both truncation variants display the three stages of aggregation reported previously for the full-length protein, AT3[U3/14Q], in Chapter 3 (Section 3.4.1), suggesting a nucleated assembly process. For simplification purposes they will be referred to as lag phase, elongation phase and stationary phase, in chronological order. Interestingly, the aggregation kinetic traces for AT3[Jos] consistently display a feature at the end of the elongation phase, where the fluorescence signal recorded overshoots the intensity level at which the reaction reaches the stationary phase (Fig. 6.7, A through C). This was not observed for AT3[U2/Jos] (Fig. 6.6, A through C) or AT3[U3/14Q] (Section 5.4.3). The significance of this feature is unknown but its presence in all traces, regardless of concentration and protein batch, suggests that it is a systematic characteristic of the assembly process, and a possible difference in assembly when the Josephin domain of ataxin-3 self-assembles without the C-terminal remainder of the protein.

To simplify the visual comparison of the aggregation kinetics at different initial AT3[Jos] concentrations, as well as between AT3[Jos] and the other two proteins studied in this thesis (AT3[U3/14Q] and AT3[U2/Jos]), the normalization performed on this data was manually adjusted to the highest intensity recorded after this feature on each trace, as opposed to the overall highest intensity (described in Section 2.3.12.2.1) utilized on the data collected with AT3[U2/Jos] (Fig. 6.6) and previously for AT3[U3/14Q] (Section 5.4.3). It is interesting to note, however, that at lower initial concentrations of AT3[Jos], the results resemble those obtained with the other ataxin-3 proteins studied, as seen in Fig. 6.7 – insets.

Fig. 6.6 – Aggregation kinetics of AT3[U2/Jos] in TXBa buffer at 37ºC, monitored by the fluorescence intensity of 20µM ThioT at 480nm. Three separate purification batches, A, B and C, were analyzed (coloured red, green and blue). For each batch of purified protein, eight concentrations were tested, at ten replicates each, in a total of 240 kinetic traces were collected. The results shown are Y-axis normalized.
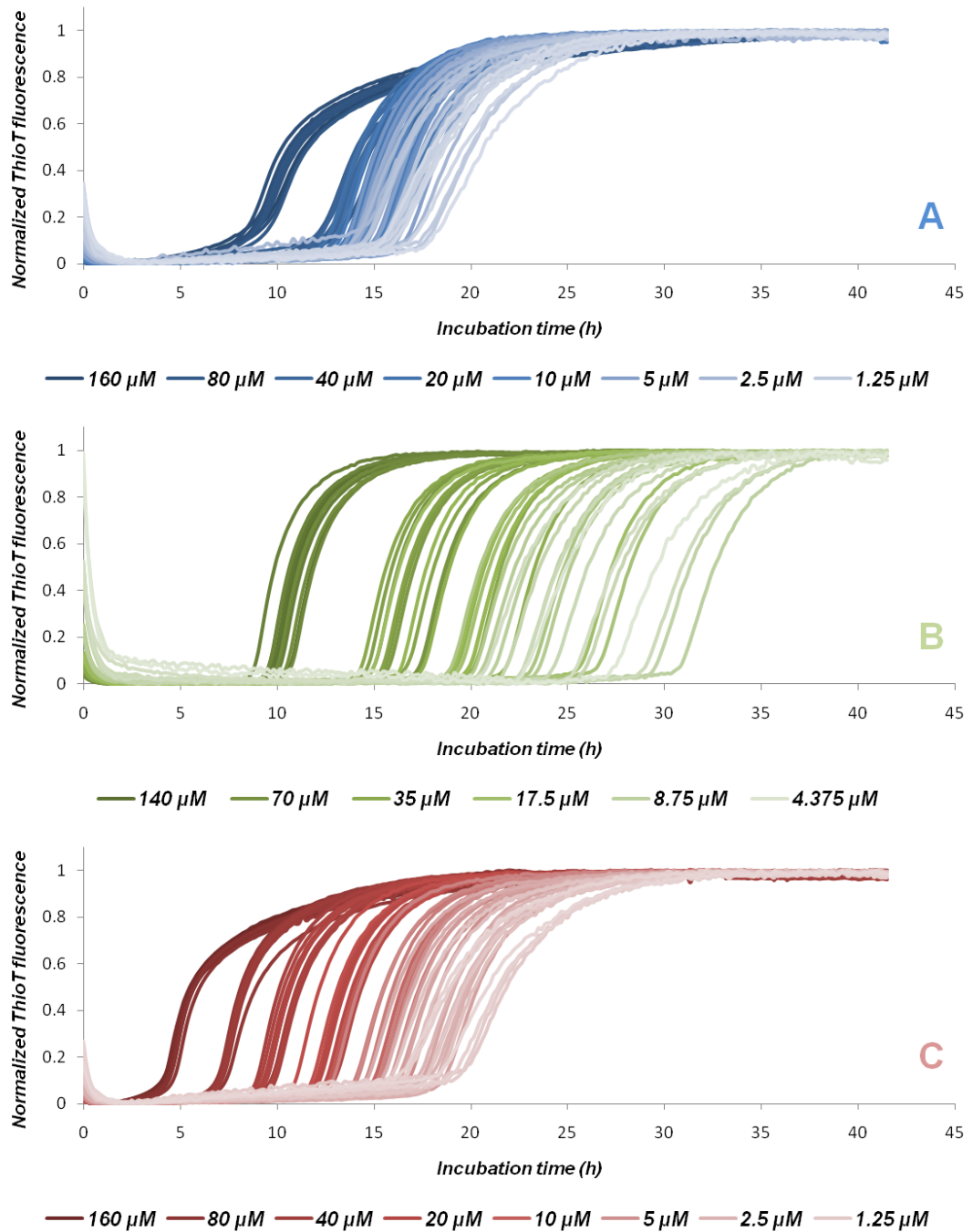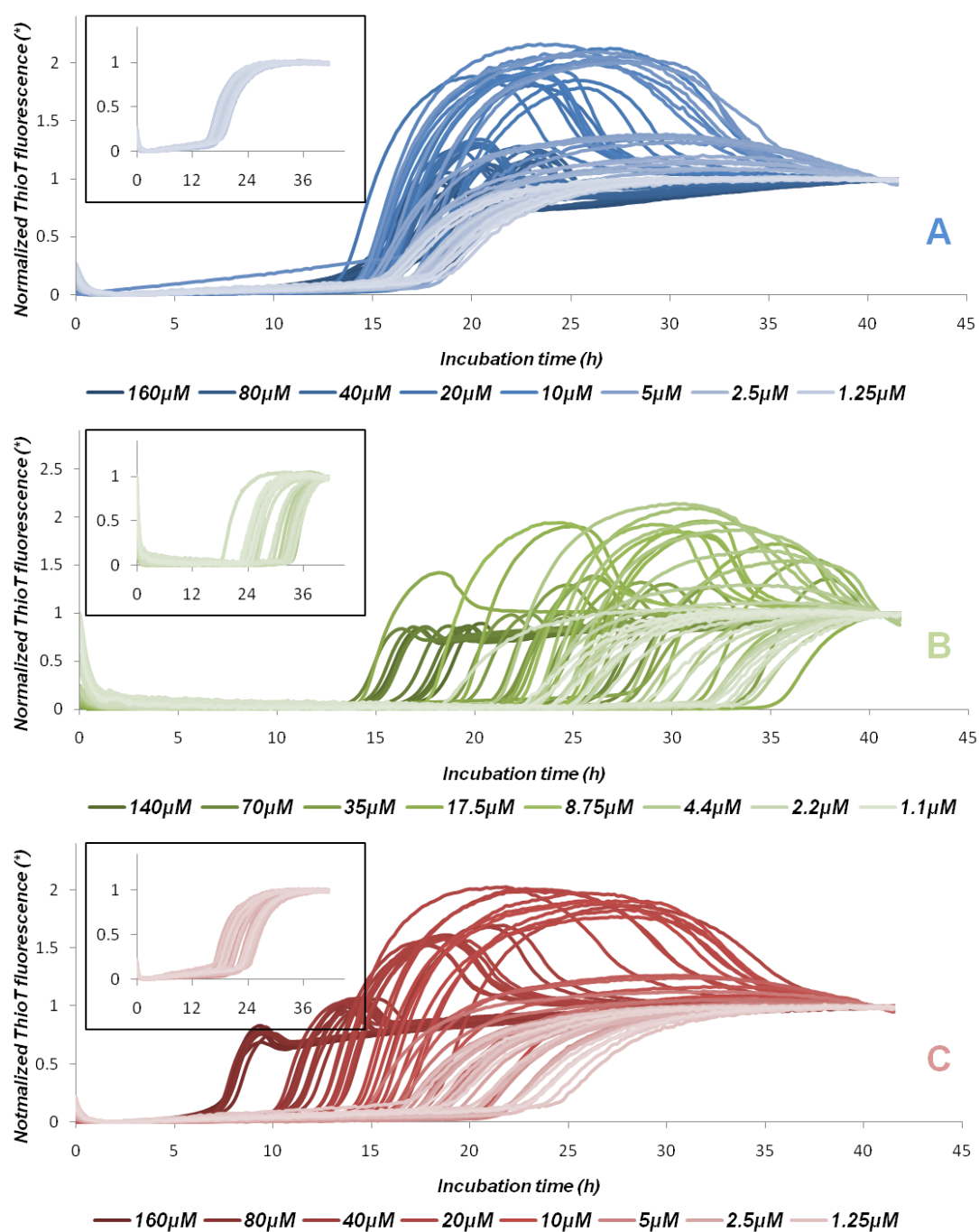
Fig. 6.7 – Aggregation kinetics of AT3[Jos] in TXBa buffer at 37ºC, monitored by the fluorescence intensity of 20µM ThioT at 480nm. Three separate purification batches, A, B and C, were studied. For each, eight concentrations were tested, at ten replicates each, in a total of 240 kinetic traces. A characteristic feature is present in nearly all traces, the only exception being the lowest concentrations tested (insets). Data was Y-axis normalized to the highest intensity value recorded after this feature for simplification purposes.

To further simplify the comparison of results between the proteins studied, values for the length of the lag phase and elongation rate of each reaction were extracted from the ThioT-monitored aggregation kinetics (shown in Fig. 6.6 and 6.7), according to the methodology described by Xue and co-workers (Xue, 2008), as previously performed for the full-length protein (Section 5.4.3 and detailed in Section 2.3.12.3). However, while the length of lag phase visibly titrates inversely with the initial concentration at AT3[2U/Jos] on all tested protein batches (Fig. 6.6), as had previously been observed for AT3[3U/14Q] (Section 5.4.3), this is not clear for AT3[Jos] (Fig. 6.7). Analysis of the extracted t_{lag} parameter highlights this difference, with results for AT3[Jos] showing a more shallow transition with initial protein concentration, when compared to the results for AT3[2U/Jos] (Fig. 6.8 – B and A, respectively), corresponding to a weak, if at all present, dependence of lag phase length on concentration. Furthermore, the extracted t_{lag} parameter shows a more disperse distribution for AT3[Jos] than for AT3[2U/Jos], suggesting a greater influence of stochastic processes in the aggregation kinetics of the former protein.

In terms of elongation rates, the aggregation kinetics of AT3[U2/Jos] show a linear dependence on protein concentration, (Fig. 6.8 – C), consistent with the results obtained previously for AT3[U3/14Q] (Section 5.4.3). On the other hand, the elongation rates across initial protein concentration for AT3[Jos] are, once again, more disperse, suggestive of a different aggregation reaction to that of AT3[U3/14Q] or AT3[U2/Jos], as seen by comparing Fig. 6.8 – C and D, or indeed the results shown in Section 5.4.3 for the full-length protein. AT3[Jos]'s elongation rates increase with protein concentration, but the data cannot be accurately fitted to a linear function, particularly for the concentrations in the midrange of those tested (Fig. 6.8 – D). Note that in linear fitting the data on Fig. 6.8 – C, as previously performed for full-length protein (Section 5.4.3), the data at higher protein concentration was not used due to its greater variability. The trendline shown on the figure represents the dependence of the elongation rates on initial protein concentrations lower than 100µM, though it is clear that at least half of the data at protein concentrations higher than that follows the same trend (Fig. 6.8 – C). Because the data were collected from three separate purification batches, it is

possible that, as raised before in Section 5.4.3 for AT3[U3/14Q], small amounts of oligomeric species are present at the start of the reaction, with an increased effect on elongation rates at higher concentrations, skewing the interpretation of the results.
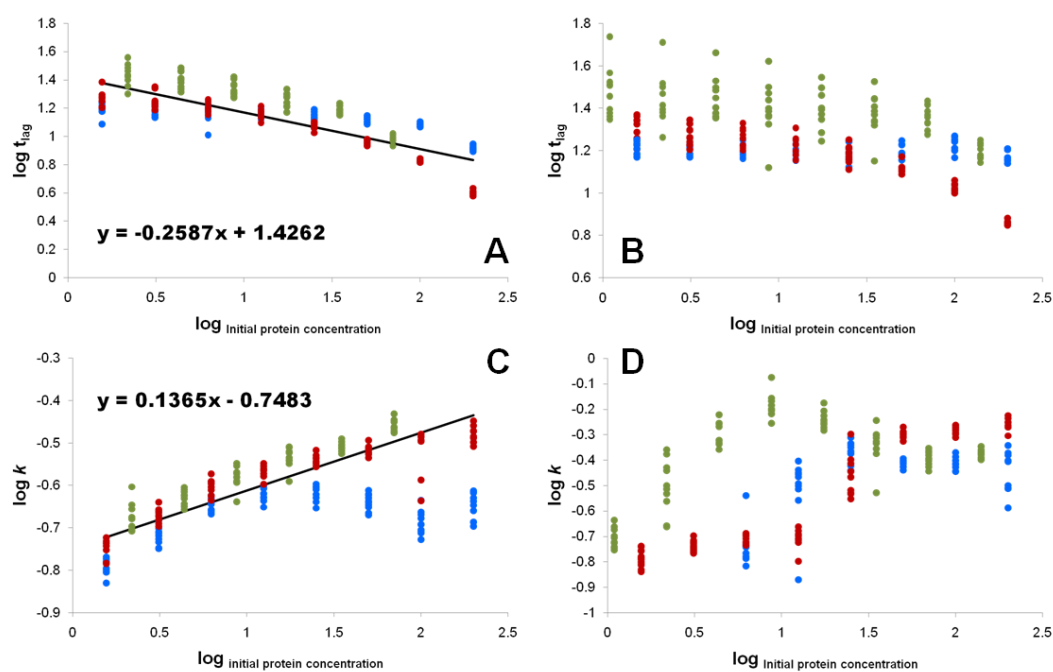

Fig. 6.8 – Comparing aggregation parameters for the assembly reactions of AT3[2U/Jos] and AT3[Jos]. Values for the length of lag phase (t_{lag}) and elongation rate (k) were extracted from the experimental data shown in Fig. 6.6 and Fig. 6.7 according to the methodology described by Xue and co-workers (Xue, 2008). These values were plotted as double logarithm plots for direct comparison, with results for AT3[2U/Jos] shown on the left and those for AT3[Jos] shown on the right. The trend of data for AT3[2U/Jos] was fitted to a linear function, as performed previously for the full-length protein (Section 5.4). The data for AT3[Jos] was too disperse to allow a description of the trend by the same procedure.

The results shown above suggest different aggregation kinetics between the truncated proteins, with AT3[U2/Jos] displaying results consistent with those of full-length protein (Section 5.4), unlike AT3[Jos]. To shed light into these observations the end-products of assembly for both truncation variants were morphologically characterized, and the dependence of a thermodynamical nucleus in the assembly reactions determined, as previously performed for the full-length protein (Chapter 5).

6.4.2 Morphological studies of aggregate species

As performed previously for AT3[U3/14Q] (Section 5.2), the species formed during each truncation variant's aggregation were monitored by negative-stain EM. The aggregation kinetics of replicate samples was monitored by ThioT fluorescence, as above (and detailed in Section 2.3.11.1.4), and sample grids prepared as described in Section 5.2 (detailed in Section 2.3.11.3.2) were imaged, in parallel, using EM. Two initial protein concentrations were studied, 80µM and 20µM, both to ensure visualization of scarcely populated aggregates and an uncluttered view of the sample (respectively).

The results showed that the species formed during the assembly of AT3[U2/Jos] resemble those formed during full-length non-expanded ataxin-3 aggregation, shown here (Chapter 5, Section 5.2) and in literature (Ellisdon, 2006). Initial samples at high protein concentration (Fig. 6.9 – **c**) show small globular aggregates of ~10nm diameter and even some elongated species of 7-10nm width and up to ~50nm length. Over time of incubation under the conditions tested, these species grow in length into worm-like fibrils of 7-10nm width and variable length (Fig. 6.9 – **d** and **e**). Interestingly though, the end-products of aggregation at this protein concentration (Fig. 6.9 – **f**) are heavily clustered, making it technically difficult to characterize them any further, while the results at the lowest concentration of AT3[U2/Jos] studied, despite equally progressing from globular species to elongated worm-like fibrils (Fig. 6.9 – **g** to **j**), do not however cluster as heavily (Fig. 6.9 – **j**). The clustering does not affect the ThioT fluorescence signal, as no significant change is visible upon the appearance of clustered fibrils.
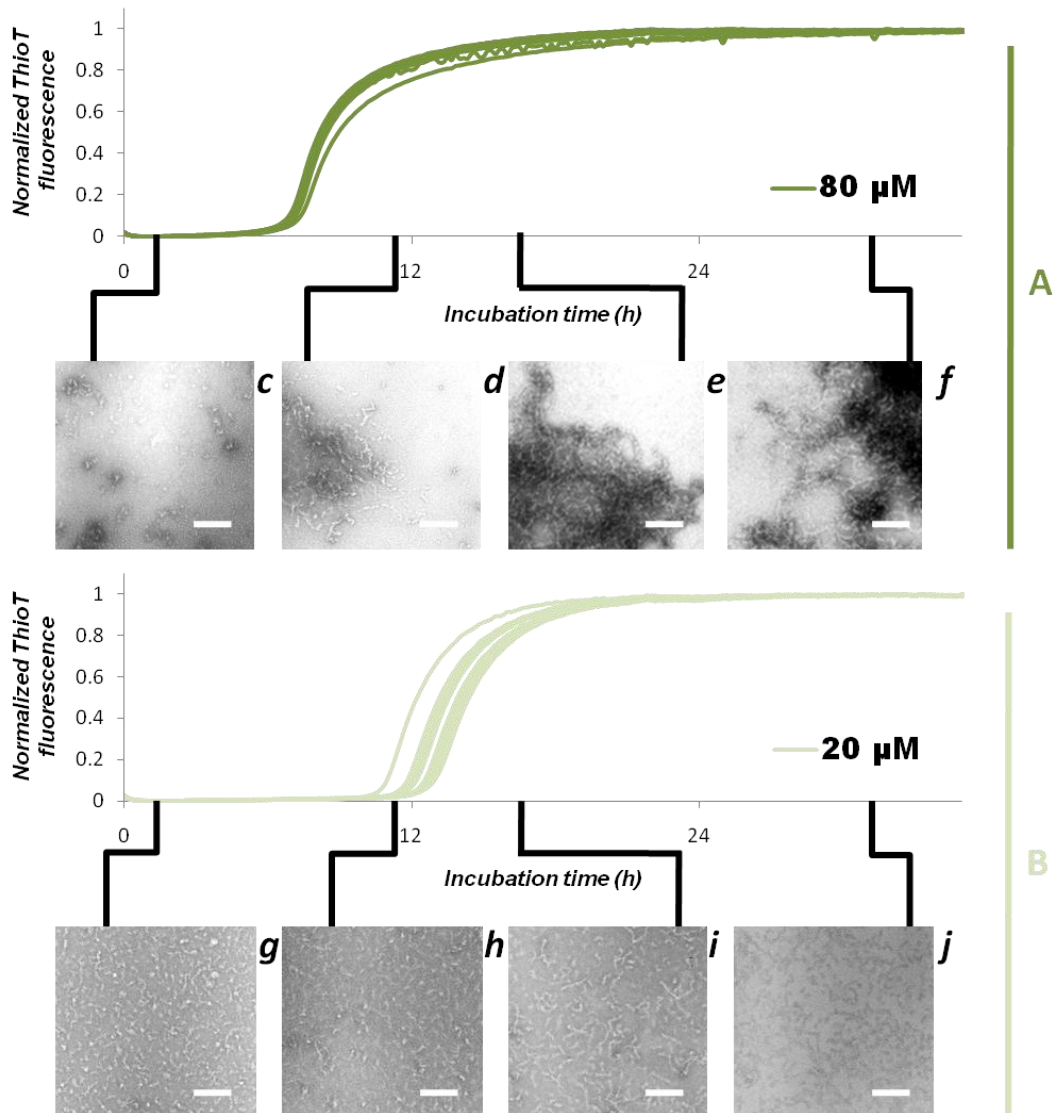
Fig. 6.9 – AT3[U2/Jos] aggregate formation over time. Monomeric sample of AT3[U2/Jos] at two initial concentrations, 80µM (A) and 20µM (B), was quiescently incubated at 37ºC in TXBa buffer, as described in the text. The progress of the polymerization reaction was followed by ThioT fluorescence (as before) and by direct visualization of the structures formed by negative-stain electron microscopy (c to f and g to j for the highest and the lowest concentrations tested, respectively). The bar on the negative-stain electron microscopy images represents 100nm.

Remarkably, and despite a clear three-phase aggregation kinetics reported by ThioT on replicate samples (as above) (Fig. 6.7 – A through C), the aggregation of AT3[Jos] does not result in the formation of visible fibrillar species (Fig. 6.10 – **f** and **j**). At initial timepoints at the highest protein concentration studied, globular species are visible (Fig. 6.10 – **c**), resembling those formed by both AT3[U2/Jos] (Fig. 6.9 – **c** and **g**) and AT3[U3/14Q] (Section 5.2), though no elongated species are

formed over the incubation time monitored (Fig. 6.10 – **c** to **f**). At the lowest concentration tested however, not even these globular species are visible, with all sample grids appearing empty (Fig. 6.10 – **g** to **j**) and indistinguishable from control grids with buffer (not shown).
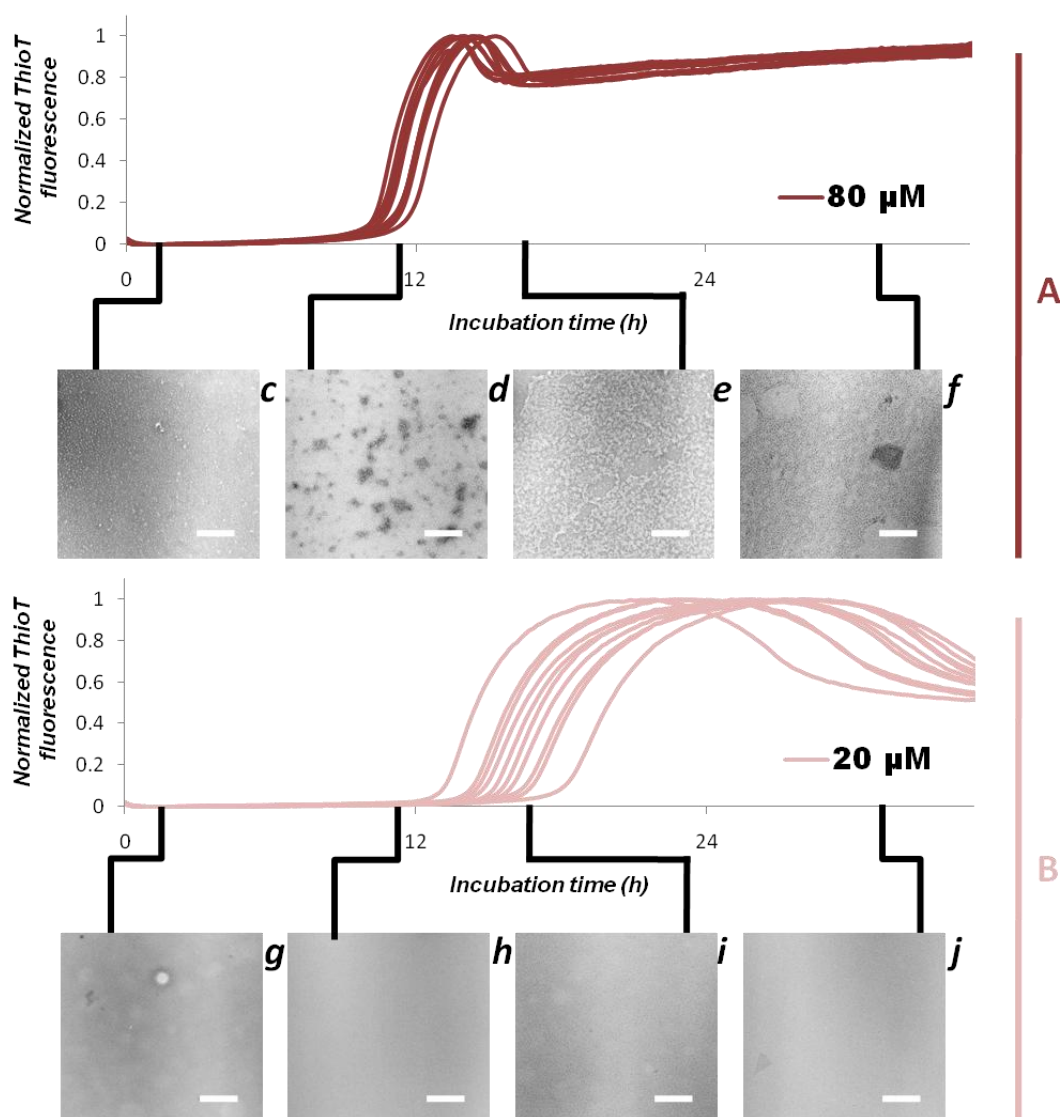
Fig. 6.10 - Aggregate formation over time. Monomeric sample of AT3[Jos] at two initial concentrations, 80μM (A) and 20μM (B), was placed under aggregation conditions, as described in the text. The progress of the polymerization reaction was followed by ThioT fluorescence (as before) and by direct visualization of the structures formed by negative-stain electron microscopy (c to f and g to j for the highest and the lowest concentrations tested, respectively). The bar on the negative-stain electron microscopy images represents 100nm.

These results, together with those shown in the previous section (Section 6.4.1) and compared to the previously obtained results with the full-length protein (Section 5.2), indicate that the aggregation of the AT3[2U/Jos] truncation variant is similar to that of the full-length protein, while the aggregation of AT3[Jos], under identical physiologically-relevant conditions tested, differs from that of both the other two proteins studied. To further distinguish and characterize the aggregation details of these proteins, seeding experiments with end-stage products (which in the case of AT3[Jos] are not visibly fibrillar) were conducted. It is important, at this point, to remember that in nucleated assemblies, seeding with end-products of aggregation abolishes or reduces the lag phase by bypassing the formation of a nucleus, distinguishing the assembly from non-nucleated (downhill) reactions, where a seeding effect is not seen (Oosawa and Asakura, 1975; Ferrone, 1999).

6.4.3 Self-seeded aggregation

To investigate the nature of the protein assembly reactions described above (Sections 6.4.1 and 6.4.2), aggregation end-product sample was prepared from protein samples at 20µM concentration, incubated quiescently at 37ºC (as detailed in Section 2.3.11.1.4). The progress of aggregation was monitored by ThioT fluorescence and, after ~36h incubation, sample was collected and pooled together (as described in Section 2.3.11.1.4). Negative-stain EM images were then taken of the products of aggregation, prior to their use in seeding experiments (Fig. 6.11 and Fig. 6.12).
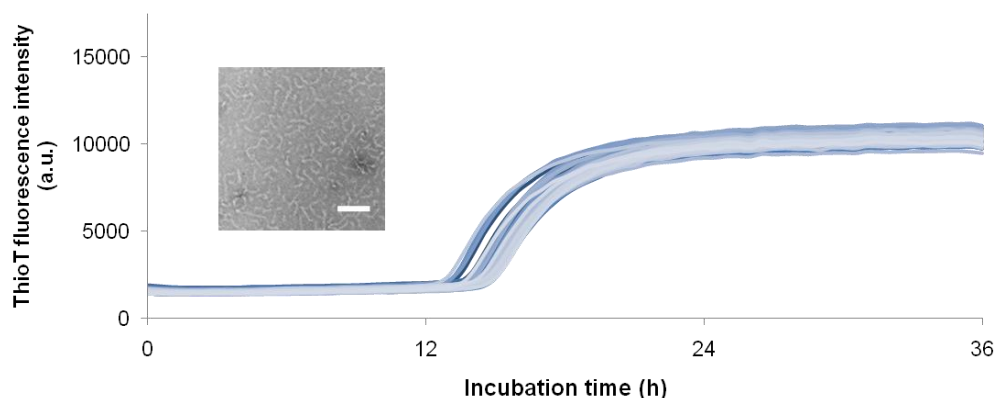
Fig. 6.11 - Production of truncated ataxin-3 proteins' seeds. 20µM monomeric AT3[2U/Jos] in TXBa buffer was incubated at 37ºC and aggregation progress was monitored by ThioT fluorescence. Sample after 36h incubation was collected and visualised by negative-stain electron microscopy (inset). The bar on the microscopy image represents 100nm.
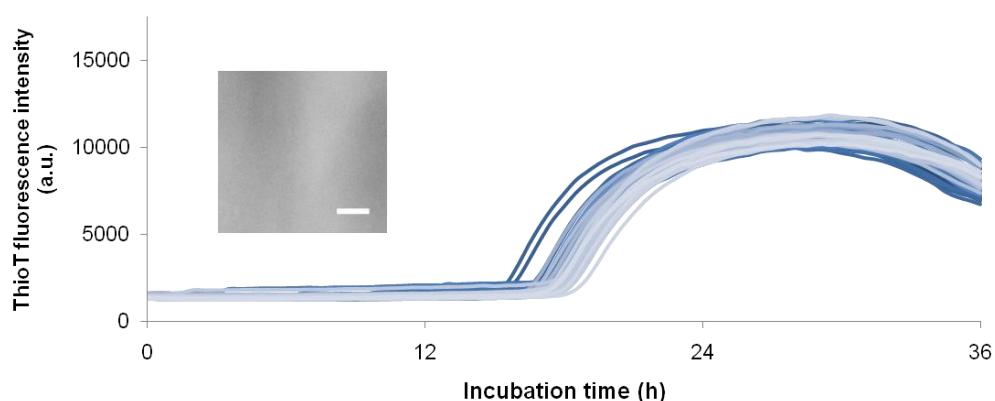
Fig. 6.12 - Production of truncated ataxin-3 proteins' seeds. 20µM monomeric AT3[Jos] in TXBa buffer was incubated at 37ºC and aggregation progress was monitored by ThioT fluorescence. Sample after 36h incubation was collected and visualised by negative-stain electron microscopy (inset). The bar on the microscopy image represents 100nm.

The ThioT monitored aggregation kinetics and the EM images collected for the seed samples of each truncation variant are consistent with previous observations (compare Fig. 6.11 and 6.12 to Fig. 6.9 and Fig.6.10, respectively). As such, these seed samples were used in seeding experiments, as previously performed with full-length protein (Section 5.3 and detailed in Section 2.3.11.2), added at different volumetric percentages to monomeric sample in the same buffer and of the same protein concentration. For practicality, the term "seeds" will be

used henceforth as a synonym of "end-products of aggregation", when used in seeding experiments.

For AT3[U2/Jos], the results of these experiments demonstrate a nucleated assembly process, much like previously shown for AT3[U3/14Q] (Section 5.3), where the addition of pre-formed aggregates of the protein reduced the lag phase length and affected the initial rates of aggregation (Fig. 6.13). This reduction of the lag phase titrates well with the percentages of end-stage sample added, as expected for a nucleated growth. However, as previously noted for the full-length protein (Section 5.3), even with 30% (v/v) seeds added the lag phase is not fully ablated.

Fig. 6.13 – Self-seeding AT3[2U/Jos] aggregation. Reaction mixtures of 20µM monomeric AT3[2U/Jos] were prepared in TXBa buffer and supplemented with varying volumetric percentages of AT3[2U/Jos] seed sample, formed under identical conditions, keeping the total mass of protein per reaction mixture identical. Quiescent aggregation at 37ºC was monitored by ThioT fluorescence, as before. Unseeded reaction mixtures were also prepared as control (shown in red). Aggregates from seeded reactions with 30% (v/v) end-products of aggregation were visualised by negative-stain electron microscopy – inset. The bar on the microscopy image represents 100nm.

For AT3[Jos], however, irrespective of the volume percent of end-stage sample added, only minimal differences in the lag phase length result (Fig. 6.14). Overall, the results with this protein are more consistent with a non-nucleated assembly process, adding to the accumulated evidence for a different assembly between AT3[Jos] and AT3[U2/Jos] or AT3[U3/14Q] (Fig. 6.13 and 6.14, compared

to Fig. 5.3). Accordingly, the products of self-seeding reactions of monomeric AT3[U3/14Q] and AT3[U2/Jos] are indistinguishable from those formed from unseeded reactions, as expected for a nucleation-driven assembly. The end-products of AT3[Jos] aggregation are once again not visible by negative-stain electron microscopy (Fig. 6.10 and Fig. 6.14), suggesting no change in the assembly with or without seed sample present at the start of the reaction.
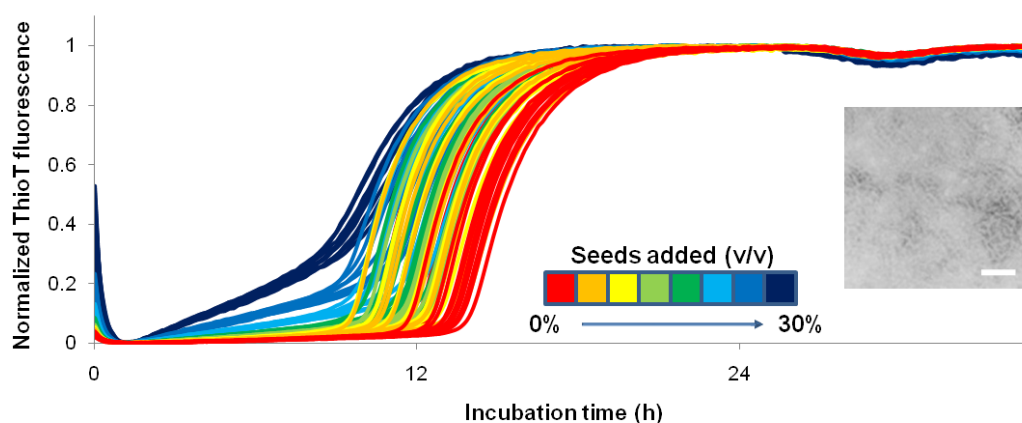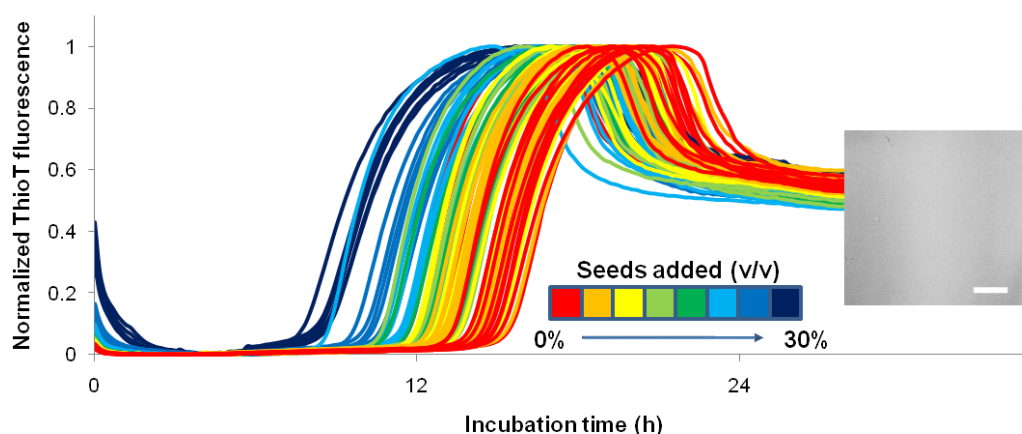

Fig. 6.14 – Self-seeding AT3[Jos] aggregation. Reaction mixtures of 20µM monomeric AT3[Jos] were prepared in TXBa buffer and supplemented with varying volumetric percentages of AT3[Jos] seed sample, formed under identical conditions, keeping the total mass of protein per reaction mixture identical. Quiescent aggregation at 37ºC was monitored by ThioT fluorescence, as before. Unseeded reaction mixtures were also prepared as control (shown in red). Aggregates from seeded reactions with 30% (v/v) end-products of aggregation were visualised by negative-stain electron microscopy – inset. The bar on the microscopy image represents 100nm.

Further characterisation of nucleated reactions includes mathematical estimations of thermodynamical nucleus size, as performed for AT3[U3/14Q] in Section 5.5. These calculations do not apply to non-nucleated processes and as such they were not performed for AT3[Jos].

6.4.4 Estimating the thermodynamical nucleus size

From the aggregation kinetics data collected at varying initial protein concentration and according to both Ferrone's and Oosawa's analysis, the size of the thermodynamical nucleus was estimated as performed previously for AT3[U3/14Q] (Section 5.5 and detailed in Section 2.3.12.4). Details of the aggregation kinetics such as the lag phase length, midpoint of the reaction and elongation rates were extracted from the experimental data shown in Section 6.4.1, using the methodology developed by Xue and co-workers (Xue, 2008) (detailed in Section 2.3.12.3).

6.4.4.1 the Oosawa method

The first of the two methods used to estimate the thermodynamical nucleus size presented here, and the simplest to calculate from experimental data, was previously discussed in Section 5.5.1 (detailed in Section 2.3.12.4.1). Accordingly, the slope of a log-log plot relating the midpoint of the reaction (t_{50}) and the initial protein concentration was constructed (Fig. 6.15), and the slope of a linear fit to the data used to estimate the thermodynamical nucleus size (as described in Section 2.3.12.4.1).

$$y = -0.2049x + 1.4433$$
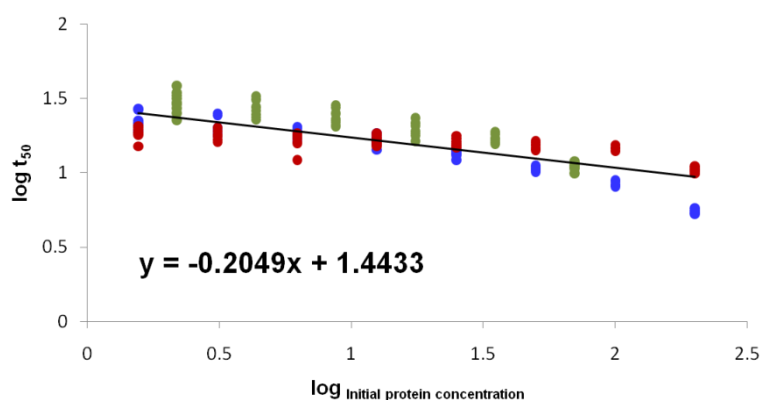
Fig. 6.15 – Thermodynamical nucleus size estimation for AT3[U2/Jos] aggregation, according to Oosawa's method (Ooosawa, 1975). The midpoint of the assembly reaction, t_{50}, extracted from the experimental data using the strategy described by Xue and co-workers (Xue, 2008), is plotted against the initial protein concentration in a log-log plot. A linear fit of the data is shown graphically as a black line, with the corresponding equation shown at the bottom left, also in black.

Solving the equation of the method, the nucleus is estimated to be monomeric, in agreement with the results shown for AT3[U3/14Q] (Section 5.5.1), suggesting a similar assembly process for both proteins.

6.4.4.2 the Ferrone method

Estimation of the size of the thermodynamical nucleus by Ferrone's method requires further mathematical processing of the experimental data, as described in Section 5.5.2 (and detailed in Section 2.3.12.4.2). A log-log plot relating the initial polymerization rate and the initial protein concentration is shown in Fig. 6.16 and the slope of the linear fit shown (Fig. 6.16 – equation) used to calculate the thermodynamical nucleus size for the assembly of AT3[U2/Jos]. The results suggest a monomeric nucleus, consistent with the previous analysis (Section 6.4.4.1) and the estimation of the thermodynamical nucleus size for the full-length protein, AT3[U3/14Q] (Section 5.5.2).

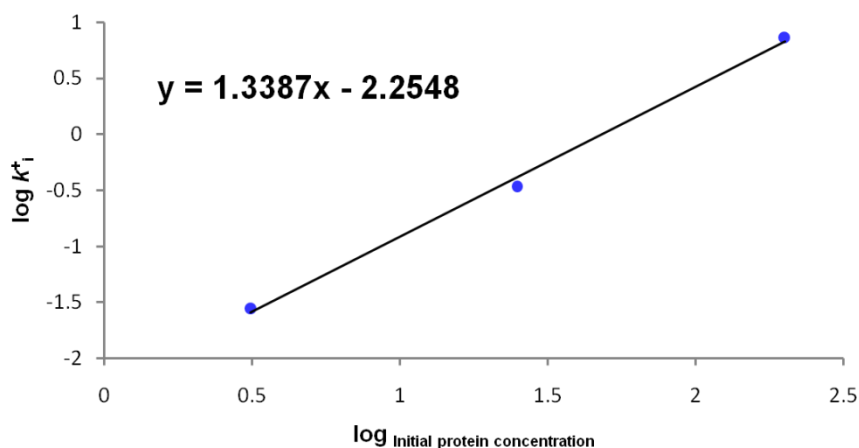
$$y = 1.3387x - 2.2548$$

Fig. 6.16 – Thermodynamical nucleus size estimation for AT3[U2/Jos] aggregation, according to Ferrone's method (Ferrone, 1999). The initial rate of elongation, k^+_i, was plotted against the initial protein concentration in a log-log plot, and a linear fit is shown graphically as a black line, with the corresponding equation shown at the top left, also in black.

All the results presented thus far in this chapter suggest that AT3[U3/14Q] and AT3[U2/Jos] share a common assembly mechanism, that is different from the one through which AT3[Jos] self-associates. However, given the multitude of assembly pathways available at any given condition, it is to be expected that some characteristics would be shared across pathways (see Section 1.2.2). Further evidence in support or disclaim of a shared nucleated assembly pathway can be obtained by cross-seeding reactions, i.e., by seeding the reaction of one protein with the end-products of another's aggregation. Only when the assembly pathways are similar can these seeds be elongated.

6.5 Cross-seeding experiments

Seeds of all three proteins were used to determine their self-seeding properties in previous sections of this thesis (Sections 5.3 and 6.4.3). Here, replicate seed samples were used to assess the cross-over of aggregation-competent species between the ataxin-3 proteins produced. All samples of seeds were confirmed, by negative-stain electron microscopy, to be unaltered over the time between the experiments. Each experiment shown below includes unseeded reactions to act as an internal control of seeding effects. In addition, the self-seeding results for each protein (Section 5.3 and 6.4.3) are shown again, for comparison purposes.

6.5.1 Cross-seeding of monomeric AT3[U3/14Q]

Monomeric AT3[U3/14Q] was prepared at 20µM concentration in TXBa buffer, as previously described in Chapter 5 (and detailed in Section 2.3.11.2). Volumetric percentages of seed sample from each of the other two ataxin-3 proteins studied, formed under identical conditions, were adjusted by dilution with AT3[U3/14Q] monomeric sample and aggregation monitored by the fluorescence

intensity of 20μM Thioflavin T (ThioT) at 480nm, as previously described for self-seeding experiments (Section 5.3 and detailed in Section 2.3.11.2). From the seeding experiments shown below (Fig. 6.17), it is clear that AT3[U3/14Q] can be seeded with products of its own quiescent aggregation (as previously shown) or those of AT3[U2/Jos] aggregation (Fig. 6.17 – A and B, respectively). Lag phases are significantly, and similarly, shortened in response to the presence of each of these proteins' seeds. The results suggest that the full-length protein and the longest truncation variant (AT3[U3/14Q] and AT3[2U/Jos], respectively) share similar or identical assembly pathways. On the contrary, AT3[Jos] does not originate end-products that can seed the assembly of monomeric AT3[U3/14Q] (Fig. 6.17 – C). This result is not entirely surprising, since the results in the previous section (Section 6.4.3) showed that the aggregation of AT3[Jos] does not follow a nucleated assembly.

As with the self-seeding results, the effect of cross-seeding monomeric AT3[3U/14Q] with pre-formed aggregates of AT3[2U/Jos] titrates with the amount of seeds added, with even the highest amount added at the start of the reaction (30% (v/v)) not resulting in a complete ablation of the lag phase. Consistently, both unseeded, self-seeded and cross-seeded reactions of monomeric AT3[3U/14Q] result in worm-like fibrillar aggregates, indistinguishable from each other (compare EM insets on Fig. 5.2 and Fig.6.17). Since seeding occurs *via* the elongation of pre-formed aggregates, it is not surprising that the end-stage products of seeded reactions resemble each other. In the case of cross-seeding with AT3[Jos], the monomeric protein present is likely responsible for determining the assembly reaction, leading to similar end-products of aggregation to those formed in unseeded reactions. Had seeds from AT3[Jos] been able to be extended by monomeric AT3[3U/14Q], the details of aggregation would follow those previously shown for AT3[Jos] unseeded and self-seeded reactions, with fibrillar aggregates not visibly formed.

Fig. 6.17 – Quiescent assembly kinetics of 20μM monomeric AT3[U3/14Q] in the absence or presence of varying amount of seeds of ataxin-3 proteins at the start of the reaction. A) seeded with varying amounts of AT3[U3/14Q] seeds (reproduced from Fig. 5.3). B) seeded with varying amounts of AT3[U2/Jos] seeds. C) seeded with varying amounts of AT3[Jos] seeds. The progress of aggregation in TXBa buffer at 37ºC was monitored by the fluorescence intensity of 20μM ThioT at 480nm. Aggregation products collected after ~41h incubation were visualized by negative-stain EM and are shown as insets. Bar on the EM images represents 100nm.

6.5.2 Cross-seeding of monomeric AT3[U2/Jos]

Prepared as mentioned above (detailed in Section 2.3.11.2), monomeric AT3[U2/Jos] at 20µM concentration in TXBa buffer was incubated in order to allow aggregation to take place. In parallel, samples were incubated with varying volumetric percentages of seed samples from AT3[U3/14Q], AT3[2U/Jos] and AT3[Jos].

Not surprisingly, since the previous section (Section 6.5.1) indicated that AT3[U3/14Q] and AT3[U2/Jos] shared a common assembly pathway, the results obtained from the cross-seeding of AT3[U2/Jos] monomer are identical to those previously shown for the full-length protein. End-stage products of seeded and unseeded reactions with AT3[U2/Jos] are fibrillar and indistinguishable from each other (compare EM insets on Fig. 6.11 and 6.18) and from those formed by monomeric AT3[U3/14Q] (compare Fig. 6.17 and 6.18). The results strengthen the conclusions taken from the previous section: AT3[U3/14Q] and AT3[U2/Jos] share a nucleated assembly reaction with each other, that is not shared with AT3[Jos].
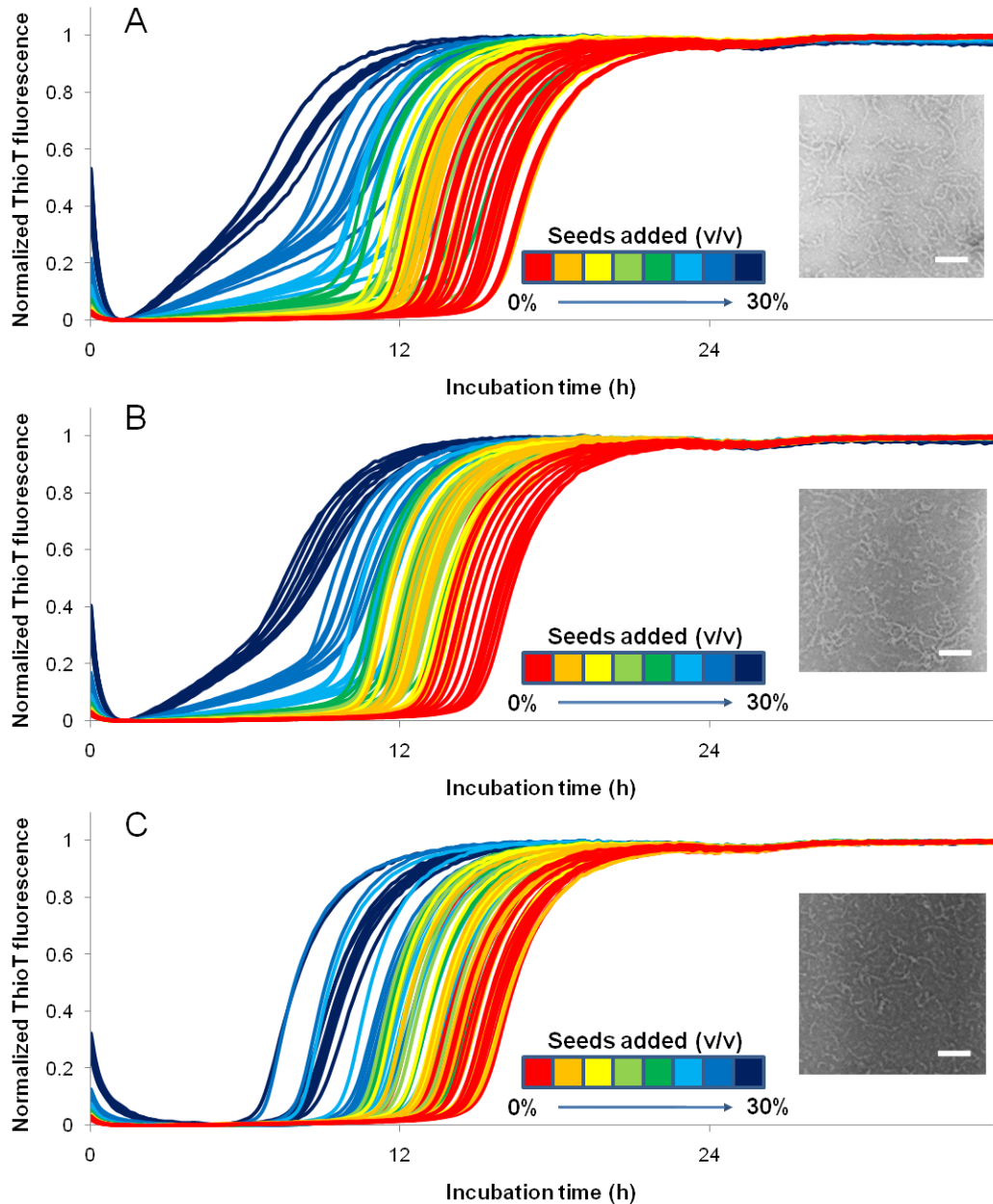
Fig. 6.18 – Quiescent assembly kinetics of 20µM monomeric AT3[2U/Jos] in the absence or presence of varying amount of seeds of ataxin-3 proteins at the start of the reaction. A) seeded with varying amounts of AT3[U3/14Q] seeds B) seeded with varying amounts of AT3[U2/Jos] seeds (reproduced from Fig. 6.13). C) seeded with varying amounts of AT3[Jos] seeds. The progress of aggregation in TXBa buffer at 37ºC was monitored by the fluorescence intensity of 20µM ThioT at 480nm. Aggregation products collected after ~41h incubation were visualized by negative-stain EM and are shown as insets. Bar on the EM images represents 100nm.

6.5.3 Cross-seeding of monomeric AT3[Jos]

As mentioned above (Section 6.4.3), AT3[Jos] displays characteristics that point to a non-nucleated process of assembly. Whether this construct is able to elongate seeds of AT3[2U/Jos] and AT3[U3/14Q] remained an open question. Accordingly, if the Josephin domain (the only portion of the protein comprised in AT3[Jos]) is directly involved in the fibril core of ataxin-3 end-products of aggregation, this variant should have the inherent capacity to assemble into fibrils if provided with seeds. To test this possibility, cross-seeding experiments of monomeric AT3[Jos] with end-stage products of AT3[U3/14Q] and AT3[U2/Jos] were conducted (Fig. 6.18), and compared with the previous self-seeding results for the protein (Fig. 6.19 – C, reproduced from Fig. 6.14).

The results obtained are quite puzzling, with a clear reduction of lag phase length in monomeric AT3[Jos] reactions seeded with pre-formed aggregates of both AT3[U3/14Q] and AT3[2U/Jos] (Fig. 6.19 – A and B, respectively), contrasting with the self-seeding results (Fig. 6.19 – C). Interestingly though, despite a visible seeding effect, the final products of aggregation from monomeric AT3[Jos] seeded reactions, much like those formed from unseeded reactions, are not visible by negative-stain electron microscopy, indicating that the worm-like fibrils provided at the start of the reaction as seeds are not being extended (Fig. 6.18 – A and B, insets). A possible explanation for these results is a surface-catalysed nucleation mechanism, similar to that described for sickle cell haemoglobin (Ferrone, 2002) or the Islet amyloid peptide (Ruschak and Miranker, 2007), where the seeds provide a scaffold to the polymer species formed, as opposed to an elongation interface.
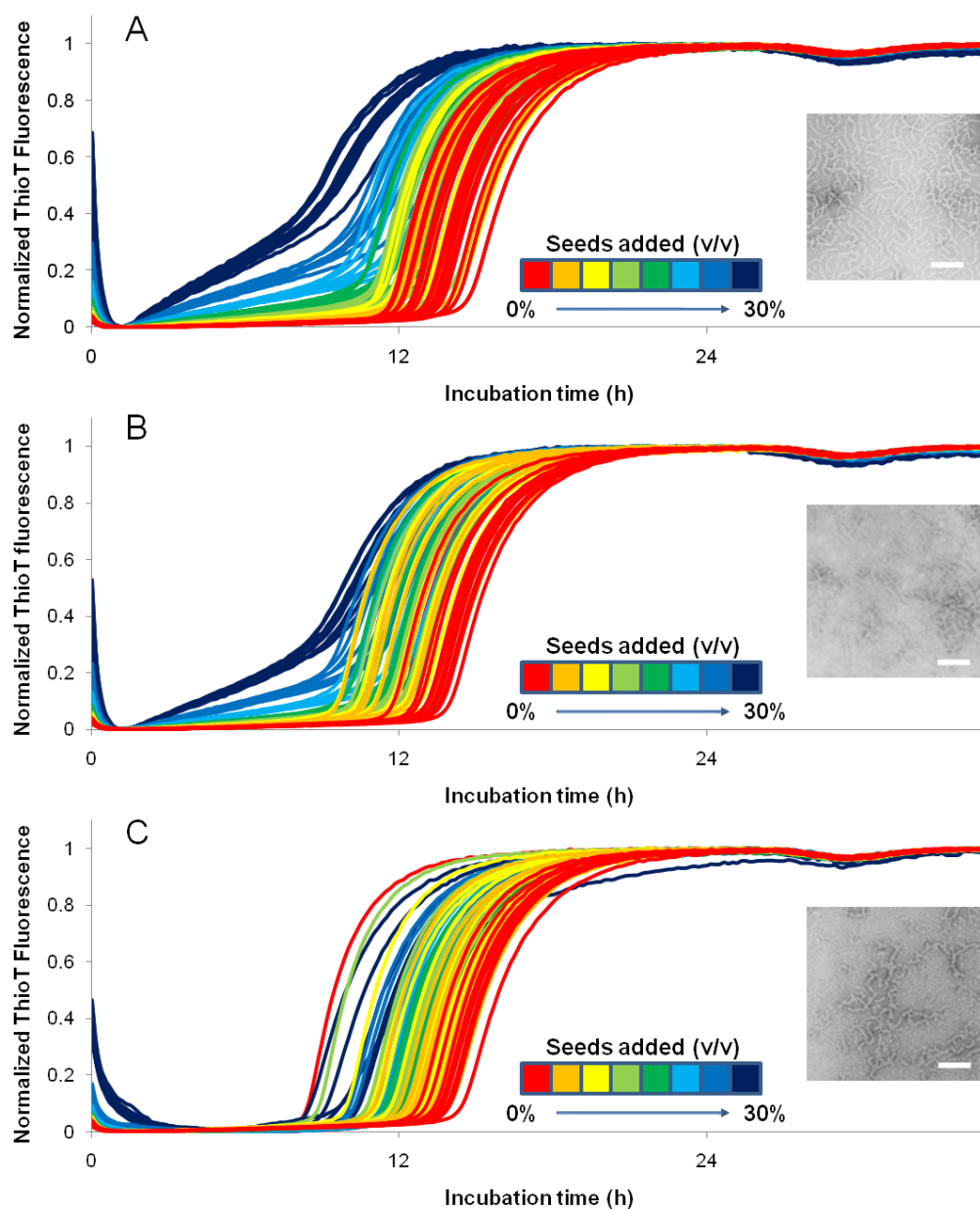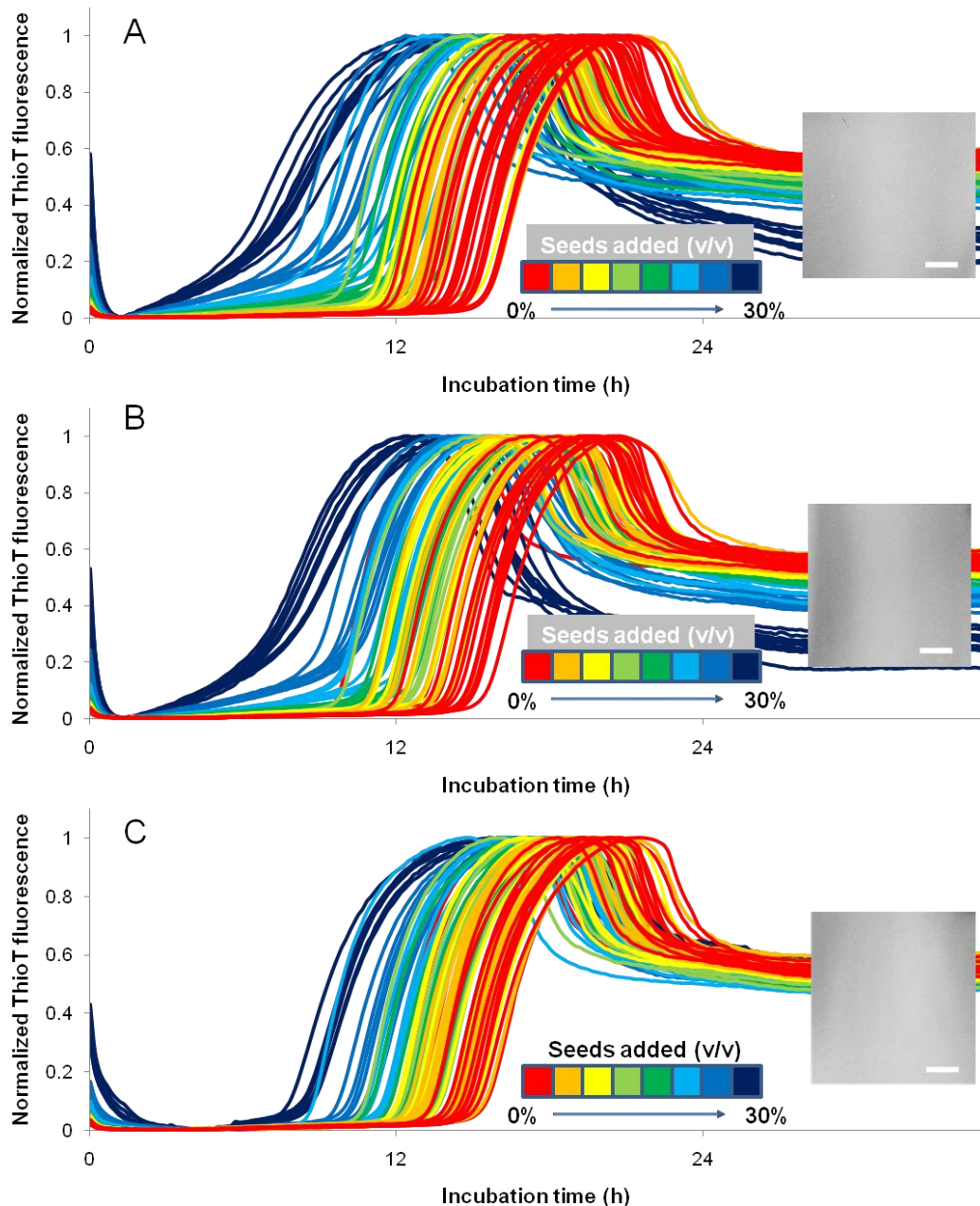
Fig. 6.18 – Quiescent assembly kinetics of 20µM monomeric AT3[Jos] in the absence or presence of varying amount of seeds of ataxin-3 proteins at the start of the reaction. A) seeded with varying amounts of AT3[U3/14Q] seeds. B) seeded with varying amounts of AT3[U2/Jos] seeds. C) seeded with varying amounts of AT3[Jos] seeds (reproduced from Fig. 6.14). The progress of aggregation in TXBa buffer at 37ºC was monitored by the fluorescence intensity of 20µM ThioT at 480nm. Aggregation products collected after ~41h incubation were visualized by negative-stain EM and are shown as insets. Bar on the EM images represents 100nm.

6.6 Conclusion

The results presented in this chapter demonstrate the relative importance of ataxin-3's domains in the protein's self-assembly mechanism. The truncation variants of ataxin-3, prepared as described in this chapter, do not show significant conformational or functional differences from full-length protein in their monomeric forms, thus validating the comparison of aggregation details between them. In addition, this similarity between the monomeric proteins' characteristics suggests a lack of significant interaction between the N- and C-terminal moieties of the full-length protein, as previously suggested in literature (Bevivino, 2001; Masino, 2003).

The full-length protein, AT3[U3/14Q], was shown before (Chapter 5) to aggregate into worm-like fibrillar material *via* lag-dependent kinetics, in a nucleated assembly process with a thermodynamical nucleus estimated to be a monomer. In contrast, AT3[Jos], a truncation variant comprising only the Josephin domain of ataxin-3, was shown here to assemble *via* a non-nucleated mechanism, despite retaining apparent lag-dependent kinetics. While a lag-dependence can be explained by the existence of secondary processes (reviewed in Wang 2010) to the main assembly reaction (such as multiple competing pathways or high off-rates of the initial species formed), the absence of a self-seeding effect from the pre-formed aggregates added at the start of the reaction indicates the absence of a thermodynamical bottleneck, a nucleus. In addition, the end-stage products of AT3[Jos] assembly are not visibly fibrillar. Together, the results suggest that AT3[3U/14Q] and AT3[Jos] do not assemble through similar pathways. Accordingly, cross-seeding results show that AT3[Jos] seeds do not change the aggregation rates or structures formed from monomeric AT3[3U/14Q] aggregation, and thus do not bypass the nucleation phase of the assembly. However, cross-seeding experiments with monomeric AT3[Jos] seeded with AT3[3U/14Q] end-products showed an increase in the initial aggregation rates with increasing seeds added. Interestingly, the end- products of these reactions were not visibly fibrillar, indicating that the

AT3[3U/14Q] seeds provided are not being extended, but rather accelerating the aggregation in some other way. One possible explanation is a surface-catalysed mechanism (Ferrone, 2002; Ruschak and Miranker, 2007), where the full surface of the fibrillar material provided as seeds works as a scaffold from which aggregates grow.

By contrast, the longest truncation variant of ataxin-3 prepared, AT3[U2/Jos], comprising the Josephin domain and the tandem UIMs, self-assembles *via* a nucleated pathway that is shared with the assembly of full-length protein under physiologically-relevant conditions, as shown by self- and cross-seeding experiments. The ThioT-monitored unseeded aggregation kinetics of AT3[2U/Jos] and AT3[3U/14Q] show similar lag-dependent reactions that form worm-like fibrilar products of similar dimensions. Adding to the insight provided by Ellisdon and co-workers (Ellisdon, 2006), who showed that the early steps in ataxin-3 aggregation are not dependent on the length of the polyQ region, the similarity of aggregation parameters for full-length ataxin-3 and the AT3[2U/Jos] truncation variant, strongly suggests that neither the polyQ region nor the remainder C-terminal portion of ataxin-3 (which includes the third UIM feature) are required for the early steps of ataxin-3 fibril formation. Instead, cross-seeding results with monomeric AT3[3U/14Q] seeded with AT3[2U/Jos] end-stage products show that the structures formed by the latter protein are on-pathway to fibril formation by the former. The inverse experiment, where monomeric AT3[2U/Jos] was seeded with AT3[3U/14Q] end-products, revealed similar results, suggesting a shared aggregation pathway by the two proteins. Taken together, these results lead to the conclusion that the stretch of residues containing the polyQ tract and the third UIM of the full-length protein, along with the remainder of the C-terminal stretch, is not directly responsible for determining the assembly into fibrillar species. Work by Harris and co-workers (Harris, 2010) showed a differential aggregation between the 2UIM isoform and the 3UIM isoform of ataxin-3 which, according to the results presented in this thesis, is not due to the characteristics of the 3UIM C-terminal moiety, but rather the alternatively spliced C-terminal sequence that characterizes the 2UIM isoform. Instead, it is the ~80 residues stretch, C-terminal from the Josephin

domain, that allows the Josephin domain to aggregate on the same pathway as the full-length protein, leading to the formation of worm-like fibrils.

Chapter 7

7 – Conclusions and Outlook

The predicted main structural features of ataxin-3 include the folded N-terminal Josephin domain, followed by a relatively unstructured C-terminal tail (Masino, 2003). In the latter, and for the splice isoform of ataxin-3 studied in this thesis, three UIMs are found, along with a polyQ tract that becomes expanded to ≥~50 glutamine residues in pathological states, a common feature of polyglutaminopathies. However, for ataxin-3 (2UIM isoform) at least, work by Ellisdon and co-workers (Ellisdon, 2006) has demonstrated that this polyQ region is not directly involved in the initial steps of aggregation. Both expanded and non-expanded variants of the protein aggregate under physiologically-relevant conditions (100mM Tris pH 7.4, 80mM NaCl, 10% (v/v) glycerol, 2mM PMSF, 5mM EDTA, and 15mM BME; 37ºC; quiescent) *via* a nucleated assembly pathway, whose core nucleus, a monomer, is structurally identical, as indicated by their cross-seeding ability (Ellisdon, 2007). Nonetheless, polyQ expansion in ataxin-3 does produce fibrillar aggregates of different morphology and stability when compared to those of non-expanded protein (Ellisdon, 2006). As a possible explanation of these observations, the authors proposed a simplified model of aggregation, whereby ataxin-3 self-associates in two consecutive steps. The first, polyQ independent, is accompanied by a conformational change in the protein, particularly of the Josephin domain, leading to an increase in β-sheet content and the formation of amyloid-like fibrillar structures. These fibrils are typically soluble, worm-like in appearance, with a width of 10-12nm and variable length, consistent with those formed by the aggregation of other polyQ proteins and peptides such as Htt and Ataxin-1 (Chen, 2002 and de Chiara, 2005, respectively). The pathway leading to their formation, common to both expanded and non-expanded protein, includes an initial appearance of globular species, 5-20nm in diameter, progressively replaced as the dominant structure by short elongated structures of

10-12nm width, also called protofibrils, followed by their expansion to worm-like fibrils. At this point, a second step in the assembly, polyQ-dependent, mediates further association of these fibrils into much larger (50-100nm width) and insoluble fibrous structures. In addition, the structures formed gain an added stability, making them resistant to treatment with SDS. According to this model, the region(s) of the protein responsible for driving the initial assembly of the molecule must lie outside the polyQ tract, consistent with reports for other polyQ-containing proteins such as ataxin-1 (de Chiara, 2005) and huntingtin (Thakur, 2009). Natalello and co-workers showed that the appearance of SDS-resistant structures in the assembly of expanded ataxin-3 correlated with an increase in hydrogen-bonding between the side-chains of residues in the C-terminal part of the protein, according to constraints defined earlier in the assembly process by main-chain hydrogen bonding (Natalello, 2011). These examples stress the importance of the study of non-expanded ataxin-3 aggregation for the understanding of SCA3/MJD pathology. Whether toxicity arises from the added stability of expanded protein aggregates or directly from the initial aggregates common to expanded and non-expanded protein proteins, it is clear that non-expanded ataxin-3 is an important subject for the study of the MJD/SCA3 pathology.

Consistent with the model described above, the results presented in Chapters 5 and 6 show that the aggregation of AT3[U3/14Q] and AT3[2U/Jos] under near-physiological conditions (20mM HEPES pH 7.5, 200mM NaCl, 1mM DTT, 1mM EDTA, 0.01% NaN$_3$), result in the formation of species that are identical (over the time of incubation) to those described by Ellisdon and co-workers (Ellisdon, 2006). For both ataxin-3 protein isoforms, a monomeric thermodynamical nucleus was estimated, suggesting similar aggregation mechanisms (Ellisdon, 2007 and the work presented in this thesis). The same was shown here to hold true for the C-terminal truncation variant AT3[2U/Jos], sharing nucleation characteristics with the 3UIM isoform.

The N-terminal domain of ataxin-3 has a papain-like globular fold and is responsible for the catalytic activity of ataxin-3 (Mao, 2005). This domain has been proposed to have a role in initiating aggregation of the full-length protein under

near-physiological conditions (Chow, 2004; Gales, 2005; Masino, 2010 and 2011), supported by its ability to form dimers positive to A11 antibody (Kayed, 2003), a property in common with early oligomers of the full-length protein (Gales, 2005), and its direct involvement in the α-to-β conformational change (Masino, 2004; Gales, 2005; Ellisdon, 2006). Initial studies of the isolated Josephin domain showed its ability to self-assemble into fibrillar species when placed under high temperature or pressure (Marchal, 2003; Shehi, 2003; Masino, 2004) or in the presence of chemical chaotropes (Chow, 2004). However, in recent years, polypeptide chains have been shown to possess the capacity to self-assemble into fibrillar species under denaturing conditions such as these, and the study of protein aggregation under native-like conditions has gained increased interest (Chiti and Dobson, 2009). Extensive work in recent years by Masino and co-workers (Masino, 2010; Masino, 2011) has shown that the Josephin domain of ataxin-3 is capable of self-association into fibrillar structures morphologically identical to those formed by the full-length protein, when placed under non-denaturing conditions (20mM sodium phosphate, pH 6.5, 10mM TCEP and 0.025% NaN_3; 37ºC; quiescent). These results strengthened the hypothesis of a role for the Josephin domain in mediating early events in ataxin-3 aggregation, postulated by several groups (Masino, 2004; Gales, 2005; Robertson, 2010). Interestingly, two hydrophobic patches on the surface of this domain have been identified as aggregation-prone, and mutations in these regions shown to slow, though not abolish, the kinetics of Josephin aggregation (Masino, 2011). These sites also correspond to substrate binding sites (Nicastro, 2009) and, accordingly, the addition of a molar excess of ubiquitin molecules to the aggregation reaction was shown to have similar results to those observed with mutational studies, where the kinetics of assembly are slowed down but not abolished. Aggregates formed by the Josephin domain in the presence of ubiquitin, however, have a different morphology to those of the Josephin domain protein alone, indicating a possible accommodation of the ubiquitin molecules in the structures formed (Masino, 2011). Similarly, work by Robertson and co-workers with small heat shock proteins has shown that the aggregation of ataxin-3 can be inhibited by the interaction with alphaB-crystallin (a small heat-shock protein) at

the surface of the Josephin domain, though this same interaction is not sufficient to inhibit the aggregation of expanded protein, resulting only in slower kinetics (Robertson, 2010).

Contrary to these reports, the results shown in Chapter 6 for AT3[Jos] under physiologically-relevant conditions (20mM HEPES pH 7.5, 200mM NaCl, 1mM DTT, 1mM EDTA, 0.01% NaN$_3$; 37ºC; quiescent) show that this domain aggregates *via* a non-nucleated process, unlike that of full-length protein (as described in Chapter 5 and Ellisdon, 2006). In other words, the aggregation of the Josephin domain appears off-pathway to the assembly process of full-length protein, though the domain is nonetheless aggregation prone, and could possibly influence the kinetics of aggregation of the full-length protein (and the truncation mutant containing the Josephin domain and the two tandem UIMs) by a surface-mediated catalysis of aggregation. Instead, the results described in this thesis point to the stretch of residues between the Josephin domain and the polyQ tract in ataxin-3, as the nucleating core of aggregation. This region is thought to be partially unstructured, with two tandem Ubiquitin interacting motifs (UIMs) as the only predicted structured stretches. Since natively unfolded stretches are rarely aggregation-prone (although there are notable exceptions, reviewed by Jahn and Radford, 2008), the UIM sequences gain an increased interest in the study of ataxin-3 aggregation. Curiously, the UIM sequences of ataxin-3 (3UIM isoform), as well as those of S5a (a ubiquitin-binding protein found as a subunit of the 26S proteasome), have been shown by Miller and co-workers to inhibit the aggregation of expanded Huntingtin protein when co-expressed in cells (Miller, 2007), *via* a process that requires both a cooperativity of all three UIM sequences and an interaction with SUMO (small ubiquitin modifier), an ubiquitin-like protein known to be involved in Huntington's pathology (Steffan, 2004), although its actual role is unclear.

UIMs exist in several proteins (Hoffman and Falquet, 2001), mediating interaction with ubiquitin or ubiquitin-like molecules, as their name indicates. These sequences are intrinsically α-helical (Mueller, 2003; Swanson, 2003), and form a hydrophobic face that complements a hydrophobic patch on the surface of its interacting partners. Conserved residues in UIM sequences mediate this

interaction and, interestingly, have been found to mediate the formation of a four-helix bundle in the crystallographic structure of UIM2 of Vps27 (Fisher, 2003), though no physiologically relevant tetramer has been found. Work by Song and co-workers with the first two UIMs of ataxin-3 (Song, 2010), presented in tandem on both AT3[U3/14Q] and AT3[2U/Jos], have confirmed their α-helical character and shown a cooperativity in binding, consistent with the results mentioned above by Miller and co-workers. The linker region between the UIMs is disordered and flexible in solution (Song, 2010), adopting a more rigid conformation upon binding, and increasing efficiency of ubiquitin binding even though it is not directly involved in the binding itself. Similarly, should these UIMs be responsible for the nucleating core of ataxin-3 aggregation, as the results in this thesis suggest, a rigid structure such as described by the authors could act as a scaffold for the assembly, mediated by hydrophobic interactions between conserved residues, as seen in the crystal packing of Vcp27's UIM2 (Fisher, 2003). Furthermore, the interaction of VCP (Valosin-containing protein) with ataxin-3, mapped to the region between UIM2 and the polyQ tract (Boeddrich, 2006), could affect the aggregation of ataxin-3 by disrupting the formation of the nucleating core. As proposed by Boeddrich and co-workers, high concentrations of VCP *in vitro* lead to steric hindrance of assembly by binding directly to ataxin-3, inhibiting aggregation. However, to explain the observation that low concentrations of VCP enhance ataxin-3 aggregation (Boeddrich, 2006), other results need to be considered. The hexameric structure of VCP offers 6 possible binding sites for ataxin-3 (all in the N-terminal portion of each subunit), and was suggested by Boeddrich and co-workers to act as a catalytic surface for the conversion of ataxin-3 into an aggregation-competent species. The work presented in this thesis offers another explanation, from the data collected with the Josephin domain in Chapter 6, together with the work by Masino and co-workers in mapping aggregation prone regions on the surface of Josephin (Masino, 2011): the domain is aggregation-prone and capable of undergoing a surface-catalysed polymerization. As such, the proximity of two or more ataxin-3 molecules, instigated by their interaction with a VCP subunit, can lead to a Josephin-mediated interaction, further progressed by surface catalysis, leading to an increase in

aggregation. In addition, Boeddrich and co-workers also reported that coexpressing VCP and expanded ataxin-3 in *Drosophila melanogaster* does not lead to increased degeneration, regardless of the expression level, proposing that other interacting partners could be involved *in vivo*. Based on the work with the Josephin domain by Masino, Nicastro and co-workers (Nicastro, 2009; Masino, 2011), a good candidate for this "extra" interaction would be ubiquitin, shielding the hydrophobic patches on the surface of Josephin. This hypothesis stresses the importance of the exact conditions of aggregation and the presence of three distinct aggregation-prone regions in ataxin-3.

In fact, the results for the fibril formation of the Josephin domain obtained by Masino and co-workers (Masino, 2011), seemingly contradictory to those presented in this thesis, may actually have an explanation rooted in the different aggregation conditions used. The results in Section 5.4.2 with varying concentrations of NaCl, point to an effect on the aggregation kinetics by the absence or presence of the salt and, perhaps even more revealing, the results in Section 5.4.1 with aggregation at varying pH values, show a striking, and unexpected, correlation between the aggregation of full-length ataxin-3 at pH 6.5 and the aggregation of the Josephin domain at pH 7.5. Specifically, the aggregation kinetics of full-length protein at pH 6.5 show an overshoot of the intensity at the end of the growth phase, very similar to that observed consistently for the Josephin domain (Chapter 6). This feature, and its consistent observation, suggests a possible influence of the Josephin domain on the aggregation of full-length protein at pH 6.5, the same pH value used by Nicastro, Masino and co-workers (Nicastro, 2009; Masino, 2011).

In summary, three aggregation-prone regions in ataxin-3 are now identified, with differential influence on the assembly process of the protein. The nucleating core of aggregation is located between the Josephin domain and the polyQ region, possibly involving hydrophobic interactions mediated by conserved residues of the UIM sequences. On either side of the core there is another aggregation-prone region, with the polyQ stretch, C-terminal to it, involved in stabilizing the fibrillar aggregates once formed, likely by providing an added array of side-chain hydrogen

bonding. N-terminal to this core region, the Josephin domain provides an alternate assembly pathway that may work either in enhancing aggregation rates, by aiding in the recruitment of molecules to growing aggregates or, alternatively, slowing aggregation rates by hijacking molecules into this alternate pathway. Given its size and fold, this domain may also be involved in determining the final morphology of the fibrils by forcing a twist in the growing fibril to relieve steric tension. Much more work is needed to fully understand the role of these domains, and particular attention to the protein's role *in vivo* is suggested by both the work of Masino and co-workers on the interacting surfaces of Josephin (Masino, 2011) and the results presented in this thesis, which point to a role of the tandem UIM region in nucleating aggregation.

With the high-throughput experimental setup designed in this thesis for the reproducible determination of aggregation parameters, a more detailed understanding of the aggregation kinetics of ataxin-3 in response to environmental factors (pH, ionic strength, and molecular interactants) can be achieved. Coupled with mutational studies designed to disrupt the hydrophobic interactions of both the UIM α-helices and the substrate-binding patches on the surface of the Josephin domain, future studies could provide valuable insights into the molecular details of ataxin-3 aggregation *in vivo*, both in pathological and non-pathological states, allowing the identification of targets for therapeutical strategies. Alternatively, a shortcut to the development of these therapeutic strategies may be found using the experimental setup developed herein, and the characterization of the aggregates formed as described in this thesis. Screening of compounds for small molecules able to prevent protein aggregation or, at least, reduce aggregation rates *in vitro*, with the ultimate goal of further development from these leads, may in the future allow the patient to live his/her life without manifestation of the disease.

Chapter **8**

8 – Reference list

- Albrecht, M., D. Hoffmann, et al. (2003). "Structural modeling of ataxin-3 reveals distant homology to adaptins." Proteins **50**(2): 355-70.

- Alexandrescu, A. T. (2005). "Amyloid accomplices and enforcers." Protein Sci **14**(1): 1-12.

- Alves, S., I. Nascimento-Ferreira, et al. (2010). "Silencing ataxin-3 mitigates degeneration in a rat model of Machado-Joseph disease: no role for wild-type ataxin-3?" Hum Mol Genet **19**(12): 2380-94.

- Andersen, C. B., H. Yagi, et al. (2009). "Branching in amyloid fibril growth." Biophys J **96**(4): 1529-36.

- Arendash, G. W., M. F. Garcia, et al. (2004). "Environmental enrichment improves cognition in aged Alzheimer's transgenic mice despite stable beta-amyloid deposition." Neuroreport **15**(11): 1751-4.

- Baker, R. T. and P. G. Board (1991). "The human ubiquitin-52 amino acid fusion protein gene shares several structural features with mammalian ribosomal protein genes." Nucleic Acids Res **19**(5): 1035-40.

- Balguerie, A., S. Dos Reis, et al. (2003). "Domain organization and structure-function relationship of the HET-s prion protein of Podospora anserina." Embo J **22**(9): 2071-81.

- Baxa, U., K. L. Taylor, et al. (2003). "Architecture of Ure2p prion filaments: the N-terminal domains form a central core fiber." J Biol Chem **278**(44): 43717-27.

- Baxa, U., N. Cheng, et al. (2005). "Filaments of the Ure2p prion protein have a cross-beta core structure." J Struct Biol **150**(2): 170-9.

- Belli, M., M. Ramazzotti, et al. (2011). "Prediction of amyloid aggregation in vivo." EMBO Rep **12**(7): 657-63.

- Benseny-Cases, N., M. Cocera, et al. (2007). "Conversion of non-fibrillar beta-sheet oligomers into amyloid fibrils in Alzheimer's disease amyloid peptide aggregation." Biochem Biophys Res Commun **361**(4): 916-21.

- Berke, S. J., Y. Chai, et al. (2005). "Defining the role of ubiquitin-interacting motifs in the polyglutamine disease protein, ataxin-3." J Biol Chem **280**(36): 32026-34.

- Berryman, J. T., S. E. Radford, et al. (2009). "Thermodynamic description of polymorphism in Q- and N-rich peptide aggregates revealed by atomistic simulation." Biophys J **97**(1): 1-11.

- Bettencourt, C. and M. Lima (2010). "Machado-Joseph Disease: from first descriptions to new perspectives." Orphanet J Rare Dis **6**: 35.

- Bevivino, A. E. and P. J. Loll (2001). "An expanded glutamine repeat destabilizes native ataxin-3 structure and mediates formation of parallel beta -fibrils." Proc Natl Acad Sci U S A **98**(21): 11955-60.

- Bhandari, P. and J. Gowrishankar (1997). "An Escherichia coli host strain useful for efficient overproduction of cloned gene products with NaCl as the inducer." J Bacteriol **179**(13): 4403-6.

- Bichelmeier, U., T. Schmidt, et al. (2007). "Nuclear localization of ataxin-3 is required for the manifestation of symptoms in SCA3: in vivo evidence." J Neurosci **27**(28): 7418-28.

- Boeddrich, A., S. Gaumer, et al. (2006). "An arginine/lysine-rich motif is crucial for VCP/p97-mediated modulation of ataxin-3 fibrillogenesis." Embo J **25**(7): 1547-58.

- Boland, B., A. Kumar, et al. (2008). "Autophagy induction and autophagosome clearance in neurons: relationship to autophagic pathology in Alzheimer's disease." J Neurosci **28**(27): 6926-37.

- Boland, B., D. A. Smith, et al. (2010). "Macroautophagy is not directly involved in the metabolism of amyloid precursor protein." J Biol Chem **285**(48): 37415-26.

- Booth, D. R., M. Sunde, et al. (1997). "Instability, unfolding and aggregation of human lysozyme variants underlying amyloid fibrillogenesis." Nature **385**(6619): 787-93.

- Buchberger, A., B. Bukau, et al. (2010). "Protein quality control in the cytosol and the endoplasmic reticulum: brothers in arms." Mol Cell **40**(2): 238-52.

- Burnett, B., F. Li, et al. (2003). "The polyglutamine neurodegenerative protein ataxin-3 binds polyubiquitylated proteins and has ubiquitin protease activity." Hum Mol Genet **12**(23): 3195-205.

- Cardoso, I., C. S. Goldsbury, et al. (2002). "Transthyretin fibrillogenesis entails the assembly of monomers: a molecular model for in vitro assembled transthyretin amyloid-like fibrils." J Mol Biol **317**(5): 683-95.

- Caspar, D. L. (2009). "Inconvenient facts about pathological amyloid fibrils." Proc Natl Acad Sci U S A **106**(49): 20555-6.

- Chai, Y., S. L. Koppenhafer, et al. (1999). "Evidence for proteasome involvement in polyglutamine disease: localization to nuclear inclusions in SCA3/MJD and suppression of polyglutamine aggregation in vitro." Hum Mol Genet **8**(4): 673-82.

- Chai, Y., S. S. Berke, et al. (2004). "Poly-ubiquitin binding by the polyglutamine disease protein ataxin-3 links its normal function to protein surveillance pathways." J Biol Chem **279**(5): 3605-11.

- Chayen, N. E. and E. Saridakis (2008). "Protein crystallization: from purified protein to diffraction-quality crystal." Nat Methods **5**(2): 147-53.

- Chen, S., F. A. Ferrone, et al. (2002). "Huntington's disease age-of-onset linked to polyglutamine aggregation nucleation." Proc Natl Acad Sci U S A **99**(18): 11884-9.

- Chen, S., V. Berthelier, et al. (2001). "Polyglutamine aggregation behavior in vitro supports a recruitment mechanism of cytotoxicity." J Mol Biol **311**(1): 173-82.

- Chiti, F. and C. M. Dobson (2006). "Protein misfolding, functional amyloid, and human disease." Annu Rev Biochem **75**: 333-66.

- Chiti, F., M. Stefani, et al. (2003). "Rationalization of the effects of mutations on peptide and protein aggregation rates." Nature **424**(6950): 805-8.

- Chow, M. K., A. M. Ellisdon, et al. (2006). "Purification of polyglutamine proteins." Methods Enzymol **413**: 1-19.

- Chow, M. K., H. L. Paulson, et al. (2004). "Destabilization of a non-pathological variant of ataxin-3 results in fibrillogenesis via a partially folded intermediate: a model for misfolding in polyglutamine disease." J Mol Biol **335**(1): 333-41.

- Clamp, M., J. Cuff, et al. (2004). "The Jalview Java alignment editor." Bioinformatics **20**(3): 426-7.

- Correia, B. E., N. Loureiro-Ferreira, et al. (2006). "A structural model of an amyloid protofilament of transthyretin." Protein Sci **15**(1): 28-32.

- Cuervo, A. M. (2011). "Chaperone-mediated autophagy: Dice's 'wild' idea about lysosomal selectivity." Nat Rev Mol Cell Biol **12**(8): 535-41.

- Cummings, C. J., M. A. Mancini, et al. (1998). "Chaperone suppression of aggregation and altered subcellular proteasome localization imply protein misfolding in SCA1." Nat Genet **19**(2): 148-54.

- Cummings, C. J., Y. Sun, et al. (2001). "Over-expression of inducible HSP70 chaperone suppresses neuropathology and improves motor function in SCA1 mice." Hum Mol Genet **10**(14): 1511-8.

- Dahlgren, P. R., M. A. Karymov, et al. (2005). "Atomic force microscopy analysis of the Huntington protein nanofibril formation." Nanomedicine **1**(1): 52-7.

- Damas, A. M. and M. J. Saraiva (2000). "Review: TTR amyloidosis-structural features leading to protein aggregation and their implications on therapeutic strategies." J Struct Biol **130**(2-3): 290-9.

- Dang, L. C., F. D. Melandri, et al. (1998). "Kinetic and mechanistic studies on the hydrolysis of ubiquitin C-terminal 7-amido-4-methylcoumarin by deubiquitinating enzymes." Biochemistry **37**(7): 1868-79.

- de Chiara, C., R. P. Menon, et al. (2005). "The AXH domain adopts alternative folds the solution structure of HBP1 AXH." Structure **13**(5): 743-53.

- Dikic, I., S. Wakatsuki, et al. (2009). "Ubiquitin-binding domains - from structures to functions." Nat Rev Mol Cell Biol **10**(10): 659-71.

- Donaldson, K. M., W. Li, et al. (2003). "Ubiquitin-mediated sequestration of normal cellular proteins into polyglutamine aggregates." Proc Natl Acad Sci U S A **100**(15): 8892-7.

- Doss-Pepe, E. W., E. S. Stenroos, et al. (2003). "Ataxin-3 interactions with rad23 and valosin-containing protein and its associations with ubiquitin chains and the proteasome are consistent with a role in ubiquitin-mediated proteolysis." Mol Cell Biol **23**(18): 6469-83.

- Dougan, L., J. Li, et al. (2009). "Single homopolypeptide chains collapse into mechanically rigid conformations." Proc Natl Acad Sci U S A **106**(31): 12605-10.

- DuBay, K. F., A. P. Pawar, et al. (2004). "Prediction of the absolute aggregation rates of amyloidogenic polypeptide chains." J Mol Biol **341**(5): 1317-26.

- Durr, A., G. Stevanin, et al. (1996). "Spinocerebellar ataxia 3 and Machado-Joseph disease: clinical, molecular, and neuropathological features." Ann Neurol **39**(4): 490-9.

- Ellisdon, A. M., B. Thomas, et al. (2006). "The two-stage pathway of ataxin-3 fibrillogenesis involves a polyglutamine-independent step." J Biol Chem **281**(25): 16888-96.

- Ellisdon, A. M., M. C. Pearce, et al. (2007). "Mechanisms of ataxin-3 misfolding and fibril formation: kinetic analysis of a disease-associated polyglutamine protein." J Mol Biol **368**(2): 595-605.

- Esposito, D. and D. K. Chatterjee (2006). "Enhancement of soluble protein expression through the use of fusion tags." Curr Opin Biotechnol **17**(4): 353-8.

- Esposito, L., A. Paladino, et al. (2008). "Insights into structure, stability, and toxicity of monomeric and aggregated polyglutamine models from molecular dynamics simulations." Biophys J **94**(10): 4031-40.

- Fandrich, M., J. Meinhardt, et al. (2009). "Structural polymorphism of Alzheimer Abeta and other amyloid fibrils." Prion **3**(2): 89-93.

- Fernandez-Escamilla, A. M., F. Rousseau, et al. (2004). "Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins." Nat Biotechnol **22**(10): 1302-6.

- Ferrone, F. (1999). "Analysis of protein aggregation kinetics." Methods Enzymol **309**: 256-74.

- Ferrone, F. A., M. Ivanova, et al. (2002). "Heterogeneous nucleation and crowding in sickle hemoglobin: an analytic approach." Biophys J **82**(1 Pt 1): 399-406.

- Fisher, R. D., B. Wang, et al. (2003). "Structure and ubiquitin binding of the ubiquitin-interacting motif." J Biol Chem **278**(31): 28976-84.

- Fowler, D. M., A. V. Koulov, et al. (2007). "Functional amyloid--from bacteria to humans." Trends Biochem Sci **32**(5): 217-24.

- Fu, Y. H., D. P. Kuhl, et al. (1991). "Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox." Cell **67**(6): 1047-58.

- Fujigasaki, H., T. Uchihara, et al. (2000). "Ataxin-3 is translocated into the nucleus for the formation of intranuclear inclusions in normal and Machado-Joseph disease brains." Exp Neurol **165**(2): 248-56.

- Fujigasaki, H., T. Uchihara, et al. (2001). "Preferential recruitment of ataxin-3 independent of expanded polyglutamine: an immunohistochemical study on Marinesco bodies." J Neurol Neurosurg Psychiatry **71**(4): 518-20.

- Gales, L., L. Cortes, et al. (2005). "Towards a structural understanding of the fibrillization pathway in Machado-Joseph's disease: trapping early oligomers of non-expanded ataxin-3." J Mol Biol **353**(3): 642-54.

- Geddes, A. J., K. D. Parker, et al. (1968). ""Cross-beta" conformation in proteins." J Mol Biol **32**(2): 343-58.

- Giehm, L., N. Lorenzen, et al. (2010). "Assays for alpha-synuclein aggregation." Methods **53**(3): 295-305.

- Glabe, C. G. (2008). "Structural classification of toxic amyloid oligomers." J Biol Chem **283**(44): 29639-43.

- Gosal, W. S., I. J. Morten, et al. (2005). "Competing pathways determine fibril morphology in the self-assembly of beta2-microglobulin into amyloid." J Mol Biol **351**(4): 850-64.

- Goto, J., M. Watanabe, et al. (1997). "Machado-Joseph disease gene products carrying different carboxyl termini." Neurosci Res **28**(4): 373-7.

- Haacke, A., S. A. Broadley, et al. (2006). "Proteolytic cleavage of polyglutamine-expanded ataxin-3 is critical for aggregation and sequestration of non-expanded ataxin-3." Hum Mol Genet **15**(4): 555-68.

- Haglund, K. and I. Dikic (2005). "Ubiquitylation and cell signaling." Embo J **24**(19): 3353-9.

- Hamada, D., T. Tanaka, et al. (2009). "Competition between folding, native-state dimerisation and amyloid aggregation in beta-lactoglobulin." J Mol Biol **386**(3): 878-90.

- Hammarstrom, P., X. Jiang, et al. (2002). "Sequence-dependent denaturation energetics: A major determinant in amyloid disease diversity." Proc Natl Acad Sci U S A **99 Suppl 4**: 16427-32.

- Harding, A. E. (1993). "Clinical features and classification of inherited ataxias." Adv Neurol **61**: 1-14.

- Hardy, J. and D. J. Selkoe (2002). "The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics." Science **297**(5580): 353-6.

- Harper, J. D. and P. T. Lansbury, Jr. (1997). "Models of amyloid seeding in Alzheimer's disease and scrapie: mechanistic truths and physiological consequences of the time-dependent solubility of amyloid proteins." Annu Rev Biochem **66**: 385-407.

- Harris, G. M., K. Dodelzon, et al. (2010). "Splice isoforms of the polyglutamine disease protein ataxin-3 exhibit similar enzymatic yet different aggregation properties." PLoS One **5**(10): e13695.

- Hartley, D. M., D. M. Walsh, et al. (1999). "Protofibrillar intermediates of amyloid beta-protein induce acute electrophysiological changes and progressive neurotoxicity in cortical neurons." J Neurosci **19**(20): 8876-84.

- Hashida, H., J. Goto, et al. (1997). "Brain regional differences in the expansion of a CAG repeat in the spinocerebellar ataxias: dentatorubral-pallidoluysian atrophy, Machado-Joseph disease, and spinocerebellar ataxia type 1." Ann Neurol **41**(4): 505-11.

- Havel, L. S., S. Li, et al. (2009). "Nuclear accumulation of polyglutamine disease proteins and neuropathology." Mol Brain **2**: 21.

- Hayden, M. S. and S. Ghosh (2008). "Shared principles in NF-kappaB signaling." Cell **132**(3): 344-62.

- Hebert, D. N. and M. Molinari (2007). "In and out of the ER: protein folding, quality control, degradation, and related human diseases." Physiol Rev **87**(4): 1377-408.

- Hirschfield, G. M. and P. N. Hawkins (2003). "Amyloidosis: new strategies for treatment." Int J Biochem Cell Biol **35**(12): 1608-13.

- Hofmann, K. and L. Falquet (2001). "A ubiquitin-interacting motif conserved in components of the proteasomal and lysosomal protein degradation systems." Trends Biochem Sci **26**(6): 347-50.

- Husain-Ponnampalam, R., V. Turnbull, et al. (2010). "Expression and purification of ataxin-1 protein." J Neurosci Methods **189**(1): 30-5.

- Ignatova, Z., A. K. Thakur, et al. (2007). "In-cell aggregation of a polyglutamine-containing chimera is a multistep process initiated by the flanking sequence." J Biol Chem **282**(50): 36736-43.

- Ikeda, F. and I. Dikic (2008). "Atypical ubiquitin chains: new molecular signals. 'Protein Modifications: Beyond the Usual Suspects' review series." EMBO Rep **9**(6): 536-42.

- Jahn, T. R. and S. E. Radford (2008). "Folding versus aggregation: polypeptide conformations on competing pathways." Arch Biochem Biophys **469**(1): 100-17.

- Jahn, T. R., M. J. Parker, et al. (2006). "Amyloid formation under physiological conditions proceeds via a native-like folding intermediate." Nat Struct Mol Biol **13**(3): 195-201.

- Jayaraman, M., R. Kodali, et al. (2009). "The impact of ataxin-1-like histidine insertions on polyglutamine aggregation." Protein Eng Des Sel **22**(8): 469-78.

- Kahle, P. J. and C. Haass (2004). "How does parkin ligate ubiquitin to Parkinson's disease?" EMBO Rep **5**(7): 681-5.

- Kawaguchi, Y., T. Okamoto, et al. (1994). "CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32.1." Nat Genet **8**(3): 221-8.

- Kayed, R., E. Head, et al. (2003). "Common structure of soluble amyloid oligomers implies common mechanism of pathogenesis." Science **300**(5618): 486-9.

- Kayed, R., E. Head, et al. (2007). "Fibril specific, conformation dependent antibodies recognize a generic epitope common to amyloid fibrils and fibrillar oligomers that is absent in prefibrillar oligomers." Mol Neurodegener **2**: 18.

- Kettner, M., D. Willwohl, et al. (2002). "Intranuclear aggregation of nonexpanded ataxin-3 in marinesco bodies of the nonhuman primate substantia nigra." Exp Neurol **176**(1): 117-21.

- Kirkin, V., D. G. McEwan, et al. (2009). "A role for ubiquitin in selective autophagy." Mol Cell **34**(3): 259-69.

- Klunk, W. E., R. F. Jacob, et al. (1999). "Quantifying amyloid by congo red spectral shift assay." Methods Enzymol **309**: 285-305.

- Knowles, T. P., C. A. Waudby, et al. (2009). "An analytical solution to the kinetics of breakable filament assembly." Science **326**(5959): 1533-7.

- La Spada, A. R., E. M. Wilson, et al. (1991). "Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy." Nature **352**(6330): 77-9.

- Lee, S., E. J. Fernandez, et al. (2007). "Role of aggregation conditions in structure, stability, and toxicity of intermediates in the Abeta fibril formation pathway." Protein Sci **16**(4): 723-32.

- LeVine, H., 3rd (1999). "Quantification of beta-sheet amyloid fibril structures with thioflavin T." Methods Enzymol **309**: 274-84.

- Li, D. W., S. Mohanty, et al. (2008). "Formation and growth of oligomers: a Monte Carlo study of an amyloid tau fragment." PLoS Comput Biol **4**(12): e1000238.

- Li, S., T. H. Nguyen, et al. (1995). "Aggregation and precipitation of human relaxin induced by metal-catalyzed oxidation." Biochemistry **34**(17): 5762-72.

- Luheshi, L. M., G. G. Tartaglia, et al. (2007). "Systematic in vivo analysis of the intrinsic determinants of amyloid Beta pathogenicity." PLoS Biol **5**(11): e290.

- Macedo-Ribeiro, S., L. Cortes, et al. (2009). "Nucleocytoplasmic shuttling activity of ataxin-3." PLoS One **4**(6): e5834.

- Majhi, P. R., R. R. Ganta, et al. (2006). "Electrostatically driven protein aggregation: beta-lactoglobulin at low ionic strength." Langmuir **22**(22): 9150-9.

- Makin, O. S. and L. C. Serpell (2005). "Structures for amyloid fibrils." Febs J **272**(23): 5950-61.

- Mao, X., C. Wang, et al. (2011). "Molecular level studies on binding modes of labeling molecules with polyalanine peptides." Nanoscale **3**(4): 1592-9.

- Mao, Y., F. Senic-Matuglia, et al. (2005). "Deubiquitinating function of ataxin-3: insights from the solution structure of the Josephin domain." Proc Natl Acad Sci U S A **102**(36): 12700-5.

- Marchal, S., E. Shehi, et al. (2003). "Structural instability and fibrillar aggregation of non-expanded human ataxin-3 revealed under high pressure and temperature." J Biol Chem **278**(34): 31554-63.

- Masino, L., G. Nicastro, et al. (2004). "Characterization of the structure and the amyloidogenic properties of the Josephin domain of the polyglutamine-containing protein ataxin-3." J Mol Biol **344**(4): 1021-35.

- Masino, L., G. Nicastro, et al. (2011). "Functional interactions as a survival strategy against abnormal aggregation." Faseb J **25**(1): 45-54.

- Masino, L., G. Nicastro, et al. (2011). "The Josephin domain determines the morphological and mechanical properties of ataxin-3 fibrils." Biophys J **100**(8): 2033-42.

- Masino, L., V. Musi, et al. (2003). "Domain architecture of the polyglutamine protein ataxin-3: a globular domain followed by a flexible tail." FEBS Lett **549**(1-3): 21-5.

- Matos, C. A., S. de Macedo-Ribeiro, et al. (2011). "Polyglutamine diseases: The special case of ataxin-3 and Machado-Joseph disease." Prog Neurobiol **95**(1): 26-48.

- McLean, C. A., R. A. Cherny, et al. (1999). "Soluble pool of Abeta amyloid as a determinant of severity of neurodegeneration in Alzheimer's disease." Ann Neurol **46**(6): 860-6.

- Milhiet, P. E., D. Yamamoto, et al. (2010). "Deciphering the structure, growth and assembly of amyloid-like fibrils using high-speed atomic force microscopy." PLoS One **5**(10): e13240.

- Miller, S. L., E. L. Scappini, et al. (2007). "Ubiquitin-interacting motifs inhibit aggregation of polyQ-expanded huntingtin." J Biol Chem **282**(13): 10096-103.

- Milner-White, J. E., J. D. Watson, et al. (2006). "Amyloid formation may involve alpha- to beta sheet interconversion via peptide plane flipping." Structure **14**(9): 1369-76.

- Modler, A. J., K. Gast, et al. (2003). "Assembly of amyloid protofibrils via critical oligomers--a novel pathway of amyloid formation." J Mol Biol **325**(1): 135-48.

- Muchowski, P. J. and J. L. Wacker (2005). "Modulation of neurodegeneration by molecular chaperones." Nat Rev Neurosci **6**(1): 11-22.

- Mueller, T. D. and J. Feigon (2003). "Structural determinants for the binding of ubiquitin-like domains to the proteasome." Embo J **22**(18): 4634-45.

- Mueller, T., P. Breuer, et al. (2009). "CK2-dependent phosphorylation determines cellular localization and stability of ataxin-3." Hum Mol Genet **18**(17): 3334-43.

- Myers, S. L., S. Jones, et al. (2006). "A systematic study of the effect of physiological factors on beta2-microglobulin amyloid formation at neutral pH." Biochemistry **45**(7): 2311-21.

- Naiki, H. and Y. Nagai (2009). "Molecular pathogenesis of protein misfolding diseases: pathological molecular environments versus quality control systems against misfolded proteins." J Biochem **146**(6): 751-6.

- Natalello, A., A. M. Frana, et al. (2011). "A major role for side-chain polyglutamine hydrogen bonding in irreversible ataxin-3 aggregation." PLoS One **6**(4): e18789.

- Nerelius, C., M. Fitzen, et al. (2010). "Amino acid sequence determinants and molecular chaperones in amyloid fibril formation." Biochem Biophys Res Commun **396**(1): 2-6.

- Nicastro, G., L. Masino, et al. (2004). "Assignment of the 1H, 13C, and 15N resonances of the Josephin domain of human ataxin-3." J Biomol NMR **30**(4): 457-8.

- Nicastro, G., L. Masino, et al. (2009). "Josephin domain of ataxin-3 contains two distinct ubiquitin-binding sites." Biopolymers **91**(12): 1203-14.

- Nicastro, G., M. Habeck, et al. (2006). "Structure validation of the Josephin domain of ataxin-3: conclusive evidence for an open conformation." J Biomol NMR **36**(4): 267-77.

- Nicastro, G., R. P. Menon, et al. (2005). "The solution structure of the Josephin domain of ataxin-3: structural determinants for molecular recognition." Proc Natl Acad Sci U S A **102**(30): 10493-8.

- Nicastro, G., S. V. Todi, et al. (2010). "Understanding the role of the Josephin domain in the PolyUb binding and cleavage properties of ataxin-3." PLoS One **5**(8): e12430.

- Nichols, M. R., M. A. Moss, et al. (2005). "Amyloid-beta aggregates formed at polar-nonpolar interfaces differ from amyloid-beta protofibrils produced in aqueous buffers." Microsc Res Tech **67**(3-4): 164-74.

- Nijman, S. M., M. P. Luna-Vargas, et al. (2005). "A genomic and functional inventory of deubiquitinating enzymes." Cell **123**(5): 773-86.

- Oosawa, F. and Asakura, S. (1975). *Thermodynamics of the polymerization of proteins.* Academic Press, New York.

- Padrick, S. B. and A. D. Miranker (2002). "Islet amyloid: phase partitioning and secondary nucleation are central to the mechanism of fibrillogenesis." Biochemistry **41**(14): 4694-703.

- Paulson, H. L., M. K. Perez, et al. (1997). "Intranuclear inclusions of expanded polyglutamine protein in spinocerebellar ataxia type 3." Neuron **19**(2): 333-44.

- Paushkin, S. V., V. V. Kushnirov, et al. (1996). "Propagation of the yeast prion-like [psi+] determinant is mediated by oligomerization of the SUP35-encoded polypeptide chain release factor." Embo J **15**(12): 3127-34.

- Perez, M. K., H. L. Paulson, et al. (1999). "Ataxin-3 with an altered conformation that exposes the polyglutamine domain is associated with the nuclear matrix." Hum Mol Genet **8**(13): 2377-85.

- Petty, S. A. and S. M. Decatur (2005). "Intersheet rearrangement of polypeptides during nucleation of {beta}-sheet aggregates." Proc Natl Acad Sci U S A **102**(40): 14272-7.

- Powers, E. T. and D. L. Powers (2008). "Mechanisms of protein fibril formation: nucleated polymerization with competing off-pathway aggregation." Biophys J **94**(2): 379-91.

- Rabinowitz, J. D. and E. White (2010). "Autophagy and metabolism." <u>Science</u> **330**(6009): 1344-8.

- Ravikumar, B., C. Vacher, et al. (2004). "Inhibition of mTOR induces autophagy and reduces toxicity of polyglutamine expansions in fly and mouse models of Huntington disease." <u>Nat Genet</u> **36**(6): 585-95.

- Reinke, A. A. and J. E. Gestwicki (2011). "Insight into amyloid structure using chemical probes." <u>Chem Biol Drug Des</u> **77**(6): 399-411.

- Ricchelli, F., P. Fusi, et al. (2007). "Destabilization of non-pathological variants of ataxin-3 by metal ions results in aggregation/fibrillogenesis." <u>Int J Biochem Cell Biol</u> **39**(5): 966-77.

- Robertson, A. L., J. Horne, et al. (2008). "The structural impact of a polyglutamine tract is location-dependent." <u>Biophys J</u> **95**(12): 5922-30.

- Robertson, A. L., S. J. Headey, et al. (2010). "Small heat-shock proteins interact with a flanking domain to suppress polyglutamine aggregation." <u>Proc Natl Acad Sci U S A</u> **107**(23): 10424-9.

- Roche Applied Biosciences, "The complete guide for protease inihibition". http://www.roche-applied-science.com

- Rodrigues, J. R., C. J. Simoes, et al. (2010). "Potentially amyloidogenic conformational intermediates populate the unfolding landscape of transthyretin: insights from molecular dynamics simulations." <u>Protein Sci</u> **19**(2): 202-19.

- Ross, C. A. and M. A. Poirier (2004). "Protein aggregation and neurodegenerative disease." <u>Nat Med</u> **10 Suppl**: S10-7.

- Rotter, M. A., S. Kwong, et al. (2005). "Heterogeneous nucleation in sickle hemoglobin: experimental validation of a structural mechanism." <u>Biophys J</u> **89**(4): 2677-84.

- Ruschak, A. M. and A. D. Miranker (2007). "Fiber-dependent amyloid formation as catalysis of an existing reaction pathway." <u>Proc Natl Acad Sci U S A</u> **104**(30): 12341-6.

- Ruschak, A. M. and A. D. Miranker (2009). "The role of prefibrillar structures in the assembly of a peptide amyloid." <u>J Mol Biol</u> **393**(1): 214-26.

- Saunders, H. M., D. Gilis, et al. (2011). "Flanking domain stability modulates the aggregation kinetics of a polyglutamine disease protein." Protein Sci.

- Schmittschmitt, J. P. and J. M. Scholtz (2003). "The role of protein stability, solubility, and net charge in amyloid fibril formation." Protein Sci **12**(10): 2374-8.

- Schwartz, D. C. and M. Hochstrasser (2003). "A superfamily of protein tags: ubiquitin, SUMO and related modifiers." Trends Biochem Sci **28**(6): 321-8.

- Schwartz, R., S. Istrail, et al. (2001). "Frequencies of amino acid strings in globular protein sequences indicate suppression of blocks of consecutive hydrophobic residues." Protein Sci **10**(5): 1023-31.

- Sen, S., D. Dash, et al. (2003). "Role of histidine interruption in mitigating the pathological effects of long polyglutamine stretches in SCA1: A molecular approach." Protein Sci **12**(5): 953-62.

- Serpell, L. C., M. Sunde, et al. (2000). "The protofilament substructure of amyloid fibrils." J Mol Biol **300**(5): 1033-9.

- Shehi, E., P. Fusi, et al. (2003). "Temperature-dependent, irreversible formation of amyloid fibrils by a soluble human ataxin-3 carrying a moderately expanded polyglutamine stretch (Q36)." Biochemistry **42**(49): 14626-32.

- Shewmaker, F., R. P. McGlinchey, et al. (2011). "Structural insights into functional and pathological amyloid." J Biol Chem **286**(19): 16533-40.

- Shtainfeld, A., T. Sheynis, et al. (2011). "Specific mutations alter fibrillation kinetics, fiber morphologies, and membrane interactions of pentapeptides derived from human calcitonin." Biochemistry **49**(25): 5299-307.

- Sipe, J. D. and A. S. Cohen (2000). "Review: history of the amyloid fibril." J Struct Biol **130**(2-3): 88-98.

- Sipe, J. D., M. D. Benson, et al. (2010). "Amyloid fibril protein nomenclature: 2010 recommendations from the nomenclature committee of the International Society of Amyloidosis." Amyloid **17**(3-4): 101-4.

- Skerget, K., A. Vilfan, et al. (2009). "The mechanism of amyloid-fibril formation by stefin B: temperature and protein concentration dependence of the rates." Proteins **74**(2): 425-36.

- Smith, A. M., T. R. Jahn, et al. (2006). "Direct observation of oligomeric species formed in the early stages of amyloid fibril formation using electrospray ionisation mass spectrometry." J Mol Biol **364**(1): 9-19.

- Smith, D. P., S. Jones, et al. (2003). "A systematic investigation into the effect of protein destabilisation on beta 2-microglobulin amyloid formation." J Mol Biol **330**(5): 943-54.

- Soldi, G., F. Bemporad, et al. (2005). "Amyloid formation of a protein in the absence of initial unfolding and destabilization of the native state." Biophys J **89**(6): 4234-44.

- Song, A. X., C. J. Zhou, et al. (2010). "Structural transformation of the tandem ubiquitin-interacting motifs in ataxin-3 and their cooperative interactions with ubiquitin chains." PLoS One **5**(10): e13202.

- Stawoska, I., A. Weselucha-Birczynska, et al. (2009). "Interaction of selected divalent metal ions with human ataxin-3 Q36." J Biol Inorg Chem **14**(8): 1175-85.

- Stefani, M. (2010). "Structural polymorphism of amyloid oligomers and fibrils underlies different fibrillization pathways: immunogenicity and cytotoxicity." Curr Protein Pept Sci **11**(5): 343-54.

- Stefani, M. and C. M. Dobson (2003). "Protein aggregation and aggregate toxicity: new insights into protein folding, misfolding diseases and biological evolution." J Mol Med (Berl) **81**(11): 678-99.

- Steffan, J. S., N. Agrawal, et al. (2004). "SUMO modification of Huntingtin and Huntington's disease pathology." Science **304**(5667): 100-4.

- Stevanin, G., G. Cancel, et al. (1995). "The gene for spinal cerebellar ataxia 3 (SCA3) is located in a region of approximately 3 cM on chromosome 14q24.3-q32.2." Am J Hum Genet **56**(1): 193-201.

- Stevanin, G., P. S. Sousa, et al. (1994). "The gene for Machado-Joseph disease maps to the same 3-cM interval as the spinal cerebellar ataxia 3 gene on chromosome 14q." Neurobiol Dis **1**(1-2): 79-82.

- Stevanin, G., Y. Trottier, et al. (1996). "Screening for proteins with polyglutamine expansions in autosomal dominant cerebellar ataxias." Hum Mol Genet **5**(12): 1887-92.

- Stromer, T. and L. C. Serpell (2005). "Structure and morphology of the Alzheimer's amyloid fibril." Microsc Res Tech **67**(3-4): 210-7.

- Sugaya, K. and S. Matsubara (2009). "Nucleation of protein aggregation kinetics as a basis for genotype-phenotype correlations in polyglutamine diseases." Mol Neurodegener **4**: 29.

- Swanson, K. A., R. S. Kang, et al. (2003). "Solution structure of Vps27 UIM-ubiquitin complex important for endosomal sorting and receptor downregulation." Embo J **22**(18): 4597-606.

- Takahashi, T., S. Katada, et al. (2010). "Polyglutamine diseases: where does toxicity come from? what is toxicity? where are we going?" J Mol Cell Biol **2**(4): 180-91.

- Tanzi, R. E. and L. Bertram (2005). "Twenty years of the Alzheimer's disease amyloid hypothesis: a genetic perspective." Cell **120**(4): 545-55.

- Thakur, A. K., M. Jayaraman, et al. (2009). "Polyglutamine disruption of the huntingtin exon 1 N terminus triggers a complex aggregation mechanism." Nat Struct Mol Biol **16**(4): 380-9.

- Todi, S. V., B. J. Winborn, et al. (2009). "Ubiquitination directly enhances activity of the deubiquitinating enzyme ataxin-3." Embo J **28**(4): 372-82.

- Tomaselli, S., V. Esposito, et al. (2006). "The alpha-to-beta conformational transition of Alzheimer's Abeta-(1-42) peptide in aqueous media is reversible: a step by step conformational analysis suggests the location of beta conformation seeding." Chembiochem **7**(2): 257-67.

- Tsai, A. M., D. A. Neumann, et al. (2000). "Molecular dynamics of solid-state lysozyme as affected by glycerol and water: a neutron scattering study." Biophys J **79**(5): 2728-32.

- Uchihara, T., H. Fujigasaki, et al. (2001). "Non-expanded polyglutamine proteins in intranuclear inclusions of hereditary ataxias--triple-labeling immunofluorescence study." Acta Neuropathol **102**(2): 149-52.

- Walsh, D. M., A. Lomakin, et al. (1997). "Amyloid beta-protein fibrillogenesis. Detection of a protofibrillar intermediate." J Biol Chem **272**(35): 22364-72.

- Wang, W., S. Nema, et al. (2010). "Protein aggregation--pathways and influencing factors." Int J Pharm **390**(2): 89-99.

- Ward, R. V., K. H. Jennings, et al. (2000). "Fractionation and characterization of oligomeric, protofibrillar and fibrillar forms of beta-amyloid peptide." Biochem J **348 Pt 1**: 137-44.

- Warrick, J. M., H. Y. Chan, et al. (1999). "Suppression of polyglutamine-mediated neurodegeneration in Drosophila by the molecular chaperone HSP70." Nat Genet **23**(4): 425-8.

- Waudby, C. A., T. P. Knowles, et al. (2010). "The interaction of alphaB-crystallin with mature alpha-synuclein amyloid fibrils inhibits their elongation." Biophys J **98**(5): 843-51.

- Wilkinson, K. D., K. H. Ventii, et al. (2005). "The ubiquitin signal: assembly, recognition and termination. Symposium on ubiquitin and signaling." EMBO Rep **6**(9): 815-20.

- Williams, A. J. and H. L. Paulson (2008). "Polyglutamine neurodegeneration: protein misfolding revisited." Trends Neurosci **31**(10): 521-8.

- Williams, A. J., T. M. Knutson, et al. (2009). "In vivo suppression of polyglutamine neurotoxicity by C-terminus of Hsp70-interacting protein (CHIP) supports an aggregation model of pathogenesis." Neurobiol Dis **33**(3): 342-53.

- Williamson, T. E., A. Vitalis, et al. (2010). "Modulation of polyglutamine conformations and dimer formation by the N-terminus of huntingtin." J Mol Biol **396**(5): 1295-309.

- Winborn, B. J., S. M. Travis, et al. (2008). "The deubiquitinating enzyme ataxin-3, a polyglutamine disease protein, edits Lys63 linkages in mixed linkage ubiquitin chains." J Biol Chem **283**(39): 26436-43.

- Xue, W. F., A. L. Hellewell, et al. (2009). "Fibril fragmentation enhances amyloid cytotoxicity." J Biol Chem **284**(49): 34272-82.

- Xue, W. F., A. L. Hellewell, et al. (2010). "Fibril fragmentation in amyloid assembly and cytotoxicity: when size matters." Prion **4**(1): 20-5.

- Xue, W. F., S. W. Homans, et al. (2008). "Systematic analysis of nucleation-dependent polymerization reveals new insights into the mechanism of amyloid self-assembly." Proc Natl Acad Sci U S A **105**(26): 8926-31.

- Yang, Y., A. P. Lyubartsev, et al. (2009). "Computer modeling reveals that modifications of the histone tail charges define salt-dependent interaction of the nucleosome core particles." Biophys J **96**(6): 2082-94.

- Zandomeneghi, G., M. R. Krebs, et al. (2004). "FTIR reveals structural differences between native beta-sheet proteins and amyloid fibrils." Protein Sci **13**(12): 3314-21.