UNIVERSITAT DE BARCELONA

Biophysical study of the aggregation of the androgen receptor protein in spinal bulbar muscular atrophy

Giulio Chiesa

BIOPHYSICAL STUDY OF THE AGGREGATION OF THE ANDROGEN RECEPTOR PROTEIN IN SPINAL BULBAR MUSCULAR ATROPHY

Giulio Chiesa

Universitat de Barcelona

Facultat de Farmàcia

Institut de RecercaBiomèdica, Barcelona

2015

UNIVERSITAT DE BARCELONA

FACULTAT DE FARMÀCIA

INSTITUT DE RECERCA BIOMÈDICA, BARCELONA

PROGRAMA DE DOCTORAT EN BIOMEDICINA

BIOPHYSICAL STUDY OF THE AGGREGATION OF THE ANDROGEN RECEPTOR PROTEIN IN SPINAL BULBAR MUSCULAR ATROPHY

Memòriapresentada per Giulio Chiesa

Per optar al títol de doctor per la Universitat de Barcelona

Director i tutor Doctorand

_____ _____

Xavier SalvatellaGiralt Giulio Chiesa

Giulio Chiesa

2015

ACKNOWLEDGEMENTS

Swashbuckling, flamboyant and reckless are three adjectives that seldom go together when referring to a scientist. The transition from these three adjectives to being a real scientist has not been smooth, but no PhD ever is. Therefore, the people that accompanied me on this journey, supporting me, challenging me, helping me and also scolding me on occasions, deserve particular attention.

My parents first: FabrizioChiesa and Paola Rampoldi and all my supportive family: Guido Chiesa, LinaRampoldi, Maurizio Rampoldi, AntonellaBisioandAlessiaRampoldi. Being away does not necessarily mean being far and I always felt your warmth from here too (which is not so far, after all).

Then there is a list of people that influenced me,in one way or another,along my journey.

First of all, I would like to thank my supervisor, Prof. Xavier Salvatella, for his guidance, insight and passion and for not stopping in believing in me when things were going off the rails. He taught me the quantitative, analytical way of thinking and about rigor and attention to details. He tried to carve out my freethinking from a bundle of noisy ideas. We had to face staggering challenges, but we got over them fairly well.

I want to thank Dr. BaharehEftekharzadeh, friend and companion on this journey, for scientific and non-scientific discussions, for reality checks, and for trying to establish order in the chaos both inside and oustide the lab. Bahareh contributed to teaching me, directly and indirectly, how to be a scientist.

I would like to thank Daniele Mungianu and SoranaIftemifor their help during the months each of them passed here. I hope I have been a decent mentor.

I would like to thank Dr. Jesus Garcia for the help with the NMR experiments and for his scientific insights.

Thanks also go to Dr. Victor Buzon for his help throughout the last part of the PhD, for his reality checks and technical tips&tricks.

My gratitude also to Joan MIquelValverde for taking care of the purifications, for jokes that made the days go by easier and for his patience in putting up with egocentric compulsive and arrogant PhD students.

I would like to thank BusraTopal for the coding and the simulations, even if, in fact, I would rather thank her for being such a good friend.

I want to thank ElaSzulc for being a great friend, at least for the first few months, then things got weird and I need a second line later on. :D

I want to thank Dr. Claudio Di Sanza, for the science, the non-science, the cups of coffee, *The Knowledge* and the *Solid Data*.

I also want to thank also former members of the lab, because each of them contributed to the story and gave it some twist: Chris for the chemistry, the scientific English supervision, the beers and thoughts about life; Anna for the good mood she brought, for being such a good friendand for the example she gave for taking tough non-canonical choices; Bryn for

his intelligence and class, Santi for his wit, Maria for giving good advice even after she left (and for the protocols!), Jordi for the coffees and the streetwise tricks, Carlos for teaching me how to survive, and Eva for being an example.

Great thanks go to the Experimental Bioinformatics Laboratory and the people I hassled every day by showing up, using instruments, leaving stuff lying around, complaining, and complaining about the complaints. In particular, Dr.MontseSoler, Dr.Isabel Heath, Nuria Villegas, and Chiara Castellazzi. Thank you for your patience and helpfulness.

I would also like to thank the staff at the Mass Spectrometry Core Facility for their professionalism, kindness and precision, in particular Dr.Marta Vilaseca, Dr.Mar VIlanova and Dr.Marina Gay.

Thanks also go to the Electron Microscopy Unit of the University of Barcelona for their friendliness, professionalism and patience in teaching me how to handle an electron microscope without asking for assistance. In particular, Dr.Carmen Lopez, Gemma Martinez, and JolandaMuela.

I finally wish to acknowledge Dr. Joan Pousat the High Throughput Crystallography Platform for giving me the opportunity to play around with the DLS and helping me whenever necessary.

There are many people that I also want to thank who have been less involved in the technical day-to-day work, but whose scientific contributions have been greatly appreciated.

First of all, Dr. Natalia Carulla, for good suggestions and scientific insight during the Advisory Committee meetings, and also whenever I felt the need for an extra opinion. I truly appreciated her help.

I also want to thank Prof.Eva EstebanezPerpiñá for her expertise in AR, scientific opinions and helpfulness in being part of my Advisory Committee for these 5 years.

I also wish to thank Prof.Eduardo Soriano for neurology-oriented insights and for participating in myAdvisory Committee.

I also want to thank Dr. BenedettaBolognesi for all the help in the first part of the PhD, the good scientific discussions and the good spirit. Benni helped me a lot during my second year of PhD, in the moments when I was a bit lost.

Finally, my thanks go to Dr.Laura Nevola, for technical help and for the tips she gave me when I didn't really know what to do.

With her, I take advantage also to thank the whole Giralt lab, which was patient enough and kind enough to have me hanging around, begging for material, instruments or suggestions every other day. The same goes for the Pons lab, thank you.

I just add one last acknowledgement to Tanya Yates, who made such a great job in spare checking and fixing the orthography for a last frantic round of corrections.

PART I: INTRODUCTION

CHAPTER 1: POLYGLUTAMINE DISEASES AND POLYGLUTAMINE PROTEINS

1.1 MISFOLDING DISEASES

The modern era has seen a rise in the importance of a family of age-related diseases that are linked to neurodegeneration. A common feature of these diseases is their association with the presence aggregates comprising one or more protein species in neurons. These aggregates form what are known as amyloid plaques(1).

Alzheimer disease (AD) and Parkinson disease (PD) are two well-studied protein misfolding diseases. AD is associated with the accumulation of the amyloid-ß protein (Aß) in brain tissue(2).Aßis the product of the proteolytic cleavage of a trans-membrane protein, the ß-amyloid precursor protein (APP)(3).Another key protein that also aggregates in the brain cells of AD patients is tau, which; however,the relationship between tau andAß is still unclear (4). In contrast, PD is associated with the misfolding and aggregation of the partially folded protein α-synuclein(5, 6).

Prion diseases, like kuru and Creutzfeld-Jakob disease, are another family of pathologies that are linked to protein aggregation. In these conditions, a protein, called prion protein (PrP^{C}), undergoes a conformational change (becoming PrP^{Sc}) that radically modifies its properties. The main characteristics of PrP^{Sc} are its propensity to form insoluble aggregates and its infectiousness, meaning that it is capable of propagating its alternative conformation to other protein molecules(7).

The last class of misfolding disorders is the family of polyglutamine diseases, associated with the anomalous genetic expansion of a polyglutamine tract present in the disease-related protein. This topic will be discussed in detail further (Chapter 1.3).

While no effective treatment is available for any of these diseases, it has been possible to isolate protein aggregates with common structural and functional features from patients. Research into the structure and biophysical properties of these aggregates and the way they are formed might help us to understand how toxicity arises.

1.2 STRUCTURAL CHARACTERIZATION OF AMYLOID FIBRILS

Protein misfolding diseases are characterized by the formation of insoluble aggregates. These aggregates tend to sequester other protein molecules, a process that leads to macromolecular assemblies called amyloid fibrils(8, 9).The definition of amyloid currently adopted by pathologists is that of an unbranchedfibrillar aggregate that occurs in tissue deposits*in vivo* and that shows green birefringence when stained with the dye Congo Red(10, 11).

A more functional definition of amyloid fibrils describes them as a type of semi-crystalline aggregate composed of protein in ß-sheet conformation, stacked along the axis of elongation, organized with translational symmetry, and held together by hydrogen bonds established between the N-H and C=O groups of the backbone of the amino-acid chains(12). This definition is based mainly on the common structural properties found in fibrils formed by six distinct proteins, each of which associated with a different disease(13). In an attempt to find a correlation between toxicity and structure, considerable research efforts have been devoted to the study of fibrils using a wide range of techniques.

Single-crystal X-ray diffraction of amyloid fibrils shows a diffraction pattern, called cross-ß, characterized by a meridian reflection at 4.8Å and another diffuse reflection at ≈10Å along the equator. This means that the distance between a ß-strand and another within the same sheet is 4.8Å, while the distance between sheets is ≈10Å(14).

Amyloid fibrils have been analyzed with more sophisticated techniques, such as solid-state nuclear magnetic resonance (NMR)(15, 16), cryo-electron microscopy (cryo-EM)(17, 18), and single-crystal X-ray diffraction(14). All of these findings point towards a high level of structural organization within the fibril spine, where strands and sheets are relatively oriented according to 8 geometries (Figure 1.1), grouped into two main categories, namely parallel ß-sheets and antiparallel ß-sheets(19). The former are formed by adjacent strands oriented in the same direction, while in the latteradjacent strandsare oriented in opposite directions. Also, every strand with the same orientation presents its side chains in register with those of the following strand, meaning that each side chain will be superimposed at a distance of 4.8Å for a parallel ß-sheet or 9.6Å for an antiparallel one(2x4.8Å)(19).

This highly packed structure is kept together by hydrogen bonds between the amides of the backbone and hydrophobic interaction between side chains belonging to adjacent strands(15, 20). The high level of organization produced and the density of interactions lead to high thermodynamic stability. Once formed, amyloid fibrils are very difficult to disassemble, and strong denaturing agents and organic solvents are generally required for this purpose. It is important to note that sonication breaks the fibrils but fails to dissolve them(21, 22).

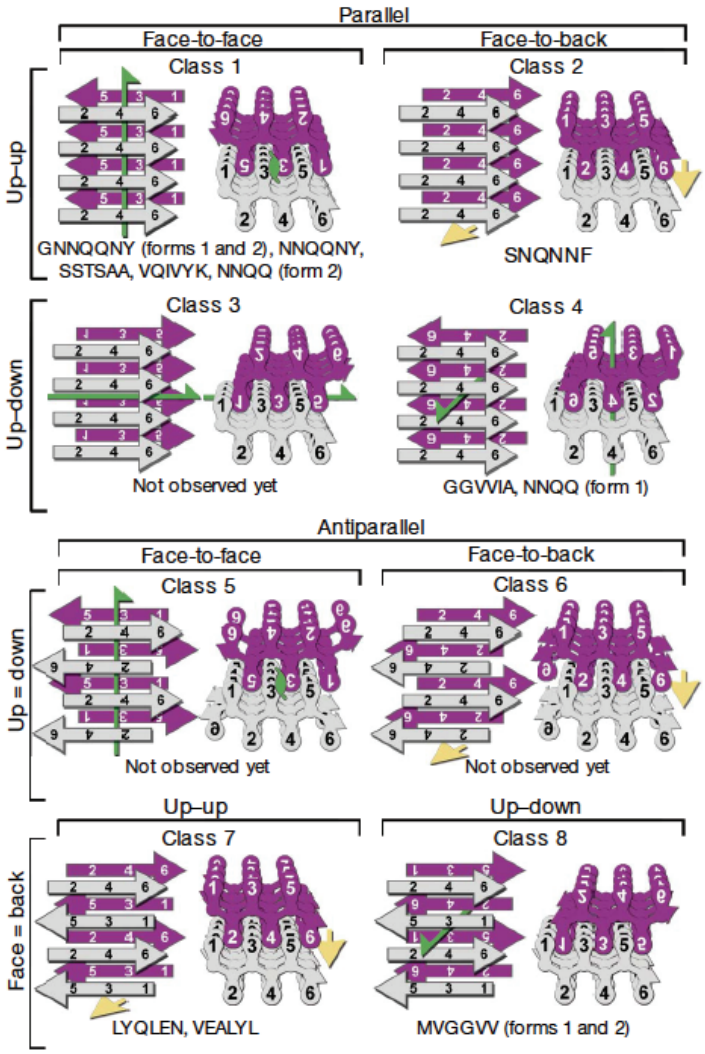


Figure 1.1.The eight classes off steric zipper. Two identical sheets can be classified by: the orientation of their faces (face-to-face vs. face-to-back), the orientation of their strands, and whether the strands within each sheet are parallel or antiparallel. Both side views (left) and top views (right) show which of the six residues of the segment point into the zipper and which point outward. Green arrows show two-fold screw axes, yellow arrows show translational symmetry. Representative protein segments are listed below each class(19)

When treated and incubated under specific conditions, many proteins and small peptides have the potential to produce amyloid fibrils(23, 24), and the fibrillar form has been considered the most stable conformation of a protein, even more stable than the native form, and it represents the global minimum in the energy landscape of many proteins, separated from the native fold by a high energy barrier(25).

It has been proposed that the amyloid fold, due to its simplicity and repeatability,was the first fold to form in the prebiotic world(26) and that other folds have evolved from it. This would explain why this potentially noxious property of proteins is still common throughout proteomes(27).

However, how this fold is acquired and how monomeric proteins assemble to build an amyloid fibril varies between protein species, and factors like the properties of the sequence(28, 29), temperature, concentration, and composition of the environment(30, 31) in which assembly occurs can play critical roles(32). Great efforts have been channeled into developing kinetic models that predict the aggregation rates of protein solutions, taking into account various physico-chemical properties, like the viscosity of the media, and sample conditions, like concentration and temperature, in order to draw general conclusions on the forces at play during aggregation reactions(33–35).

The complexity of assembly is also reflected in the difficulty to unravelhow the toxicity arises: all the molecular species produced (monomer, unfolded monomer, oligomers, fibril) can potentially be involved, but dissecting which one is actually responsible remains a challenge, as reflected by the number of contradictory findings in the literature(36–38).

1.2.1 SECONDARY STRUCTURE TRANSITIONS AND AGGREGATION

Identifying how protein molecules change their structure during the aggregation is one of the greatest challenges in the amyloid field, because of the low population of the transient species, the range of sizes produced cannot be observed by a single technique, and the structural species are highly heterogeneous.

Nevertheless, insights into the secondary structure of amyloid proteins and into the way they evolve have been made by applying a wide range of techniques. These have revealed that the critical step for the aggregation of the protein is the conversion of a disordered sequence into a ß-sheet(39). The disordered state can be caused by denaturation, like in the case of lysozyme(40, 41), transthyretin(42), and superoxide dismutase 1(43), can be intrinsic, like Aß peptide(44), or the protein can be in a partially disordered state, like ß2-microglobulin(45).

However, other proteins are described to aggregate after a double secondary structure transition. α-synuclein and islet amyloid polypeptide (IAPP), or amylin, associate as α-helical bundles and subsequently perform a slow α-to-ß transition, where the α-helix unfolds and acquires the ß-sheet conformation, forming the fibrils. Nevertheless, this interpretation has been matter of controversy, as other studies put forward the hypothesis that the α-helical conformation is actually an off-pathway intermediate that is therefore protective against the formation of fibrils(46).

1.3 POLYGLUTAMINE DISEASES

Another group of neurodegenerative disorders associated withthe incorrect folding of a specific protein species arepolyglutamine diseases, a group that include nine pathologies, the most studied of which is Huntington disease (HD)(47), and six spinocerebellar ataxias (SCA 1–3, 6, 7, 17)(48)(49), dentatorubral-pallidoluysian atrophy (DRPLA)(50), and finally spinobulbar muscular atrophy (SBMA)(51).SBMA will be described later in more detail (Chapter 2).

Except for SBMA, which is X-linked, all these hereditary diseases are dominant and gain-of-function conditions, which affect mostly neuronal cell. Each of these diseases is associated with genetic expansion on a polymorphic site located in the disease-related gene. The polymorphism involves a repetitive sequence of the codon CAG, coding for glutamine, and it is referred to as the polyglutamine (polyQ) tract.

The clinical hallmarks of this family of pathologies are the presence of aggregates of disease-related proteins inside neuronal cells, late onset of the disease, and a progression that is faster and with earlier onset for patients with longer polyQ tracts, a phenomenon called anticipation(52).It is worth noting that linearity between symptoms and polyQ tract length is not perfect. In fact, the symptoms occur only when the number of repeats is higher than a certain threshold. This threshold varies from disease to disease, as it will be discussed in more detail later in this chapter (Chapter 1.4).

The proteins of this family share little similarity in terms of sequence, cellular compartment, function, and structure(53, 54). Also, each polyQ disease affects distinct tissues and, althoughneurodegeneration is a common feature, different regions of the brain and neuron cell subtype can be affected.

In all of these diseases, the only elements in common are the presence of aggregates of proteins with an expanded polyQ tract and the development of these conditions. It is therefore possible that similar mechanisms of toxicity underlie all polyQ diseases and that these conditions are closely related to the formation of aggregates. Studying how these aggregates are formed is therefore of primary importance for understanding such diseases.

1.4 POLYGLUTAMINE EXPANSIONS AT THE DNA LEVEL

Highly repetitive sequences are present across the entire genome and are generally associated with polymorphic sites, meaning that the number of repeats can expand and contract from replication to replication. So far, more than twenty hereditary diseases have been identified to be associated with a genetic expansion, of which nine (the polyQ diseases) present this expansion in a coding region of DNA: a polyCAG sequence(55).

As a general observation, these mutations are almost always dominant and determine a gain-of-function, generally associated with a neurotoxic effect. The genetic expansions in non-coding regions are better tolerated than those in coding ones(56) and they can span between 55 and 200 units before any pathologic phenotype develops. For somatic trinucleotide repeats (TRNs), the threshold is much lower, fluctuating between 30 and 40 units, depending on the protein associated.

This genetic instability is generally related to the propensity of the highly repetitive sequence of forming non-classical DNA structures(57). Sequences like $(CNG)_n$ or $(CTTG)_n$ can form hairpins, $(GAA)_n$ repeats form triplexes, and $(CGG)_n$ and $(C_4GC_4GCG)_n$ repeats form G-quadruplexes(56). A hairpin formed by a CAG/CTG repeat (Figure 1.3) determines the slippage of the leading strand, which leads to a small-scale expansion(58, 59).Larger scale

expansion may be associated with the sum of recurrent small-scale expansions or with other mechanisms, which involve Base Excision Repair (BER) or Nucleotide Excision Repair (NER)(60).

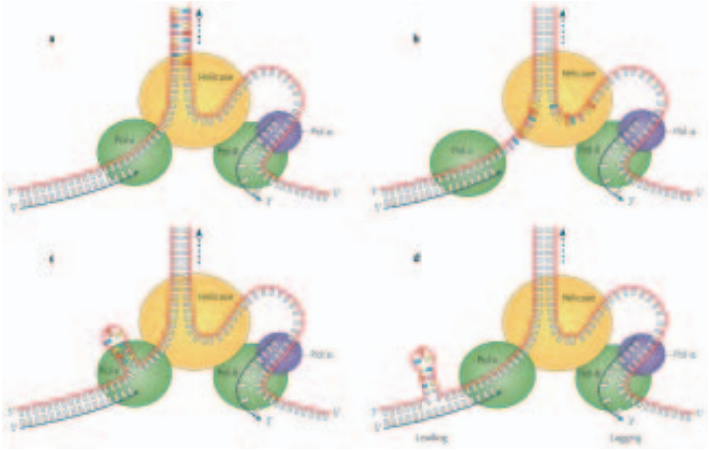


Figure 1.2 Slippage model for a change in trinucleotide repeat length(60)

As shown in Table 1, these polymorphic regions can expand and contract over generations, but,beyond a certain number of repeats, the sequence becomes unstable and extra copies are incorporated in each generation. The threshold for instability differs from that for the onset of disease and are generally not strictly related, but the distance between the two can also be a single trinucleotide, like in SCA6 (see table 1.1)..

Disease	Sequence	Location	Parent of origin of expansion	Repeat number (normal)	Repeat number (pre-mutation)	Repeat number (disease)	Somatic instability
coding TRNs							
DRPLA	CAG	ATN1 (exon 5)	paternal	6-35	35-48	49-88	Yes
HD	CAG	HTT (exon 1)	paternal	6-29	29-37	38-180	Yes
OPMD	GCN	PABPN1 (exon 1)	paternal and maternal	10	12-37	>11	Not found in tissue (hypothalamus)
SCA1	CAG	ATXN1 (exon 8)	paternal	6-39	40	41-83	Yes
SCA2	CAG	ATXN2 (exon 1)	paternal	<31	31-32	32-200	Unknown
SCA3 (Machado-Joseph disease)	CAG	ATXN1 (exon8)	paternal	12-40	41-85	52-86	Unknown
SCA6	CAG	CACNA1N (exon 47)	paternal	<18	19	20-33	None found
SCA7	CAG	ATXN7 (exon 3)	paternal	4-17	28-33	>36 to >460	Yes
SCA17	CAG	TBP (exon 3)	paternal and maternal	25-42	43-48	45-66	Yes
SBMA	CAG	AR (exon 1)	paternal	13-31	32-39	40	None found
non-coding TNRs							
DM1	CTG	DMPK (3'UTR)	maternal	5-37	37-50	z50	Yes
DM2	CCTG	CNBP (intron 1)	uncertain	<30	31-74	75-11000	Yes
FRX-E	GCC	AFF2 (5'UTR)	maternal	4-39	40-200	<200	Unknown
FRDA	GAA	FXN (intron 1)	recessive	5-30	31-100	70-1000	Yes
FXS	CGG	FMR1 (5' UTR)	maternal	6-50	55-200	200-4000	Yes
HDLZ	CTG	JPH3 (exon ZA)	maternal	6-27	29-35	36-57	Unknown
SCA8	CTG	ATXN8OS (3' UTR)	maternal	15-34	34-89	89-250	Unknown
SCA10	ATTCT	ATXN10 (intron 9)	maternal and paternal*	10-29	29-400	400-4500	Yes
SCA12	CAG	PPP2R2B (5' UTR)	maternal and paternal**	7-28	28-66	66-78	None found

*Table 1.1.List of all the TNR expansion diseases. AFF2, AF4/FMR2 family, member 2; AR, androgen receptor; ATN1, atrophin 1; ATXN, ataxin; ATXN8OS, ATXN8 opposite strand (non-protein coding); CACNA1A, calcium channel, voltage-dependent, P/Q type, alpha 1A subunit; CNBP, CCHC-type zinc finger nucleic acid binding protein; DM, myotonic dystrophy; DMPK, dystrophiamyotonica-protein kinase; DRPLA, dentatorubral-pallidoluysian atrophy; FMR1, fragile X mental retardation 1; FRAX-E, mental retardation, X-linked, associated with FRAXE; FRDA, Friedreich ataxia; FXN, frataxin; FXS, fragile X syndrome; FXTAS, fragile X-associated tremor/ataxia syndrome; HD, Huntington disease; HDL2, Huntington disease-like 2; HTT, huntingtin; JPH3, junctophilin 3; OPMD, oculopharyngeal muscular dystrophy; P, paternal; PABPN1, poly(A) binding protein nuclear 1; PPP2R2B, protein phosphatase 2, regulatory subunit B; SCA, spinocerebellar ataxia; SMBA, spinomuscular bulbar atrophy; TBP, TATA-box binding protein; TNR, trinucleotide repeat.; *smaller changes when maternal; **smaller changes when paternal.*(60)

Another phenomenon associated with the instability of these sequences is genetic anticipation, which consists ofan earlier onset of the symptoms, together with a progressively more severe disease phenotype at every generation from the first one affected(52, 61).

It has been proposed that the expansion mutation occurs during meiosis, while spermatogonia divide, or after meiosis in differentiated germ cells(62). In a mouse model of Huntington disease (HD), the polyQ expansion increases over the lifetime of the animal(63), a finding that was confirmed by screening patients over their lifetimes(64) and analysis of postmortem brain samples(65). However, it is not clear whether this further expansionis a general phenomenon as there is a lack of similar information on other diseases.

In spite of the widely reported relationship between a long polyQ chain (or a long polyCAG sequence) and the onset of these diseases, the causes of toxicity are yet to be determined. A considerable body of evidence was collected supporting the association of toxicity with a gain-of-function of the RNA transcript to undergo a not entirely well-understood conformational change, which sequesters proteins involved in the translation and splicing and in the production of small silencing RNA(55, 66–68). The main research line regarding toxicity of polyglutamine proteins addresses how they aggregate, how the aggregates interact with the cell environment, and how this interaction can lead to cytotoxicity, as will be discussed in the further sections.

1.5 BIOPHYSISCS AND STRUCTURE OF THE POLYGLUTAMINE PEPTIDE

To gain a thorough understanding of the aggregation ofpolyglutamine proteins, biophysicists have channeled great effort into dissecting this process, starting from reductionist approaches with polyglutamine peptides(69).

Glutamine is a polar residue; therefore it is counter-intuitive to think it would form amyloid fibrils, which are generally associated with hydrophobic interactions and hydrogen bonding. However, Max Perutz described the polyQ chain in an aqueous buffer as a polymer in a poor solvent (70), or not optimal for its solvation. The solvent water is excluded from the collapsed polymer chain, so that the polyQ chain is not as extended as a classical random coil, but neither establishes long-range interactions or forms a specific structure. This degree of compaction of the polyQ chain does not correlate with the number of repeats, as observed by fluorescence correlation spectroscopy (FCS)(71), fluorescence resonance energy

transfer (FRET)(72), and small-angle X-ray scattering (SAXS)(73). Another interesting feature is the very high barrier to unfolding by extension in atomic force stretching experiments(74). This behavior was explained by many simulations as the monomeric polyQ establishing preferentially hydrogen bonds between amides of the side-chains and main-chains, instead of water(75–77).

A monomeric polyQ chain is a random coil, andthere is little difference in secondary structure between long and short polyQ chains, as polyQ peptides of different sizes observed by circular dichroism and NMR are reported as random coil(78, 79). Nevertheless, simulations isolated other low populations of secondary structure, such as the α-helix, ß-turn and PPII-helix(75, 76, 80), which are supported by little experimental data to date.Recent Fourier transform infrared spectroscopy (FTIR) data show minor populations of polyQ peptides in ß-sheet conformation in a concentration- and repeatlength-dependent fashion(81).

Further solution NMR studies were carried on a polyQ tract fused with glutathione-S-transferaseto increase its solubility(82) and on the polyQ tract of Ataxin-3(83). In both cases,heteronuclearsinglequantum coherence (HSQC) experiments identified the polyQ tract as a cluster of collapsed peaks, a feature typical of resonances in disordered regions. Solid-state NMR experiments explored the structure of polyQ tracts in the fibril state, confirming its ß-sheet nature but also describing an effect of the polyproline region at the C-terminus of the polyQ peptide, which occurs exclusively in the monomeric form of the peptide. The polyproline locks the adjacent polyQ tract in a PPII helix conformation, preventing the formation of the ß-sheet; however, once the fibril is formed, none of the glutamines adopt a PPII conformation(84).

The Wetzel group performed an extensive study on the kinetics of aggregation of polyQ peptides flanked by pairs of lysines (KKQ_nKK). Applying charged residues as flanking motifs to the polyQ peptide was required for solubility in aqueous solution(85).

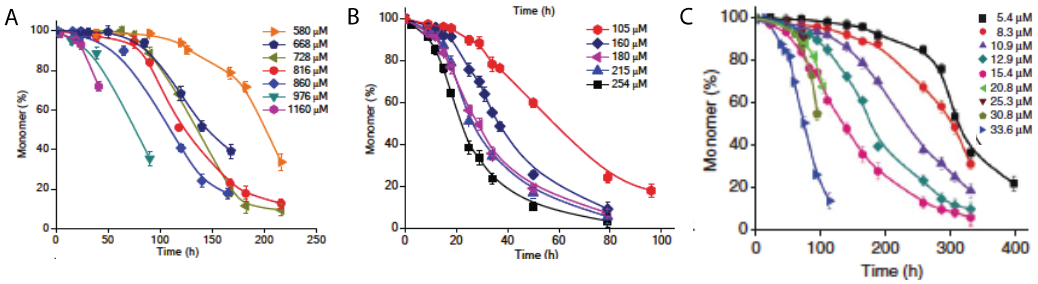


Figure 1.3 Sedimentation assay of $KKQ_{18}KK$ (a), $KKQ_{27}KK$ (b) and $KK_{37}KK$ (c) at several concentrations(86)

In vitro studies on polyQ peptides characterized their tendency to aggregate both in a repeat length- and concentration-dependent fashion, determining that the threshold after which the protein becomes toxic does not arise from different biophysical properties in the peptides of pathogenic length(86, 87). Figure 1.3 shows also that the aggregation kinetics of these peptides follows an evolution of hundreds of hours, suggesting that it is necessary to overcome a high-energy barrier for the formation of a nucleus of aggregation to occur.

The kinetics of polyQ peptideaggregation follows a classical nucleated growth polymerization mechanism (Figure 1.3)(78), which consists of a lag phase when the high kinetic barrier needs to be overcome, followed by a relatively fast elongation, which can be monitored by the disappearance of the monomer in solution after ultracentrifugation, shown in Figure 1.3.

The classical nucleated growth polymerization model has been adapted by the Wetzel group to take into account a marked polyQ dependenceonthe critical nucleus size.

They consider the critical nucleus size (n*) as the number of monomers composing the first stable oligomer, where the equilibrium between elongation and dissociation is shifted towards the elongation and the oligomer builds up to form fibrils (Figure 1.4A.)They calculated that n* decreases rapidly from n*≈4 for 24 repeats to n*≈1 for repeats equal or longer than 26. In this model, the nucleating event, if it is not a protein-protein interaction, requires a change of conformation of the monomer which entails a very high entropic cost, thus explaining the lag phase(88).

Therefore, the model proposed by the Wetzel group involves a change of conformation in the monomeric peptide from random coil to ß-turn, thus allowing the recruitment of other monomers to form the nucleus of the fibril(89) (Figure 1.5B).

This relatively simple behavior would suggest a linear progression towards a toxic phenotype correlated with the number of glutamines, but it does not account for the sudden appearance of toxicity when the protein contains a number of repeats higher than a given threshold nor does it explain why this threshold is different for each protein. In this regard, several studies have

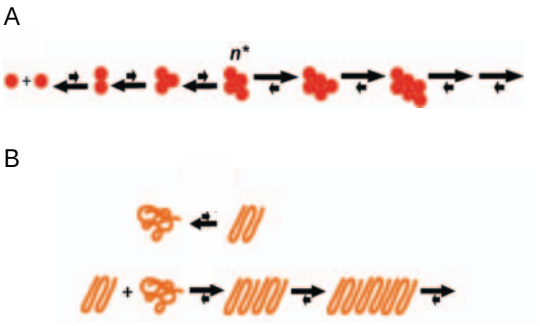


Figure 1.4. (A) Model for minimal nucleus size n; (B) model for the nucleation event with n*≈1. A random coil polyQ peptides require a ß conformation to nucleate*(69) .

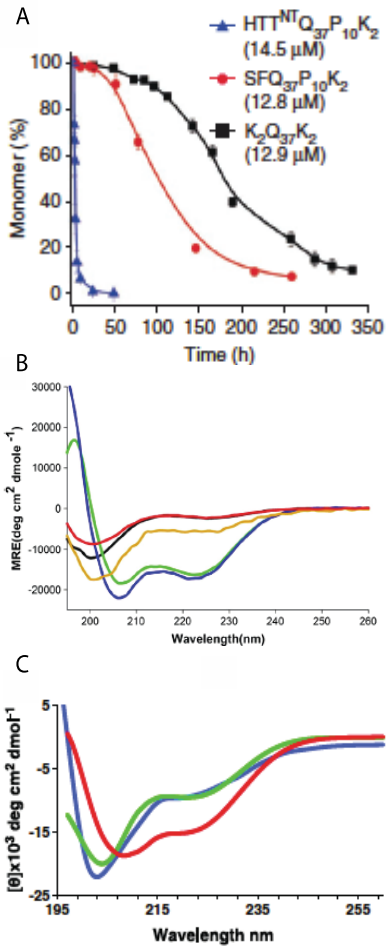


Figure 1.5. (A) sedimentation assay of htt$^{NT}Q_{37}P_{10}$ (blue), a N-terminal truncated version of the same peptide (red), and a $K_2Q_{37}K_2$ peptide (black); (B) htt^{NT} peptides at different concentrations: 1.3mM (green), 0.67mM (blue), 0.33mM (orange), 0.15mM (black), 0.07mM (red)(86)

focusedonsequence contour of the polyQ tract or the flanking regions. This topic is treated extensively in the following sections.

1.6 A WELL-ESTABLISHED MODEL FOR POLYQ: HUNTINGTIN

Most of the studies on polyQ proteins are carried out on huntingtin (Htt), which is the 350 KDa protein associated with HD. In most of these studies, a peptide corresponding to the first 30-50 residues of exon 1 of this proteinwas used as a model. This peptide is composed of a N-terminal flanking region of 17 aminoacids (htt^{NT}), the polyQ tract, and a C-terminal polyproline sequence of 10 residues, and it is thought to be the toxic element for HD(90). A normal polyQ tract for Htt is between 6 and 37 repeats, while the toxic protein contains 38 repeats or more. The structure and function of these two flanking regions have been well characterized(69, 91). These studies highlighted that htt^{NT}plays a crucial role in nucleating the aggregation of polyQ peptides and in increasing the rate of aggregation significantly (86), while the polyproline region (P_{10}) at the C-terminus hampers the self-association of polyQ peptides, as its removal results in a dramatic increase in aggregation rate and toxicity(92, 93).

A crystallographic study of the Htt peptide fused to MBP revealed that the htt^{NT} is in α-helical conformation, while the polyproline region (P_{10}) at the C-terminus adopts a PPII helix conformation. In the crystal, the short polyQ tract adopts a series of conformations, among them the α-helix (94). It is possible that the strong packing of the crystal lattice forces the helical conformation of htt^{NT}, as CD data are not completely in agreement with this finding.

A htt^{NT} peptide with a small polyQ tract in solution has been observed as a random coil, with some degree of helicity(95, 96). In this study, the helical content increased with concentration (Figure 1.6B), but also with time of incubation (Figure 1.6C). This same study describes that immediately after strong disaggregation, a solution of peptides containing htt^{NT} with a short polyQ tract is composed mainly by monomers, but with the presence of low populations of tetramers and higher molecular weight species (Figure 1.7)(97).

It is remarkable to note that placing htt^{NT}at the N-terminal or C-terminal of the polyQ tract does not change its effect, while, for the the P_{10} motif, the positionis important for its function(95). These observationssuggest that these two flanking regions have different mechanisms of action: the polyproline region would lock the polyQ in a conformation unfavorable to a transition to the ß-turn thus preventing aggregation, while the htt^{NT}, with its high self-affinity, would dramatically increase the local concentration of the polyQ chain, encouraging the nucleation event.

The Wetzel group proposed another model of a multi-step oligomerization mechanism (Figure 1.7), where a monomer adopts two conformations of htt^{NT}: the random coil conformation is in equilibrium with the α-helical, but the α-helical monomer has high affinity to oligomerize and can form transient oligomeric species (they identify the first one as a tetramer(97)) that are stabilized by coil-coils interactions. Once this α-helical bundle is formed, it can recruit more monomer, until the local concentration of the polyQ chain is

high enough to overcome the energy barrier and reorganize as ß-strand. This irreversible event determines the formation of the fibrils and their subsequent elongation.

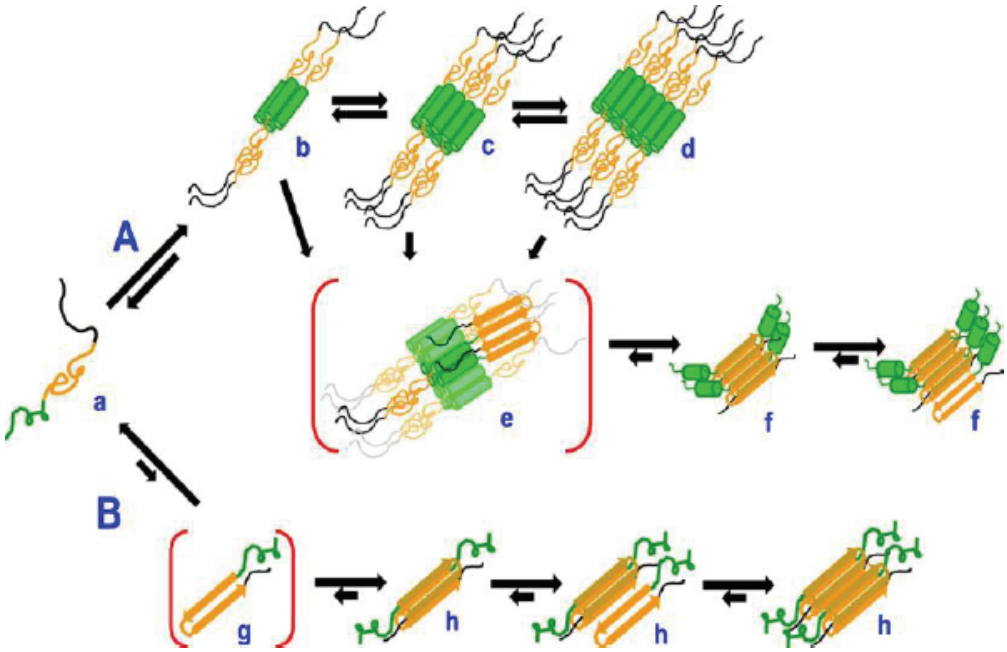


Figure 1.6 Multistep oligomerization mechanism of the exon1 polyQ peptide of Htt. A is the reaction constant between a, the random coil monomer, and b, the helical bundle. B is the reaction constant between a and g, the highly unfavorable monomer in ß-turn conformation. Species b, c and d are in equilibrium, but when the oligomer e is formed, the high local concentration allows the high energy barrier to be overcome, thus allowing ß-sheet formation, and the equilibrium is shifted towards the elongation of the fibril(349).

1.7 THE ROLE OF THE FLANKING REGIONS

It is clear that flanking regions are critical in the nucleation process for huntingtin, but similar scenarios have also been described for other systems. Ataxin-3, associated with another polyQ expansion disease, contains a highly aggregation-prone domain (the Josephin domain) that can form worm-like fibrils that are not resistant to SDS. Recombinant proteins composed by the Josephin domain linked to short polyQ tracts aggregate with similar rates, while longer polyQ tracts determine a higher rate of aggregation of these proteins(98). Furthermore, the SDS-soluble aggregates formed by the Josephin domain are observed to evolve to a more stable SDS-resistant fibril when the ataxin-3 protein carries a polyQ tract (99). Experiments performed on a model system demonstrated the importance of the flanking regions, by associating several different N- and C-terminal flanking regions to polyQ tracts of different lengths and controlling their aggregation rates(100). This study revealed that the insertion of a folded domain at C-terminal of the polyQ tract prevents the aggregation of the protein, whereas polyQ tracts placed towards the N-terminal or C-terminal endsof the polypeptide (also as a consequence of proteolytic cleavage) result ina higher propensity to aggregate. In addition, the presence of a long polyQ tract within two folded domains of a protein can destabilize the folding of the host protein and trigger its aggregation(101, 102).

The Hendrikson group performed seminal work in identifying common patterns of the flanking regions(103). Using a bioinformatics approach, they found that Q/N-rich proteins (of which polyQ proteins are a subgroup) are particularly enriched in coil-coil domains or motifs, and these motifs are generally in proximity to Q/N-rich regions (such as polyQ). They then validated this finding through biophysical techniques and mutagenesis. Furthermore, an analysis of the sequence biases conserved throughout eukaryotes highlighted an over-representation of aminoacids like Leu, Pro and His in the regions flanking the polyQ tract, while Gly, Asp and Cysare under-represented(104). This could be explained by the propensity for the former aminoacids to form secondary structures (many Pro residues may force PPII helices, while many Leu residues α-helices), while Gly, Asp and Cyspromote ß-turns or introduce higher degrees of freedom, favoring the random coil structure. All together, these data point tothe relevance of the coil-coil interaction for the oligomerization of polyQ proteins, which can be read both as a mechanism by which the flanking regions act in promoting the oligomerization in all the polyQ diseases, and may explain the physiological functions of polyQtracts.

1.8 THE PHYSIOLOGICAL ROLE OF THE POLYQ TRACT

Given the instability of the DNA sequence and the propensity of proteins that carry a polyQ tract to aggregate and form fibrils,why are these regions not disfavored by evolution? Whilethe answer remains unclear, for the polyQ tractto have been conserved throughout eukaryotes(105) and to be widespread in the human genome(106), this tractclearly must have a physiologic function. The observation that the toxicity of polyQ diseases is often caused by the sequesteringfrom the cell environment of important proteins that interact with the host protein(107, 108)and its association with coil-coil prone regions point towards a critical function of the polyQ tract in strengthening or establishing protein-protein interactions. This hypothesis is also supported by the relationship found between the number of glutamines in the polyQ tract of androgen receptor (AR) and the insurgence of prostate cancer, where a low number of repeats is considered a risk factor for contracting this type of cancer(109), while a high number of repeats is associated with spinal bulbar muscular atrophy (SBMA), as will be discussed more in detail in the next chapter.

A possible explanation is that the polyQ tract extends a coil-coil interaction on a broader surface, becoming part of the coil-coil itself. In this relationship, the length of the polyQ tract determines the strength of the interaction, while the flanking regions provide the specificity. This delicate system is also highly aggregation-prone, but it is regulated by the cell clearance machinery, and the fast aggregation of proteins with polyQ tracts longer than a certain threshold might overload it, thus causing toxicity(105).

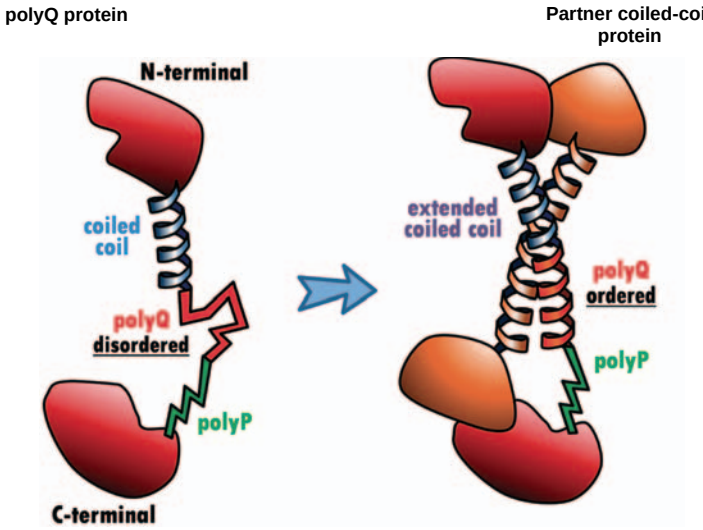


Figure 1.7 Model of physiological function for a polyQ protein. The polyQ acquires a helical conformation and strengthens the coil-coil with the partner. (105)

1.9 PROTEOSTASIS AND PROTEIN AGGREGATION

Native proteins in their physiological environment are metastable(25) and are kept folded and functional by sophisticated housekeeping processes in the cell. This active process is called protein homeostasis, or proteostasis, and it is the combination of genetic regulation and cellclearance. The genetic regulation involves a set of genes that controls the production of proteins, switching their expression on and off. Cellclearance is obtained through a network of proteins that maintain client proteins in solution and active, or commit them to degradation when aggregated, erroneously expressed or misfolded(110).

The relationship between proteostasis and misfolding diseases has been extensively studied using animal models of polyQ diseases, as they are age-related and can be controlled by tuning the severity of the effects with the number of glutamines expressed (111).The age-dependence has been demonstrated to be widespread in every animal model studied, but the relative time of onset and severity with respect to the number of glutamine repeats are organism-specific(112, 113).

The proteostasis network comprises three main classes of proteins, which perform three different tasks respectively: chaperonins, molecular chaperones, and ubiquitin-proteasome system (UPS).

Chaperonins are multiprotein complexes organized as a ring with a central cavity that can host a single polypeptide chain. They assist protein folding by an ATP-mediated process, moving the polypeptide through the pore and allowing it fold upon exit(114). Molecular chaperones prevent protein aggregation by binding to exposed hydrophobic patches, refolding aberrantly folded proteins, and actively dissociating protein aggregates. Among molecular chaperones, heat shock proteins (Hsp) are the most widely studied(115).

Small heat shock proteins (sHsps) bind to misfolded proteins by hydrophobic interactions and cooperate with the Hsp70 family of chaperones to actively refold through an ATP-

dependent process. Another complementary protein called nucleotide exchange factor (NEF) restores ATP and the system(116). Hsp70 can also form a complex with Hsp90 and a set of cofactors to actively promote the folding of proteins involved in signal transduction(117) and assists the targeting of aggregates to the lysosome, by recruiting p62 ubiquitin adaptor, through the cofactor Bag-3(118). All of these elements together cooperate to maintain proteins folded, to prevent their aggregation, and to actively disassemble aggregates(119)

The UPS is a broad network that comprises more than 600 proteins in the human proteasome, both present in the cytosol and the nucleus. Its role, in parallel with autophagy, is to clear the cell by degrading proteins that are in excess or terminally misfolded(120, 121). The UPS and molecular chaperone machinery work in concert, as the ubiquitin ligase CHIP binds to Hsp70 and Hsp90 and proteins in a non-native fold are more probable to spend time in the Hsp-bound state and being recognized by CHIP(122) or being targeted to the lysozyme through ubiquitin-independent pathways associated with Hsp70(123).

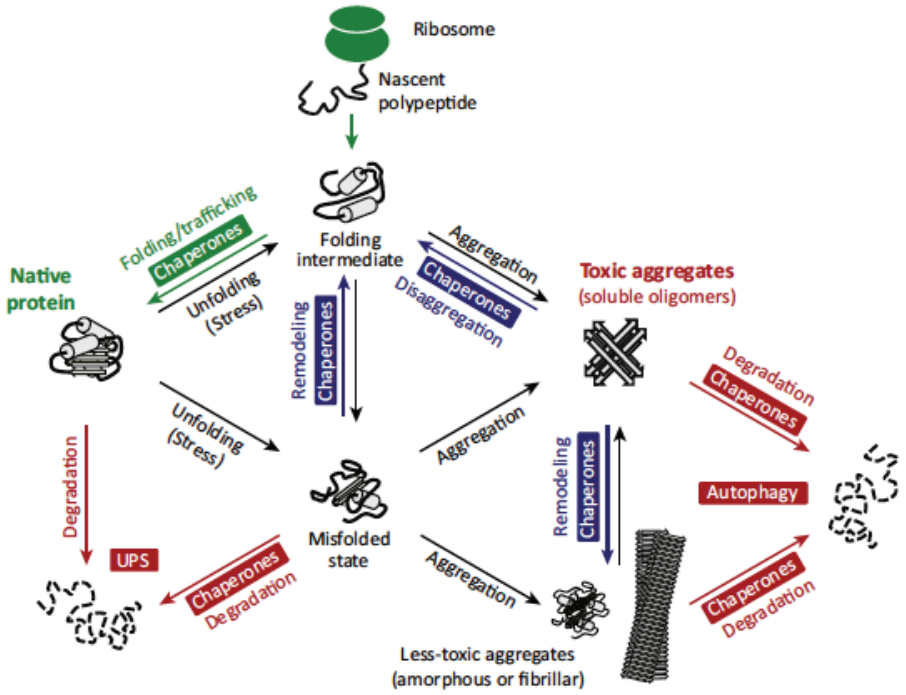


Figure 1.8 Representation of the proteostasis network. In green the processes associated to synthesis and trafficking of new proteins. In blue the remodeling and refolding of misfolded proteins and the disassembly of aggregates mediated by molecular chaperones. In red, degradation processes through UPS or autophagy. Toxic aggregates may be converted to less toxic, more stable species(110).

Therefore, proteostasis is maintained robust network with high redundancy, which regulates the accumulation of aberrant protein species and aggregates. However, there is increasing evidence of an association between ageing and a decrease in efficiency of the proteostasis(124, 125). It has been demonstrated that Hsp70 is up-regulated less efficiently in senescent fibroblast cultures taken from various animal species(126–128). It is possible that, with a decreased capacity of the proteostasis network, aggregated species are

sequestered by the proteasome, but cannot be degraded and consequently accumulate as ubiquitylated species, stalling the system and preventing other proteins from being targeted for degradation(129, 130).

These findings point to new opportunities for pharmacological intervention, by blocking or delaying the aggregation of misfolded species and by extending the capacity of the proteostasis network to cope with aberrant folding and accumulation of proteins. Approaches in this regard have attempted with small-molecules, aimed to activate cytosolic stress response pathways, such as overexpression of molecular chaperones(131) or induction of autophagy(132), and by increasing the ability of the UPS to dispose of proteins(133).

CHAPTER 2: SPINAL BULBAR MUSCULAR ATROPHY OR KENNEDY DISEASE

2.1 INTRODUCTION

Among polyglutamine diseases, spinal bulbar muscular atrophy (SBMA, or Kennedy disease) was the first to be identified as associated with the expansion of a polyQ tract(134, 135). The disease has late onset and is characterized by dysarthria, dysphagia, wasting and fasciculation of the tongue, weakness of the proximal muscles, and absence of tendon reflexes. However, life expectancy is not reduced(51, 136).

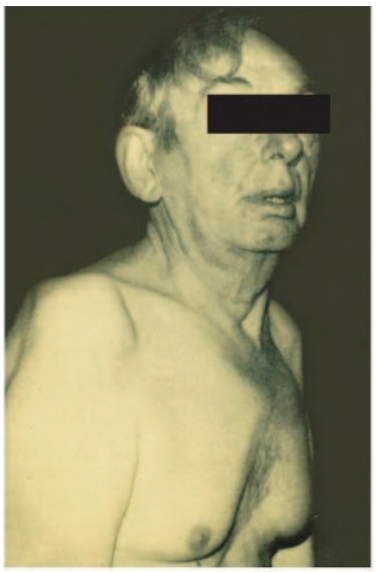


Fig 2.1 Patient with SBMA. Note the atrophy of the proximal upper limb musculature and the presence of gynecomastia(138)

SBMA is a hereditary disease that occurs in 1 in every 50000 males, with X-linked heritability(51, 137) in Caucasian and Asian populations, with no cases observed in African and Aboriginal populations(138) and a slight preponderance in the Japanese population. However, this higher rate of SBMA in the Japanese population is attributed to a founder effect, instead of a higher propensity to develop genetic expansion(139).

SBMA is also characterized by androgen insensitivity, which can occur independently of the neuro-muscular degeneration and is manifested as breast enlargement, reduced fertility, and testicular atrophy(51). Furthermore, the penetrance of this disease is strongly male-related, with milder and later symptoms in heterozygous women(140) and a negative correlation between levels of testosterone in patients and progress of the disease(136).

The pathophysiology of SBMA strongly resembles that of amyotrophic lateral sclerosis (ALS). Both diseases lead to degeneration of muscle tissue and motor neurons and this similarity has led to misdiagnosis of ALS in SBMA patients. In many cases of SBMA, the symptoms associated with androgen insensitivity were present before the onset of neurodegeneration(141–143) and therefore could not be used to differentiate this disease from ALS.Indeed, the only feature allowing the discrimination of these two diseases is the rate of progress, with ALS evolving faster and leading to death. The increase in genetic testing prevents misdiagnosis(144).

From the analysis of the symptoms and the heritability of the disease, early research into SBMA focused on studying mutations on the AR gene that could cause the disease. The polymorphic polyCAG tract in exon 1 immediately attracted the attention of the scientific community. Observations of clinical cases disclosed that SBMA patients harbor 37 to 66 repeats in this region of DNA, while healthy individuals have 13 to 34 repeats, thereby indicating a strong correlation between the number of repeats and onset of the disease(51, 135, 145).

Biopsies of SBMA patients reveal degeneration of both muscle cells and motor neurons (146), a process associated with the presence of nuclear inclusions (NIs), or nuclear aggregates, which are positively stained by antibodies against AR(147, 148).NIs found in muscles and the spinal cords of patients contain AR with an expanded polyQ tract. These aggregates are SDS-insoluble and can be dissolved by formic acid(147, 149). Aggregates of AR are also spread between nucleus and cytoplasm in cells from other tissues; however, only NIs correlate well with toxicity(150).

2.2 MODEL SYSTEMS OF SBMA

SBMA has been studied using several model systems. Notably, the Sobue group developed mouse models expressing an AR gene with polyQs of more than 90 residues(145). In this model, as for others(151, 152), transgenic AR is expressed in the presence of endogenous AR, therefore the mice do not show particular symptoms of androgen insensitivity.

The effect of testosterone levels on the severity of the disease was studied in a SBMA mouse model, both by competition with androgen antagonists(153, 154) and by increasing the testosterone levels in the organism(155). The action of antagonists like flutamide(154), as well as surgical castration(151), partially rescue the SBMA phenotype in mice. Flutamide is an AR antagonist extensively used for prostate cancer treatment in the current pharmacopeia(156, 157).

Conversely, testosterone levels higher than physiological levels do not lead to an increase in the severity of the disease, meaning that AR is saturated with testosterone in concentrations that are close to the physiologic levels for this ligand (158). Another mouse study outlined that the onset of cytoplasmic oligomers of AR protein can be found in the tissues before the manifestation of the symptoms, thus supporting the hypothesis that the aggregation starts in the cytoplasm(152).

Another valuable animal model usedin this line of research is *Drosophila melanogaster*, a fly extensively used to study polyQ diseases(159–161), as the target gene can be expressed in specific tissues and it is relatively quick and simple to obtain several fly lines for screening mutants or conditions. A seminal work demonstrated that the full-length version of AR is necessary to produce the toxic phenotype in *Drosophila*and confirmed that AR binding to testosterone is necessary for triggering toxicity in flies expressing the protein in different tissues(162). That study also described that both the nuclear translocation of AR and its binding to the DNA are necessary conditions but not sufficient for the onset of toxicity. They finally put forward the hypothesis that AR with an expanded polyQ establishes a different set of interactions than physiologic AR. This new network of interactions might be the cause of toxicity (162).

A number of cell lines can be used to study how the cell environment changes in response to the expression of polyQ AR. An immortalized motor neuron cell line (NSC34) transfected with polyQ AR show lower survival and higher caspase-3 (Cas-3) activation than controls (149). The same cell line was used to study the perturbation in the expression of polyQ AR when induced by testosterone and the effects on its own transcript(163).

Cells that lack AR, like PC12 cells, were transfected with wild-type (wt) and polyQ AR to study the relationship of this protein with the proteasome(164, 165).Stable transfection of these cells with AR under an inducible promoter was used to demonstrate that cell death

and the formation of aggregates occur only after inducing the expression of AR and only in the presence of testosterone(166).

Muscle degeneration was studied by producing stable cell lines expressing normal and polyQ AR from C2C12 cellsand a stable cell line was developed from muscle tissues of SBMA patients, proving that cytotoxicity is not associated with only motor neurons, but also with muscle tissue (167, 168).

More recently, SBMA-induced pluripotent stem cells (IPSCs), developed from patient tissues, showed no interference in the differentiation into the motor neuron lineage(169). In addition, a line of mesenchymal stem cells (MSCs) was derived from adipose tissue of SBMA patients. MSCs offer the advantage of not requiring viral infections to establish the cell line and are increasingly preferred to IPSCs (170). That study reported defects in differentiation to adipocytes and reduced cell viability in SBMA MSCs, thereby pointing to defects in tissues other than spinal cord and muscle contributing to the SBMA phenotype.

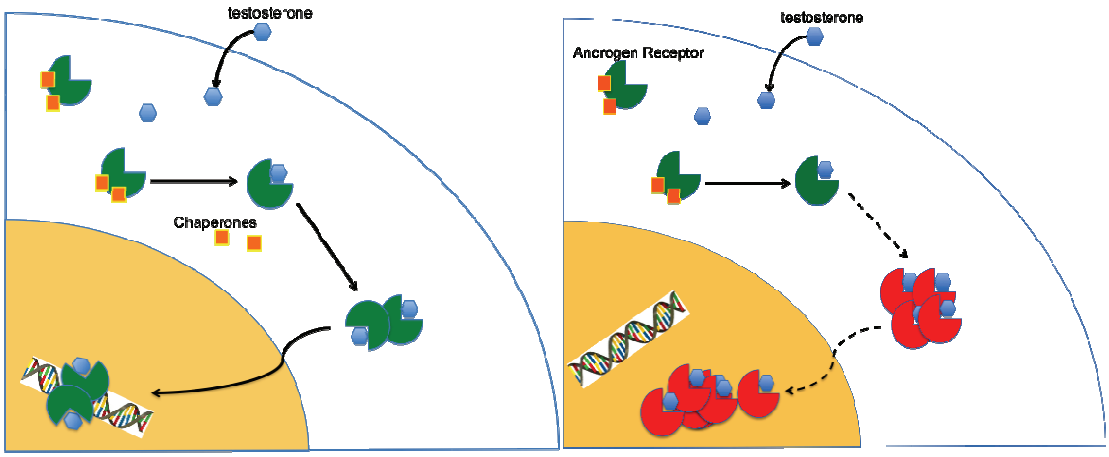


Figure 2.2 Physiological (green) and pathogenic (red) function of androgen receptor. The protein in the cytoplasm is a monomer in complex with molecular chaperones. The binding of testosterone determines the dissociation of the chaperones and triggers the oligomerization of polyQAR, which leads to the formation of aggregates in the nucleus by a mechanism as yet not understood.

2.3 NUCLEAR INCLUSIONS AND MECHANISMS OF TOXICITY

Immunohistochemistry of SBMA patients using AR antibodies revealed the presence of AR in NIs (figure 2.3), particularly in anterior horn neurons and in myocytes. The abundance of inclusions correlates well with the length of the polyQ tract (147, 148). Together with NIs, other aggregates were found in a wide range of tissues, mostly in the cytoplasm or the Golgi apparatus. Indeed, some studieshave put forward the hypothesis that not only NIs areinvolved in the development of the pathology (150).

Nevertheless, the observation of NIs correlates with cell death in cultured cells(171), and the amount of polyQ aggregates in animals correlates well with the progression of the disease(145). Moreover, experiments with HD mouse models demonstrate that suppressing the expression of the transgene of a conditional HD mouse mutant determines the clearance of NIs, reverting the severity and followed by recovery from the disease(172). For all of these

reasons, it is commonly believed that NIs are tightly associated with toxicity. However, it remains unclear whether the toxicity is attributable to NIs or to some other intermediate species on-pathway for their formation.

Some studies report an oligomeric species associated with the toxicity in cells (173) or describe a truncated version of the polyQ AR as independently activating an apoptotic cascade in a neuronal cell line(174). Also, others propose that the aggregates in fact have a protective function, as they remove the neurotoxic particles from solution(175)

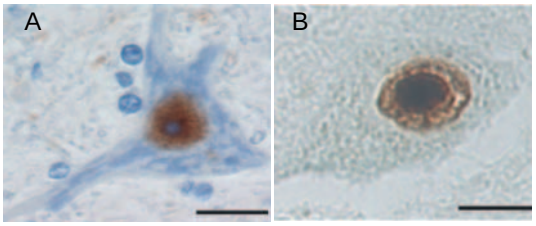


Fig. 2.3 NIs of AR. A) Web-like pattern of nuclei in the presence of aggregates of AR in anterior horn neurons. B) AR aggregates stained using 1C2 antibody in an anterior horn neuron. Adapted from (150)

However, these controversies are outweighed by the increasing number of studies that report a correlation between NIs and toxicity.Therefore we centered our interest in discussing how the aggregates perturb the cell environment and what their properties are.

The mechanisms proposed for explaining the toxic effect of the fibrils range from the deterioration of the cell clearance system to the degeneration of interactions with diverse proteins. It cannot be excluded that many of the mechanism here listed coexist and contribute to overall cell toxicity (Figure 2.4).

2.3.1 AGGREGATES INHIBIT CELL CLEARANCE MECHANISMS

Cell clearance is accomplished by two parallel mechanisms: proteasome-associated proteolysis and autophagy. The proteasome processes misfolded and short-lived proteins and specifically recognizes proteins marked by ubiquitination(121).

As previously mentioned, impaired proteasome function is generally associated with neurodegeneration and misfolding diseases(176–178). The cell clearance system is considered key in reverting the effect of the disease in an HD animal model, where expression of the transgene in a conditional mouse mutant of HD was suppressed(172).

Furthermore, AR presents several sequences recognized by the proteasome complex or its cognate proteins, and overexpression of proteins associated with the proteasome lead to reduced toxicity in cell models of SBMA(179, 180).

Several Hsps cooperate with the proteasome to regulate aggregation. In cell and mouse models of SBMA, the administration of compounds that increase the expression of Hsp70, Hsp90 and Hsp109 lead to a decrease in cell death(181).

However, it has been also observed that an inhibitor of Hsp90, a chaperone that stabilizes client proteins in solution, ameliorates the phenotype of an SBMA mouse model. This can be explained by a selection of the pathway for proteasome-related degradation of the polyQ AR, in the absence of chaperones that stabilize the monomer(182).

Taken together, the proteasome and chaperone system collaborate to maintain the cell clearance of aggregates, whether by preventing the formation of oligomers or by targeting the aggregates already formed for degradation(110). The observation that this process works less efficiently with age(183) may explain the late onset of the disease, as, with time, the organism finds it increasingly more difficult to efficiently remove the aggregates, thus giving rise to their accumulation.

In addition, the eukaryotic proteasome is not able to degrade polyQ peptides. These polypeptides are released and accumulate and can be disposed of only through autophagy(129).

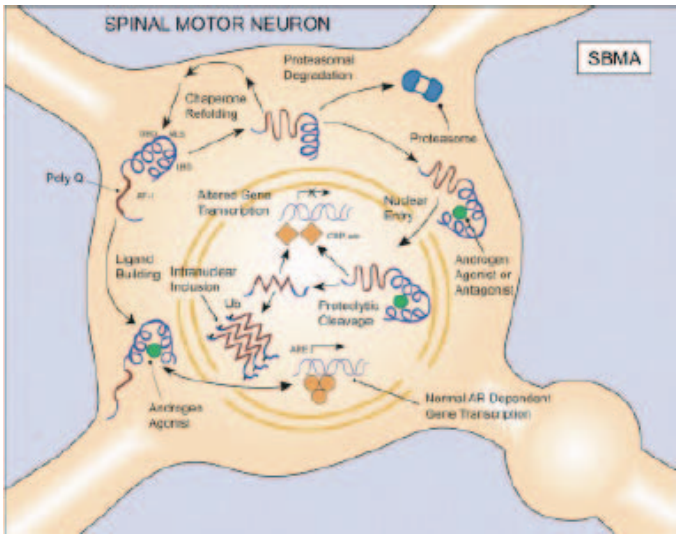


*Figure 2.4*Schematic representation of SBMA cellular pathogenesis. When polyQ-AR translocates into the nucleus, the protein forms NIs, which can sequester other nuclear proteins, like CBP, thus altering the gene transcription. The chaperone machinery and the proteasome contribute to keeping AR in the right fold or to degradingmisfolded or aggregated AR. Toxicity may arise because of inability to degrade the aggregates formed(184).

2.3.2 AGGREGATES SEQUESTER THE TRANSCRIPTION MACHINERY

Q/N-rich proteins can form strong interactions with polyQ proteins by coil-coil formation (103, 105, 185).Q/N-rich proteins are characterized by an abundance of Gln and Asn residues in their sequences, often organized in long runs of Q/N repeats and generally associated with yeast prions(186). A polyQ protein can be considered a Q/N-rich protein, and the involvement of a polyQ tract in coil-coil interactions has been already described in Chapter 1.8.

A longer polyQ tract might strengthen some interactions, and it is therefore possible that aggregates of AR accumulating in the nucleus sequester natural partners of this receptorby their polyQ tract, thus altering the transcriptional machinery. An example can be found in the interaction between polyQ AR and the CREB-binding protein (CBP)(187), another polyQ protein. This observation suggests that many other molecules interacting with AR may be

sequestered in the aggregates. To support these hypotheses, microarrays of cells expressing polyQ AR were performed with cell lines belonging to different types of tissue(188, 189). These studies identified a different up-regulation and down-regulation of the set of genes controlled by AR, together with a decrease in the transcriptional properties of AR with an increase in polyQ tract length.

However, results concerning gene expression must be analyzedcarefully, separating the relationships of cause and effect between dis-regulation and presence of aggregates. The expression pathways can also be altered in association with a generic stress response as a result of the intrinsic toxicity of the aggregates, which has no correlation with sequestering proteins belonging to the transcriptional machinery.

2.3.3 AGGREGATES HAVE ALTERED INTERACTIONS BETWEEN AR AND ITS PROTEIN PARTNERS

More generically, given that AR has more than 236 interacting proteins(190), it is feasible that each interaction between the N-terminal transactivation domain (NTD) of ARand its natural partners are compromised, not only in the nucleus but also in the cytoplasm.

Changes in interactions associated with the length of the polyQ tract have been widely reported. The deacetylase SIRT-1 is reported to interact with the polyQ AR and modulate its aggregation and toxicity(191). Also, the cytochrome c oxidase subunit Vb (COXVb) interacts with the polyQ AR with more affinity than wt AR and colocalizes in polyQ-AR aggregates but not in those of wt AR. These observations point to an association between the disruption of the cytochrome and cytotoxicity(192).

These observations are consistent with a scenario in which the polyQ tract is involved in protein-protein interactions and its length is critical in favoring or disfavoring certain interactions, by modifying the secondary structure of the protein(193) or by regulating the strength of interaction with a coil-coil partner(103, 194).

2.3.4 MUSCULAR AND NEURONAL DEGENERATION

Neurodegeneration is a specific feature of protein misfolding diseases, andin HD aggregates sequester proteins from the neuron environment, thus hindering trafficking and leading to neuronal cell death. In addition, animal and cell models of SBMA show impaired axonal transport(195–197).

However, degeneration in the myofibrils organization of the muscle cell has been proposed as another cause of toxicity. AR aggregates present a different localization in muscle tissues both in mouse models, and in SBMA patients. Also, muscle tissue of SBMA patients present marks of denervation and myopathy, like fiber atrophy, fiber splitting, and the presence of large fibers with central nuclei. These features also positively correlate with the degree of disability(146). Similar results were also found in SBMA mouse models(198).

This new mechanism of toxicity might be associated with the localization of the AR protein in skeletal muscle cells(168). Experiments in C2C12 cells revealed that AR also localizes at the level of mitochondria and microsomes(167). An interaction with Caveolin-1 was also reported, suggesting an additional localization in the plasma membrane(199)

All the listed mechanisms described to explain the toxicity should not be taken as alternative or mutually exclusive hypotheses, but instead as several aspects of a cytotoxic process in which each of them may be involved at different stages of the disease(184).

2.4 NUCLEAR INCLUSIONS AND PROTEOLYTIC CLEAVAGE

In each of these studies, particular care is dedicated to the nature and composition of the NIs that are systematically found in every system, due to their tight relationship with the onset of the toxicity.

Nevertheless, the structure, organization and properties of these aggregates are not understood. It is known that antibodies aimed at recognizing the C-terminal domain of AR fail to stain the aggregates, whereas antibodies directed against the N-terminus produce positive staining in immuno-histochemistry experiments(148). These observations suggest that the N-terminal AR is the main component of the aggregates.

For this reason, determining the effect of proteases that target AR is of considerable importance. A line of research addressed the role of the apoptotic caspase-3 (Cas-3), involved in a proteolytic cleavage of AR that produces three fragments. The N-terminal fragment of 17-20KDa corresponds to the first 160-200 residues (depending on the number of glutamines), which is the region containing the polyQ tract(200, 201).

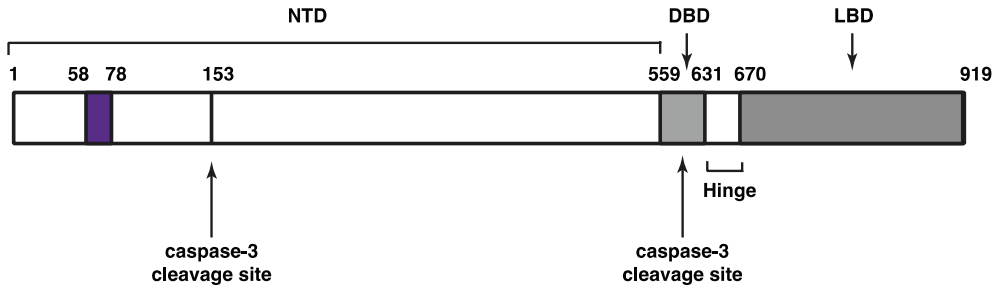


Figure 2.5Cleavage sites for caspase-3 in androgen receptor(200)*. The N-terminal fragment contains the polyQ tract and is considered the smallest fragment necessarily present in the aggregate.*

This study associates the cleavage of Cas-3 with an increase in toxicity of the polyQ AR(200), so that a toxic fragment is released and freely interacts with other fragments to produce the first oligomers. Other studies describe this proteolytic cleavage as a late event in the progression of the toxicity, thus challenging this hypothesis(202). Nevertheless, it is true that Cas-3 is the apoptotic protease that cleaves AR at the most N-terminal position, thus determining the smallest fragment produced in a biological context.

It is therefore reasonable to assume that the AR fragment cleaved by Cas-3 would be the fraction of protein necessarily present in the NIs. The characterization of the AR aggregation using this fragment is therefore a reasonable model system.

CHAPTER 3: ANDROGEN RECEPTOR

3.1 GENERAL INTRODUCTION

The protein AR belongs to the steroid nuclear receptor (NR) superfamily, which also includes receptors for glucocorticoids (GR), progesterone (PR), estrogen (ER), and other thyroid hormones, as well as prostaglandins and some vitamins. The activation of these transcription factors is mediated by a specific ligand(203, 204), andthereforethey act as a critical link between external stimuli and the transcriptional responses in the cell.

NRs share a common domain organization, which consists of the following: an N-terminal domain, which is highly variable in sequence composition and length; a DNA-binding domain (DBD), which is the most conserved region; and a ligand-binding domain (LBD) at the C-terminus. The N-terminal region of these receptors is predicted to be intrinsically disordered and presents one or more transactivation units. The DBD is well-conserved across the members of the family and various species and is also involved in dimerization and DNA-binding. Between the DBD and LBD, there is a flexible and poorly conserved region that acts as a hinge and contains a nuclear localization signal (NLS), for the translocation to the nucleus. The LBD of nuclear receptors has a highly conserved structure, whereas the sequence is only moderately conserved(205).

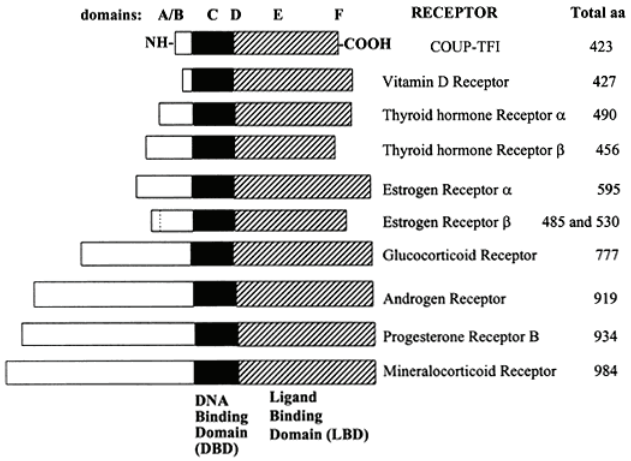


Figure 3.1Relative lengths of several members of the steroid/nuclear hormone receptor superfamily, shown schematically as linearized proteins with common structural and functional domains. Variability between members of the steroid hormone receptor family is primarily due to differences in the length and amino acid sequence of the amino (N)-terminal domain(206)

3.2 GENERAL BIOLOGY OF ANDROGEN RECEPTOR

Comprising 919 residues (110KDa), human AR is constitutively expressed and associated with the development of the male sexual phenotype and other pathways mediated by testosterone (207, 208). Common to all nuclear receptors, AR consists of an N-terminal transactivation domain (NTD), a DBD, and a LBD at the C-terminus. The AR gene is located on

the X chromosome, in position q11-12, and is organized in 8 exons: exon 1 codes for the NTD, two exons are dedicated to DBD and the remaining five exons code for the LBD (Figure 3.1).

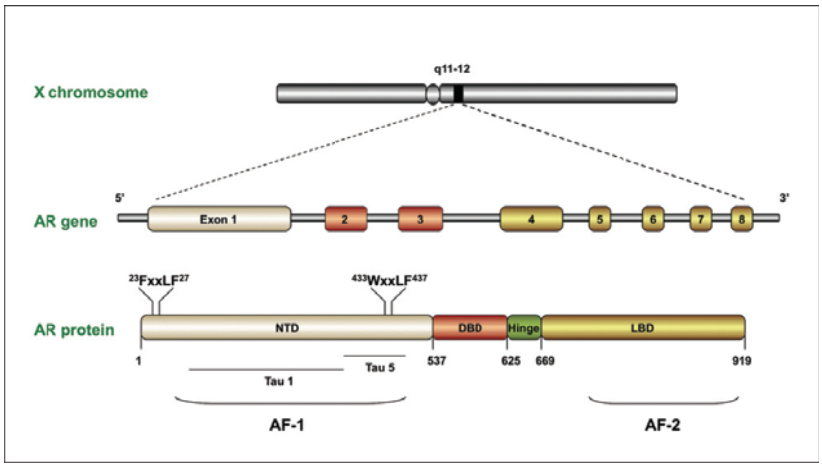


Figure 3.2 Schematic representation of AR gene on X chromosome and relative transcript. Important motifs are highlighted.(391)

Under physiological conditions, AR is present in the cytoplasm as a monomer associated with molecular chaperones, such as Hsp40, Hsp70 and Hsp90. The binding of testosterone or di-hydrotestosterone (DHT) to LBD causes a change of conformation. In this process, the NLS in the hinge region is exposed, the chaperones are released, and the protein dimerizes. Once the protein translocates to the nucleus as a dimer, it binds to a DNA sequence, called the androgen response element (ARE), and recruits other transcription factors for initiating transcription(209). AREs are specific sequences located close to genes controlling sex differentiation and cell homeostasis in tissues such as bones(210) and neurons(211).

3.3 DOMAIN ORGANIZATION

AR, like all the members of the NR family, has three clearly defined domains: the NTD spans residues 1 to 558, the DBD residues 559 to 622, the hinge region residues 623 to 669, and the LBD residues 670 to 919. This numbering refers to the GenBank version of human AR (GenBank: AH002607.1).

Figure 3.3 3D representation of AR. The structure of LBD (670-919) is PDB code 2AM9, for DBD (559-622) is PDB code 1R4I. NTD (1-558) and hinge (623-669) are an artistic representation.

The Protein Data Bank (PDB) holds 51 structures of the LBD (37 from human, 8 from chimpanzee, 5 from rat and 1 from mouse) and 1 structure produced from the DBD (PDB code 1R4I), but no structural information is available for the NTD, due to its disordered nature (Figure 3.4).

The following sections describe the properties of each domain, with a particular emphasis on the NTD.

3.4 N-TERMINAL TRANSACTIVATION DOMAIN

The NTD is the least conserved domain in NRs and the largest of the family. It constitutes more than 60% of the AR sequence, of which only 15% is conserved with other NRs(212). A substantial fraction of residues in the NTD is composed of highly repetitive sequences, which are also called low-complexity regions. They include the polyQ tract in residues 58 to 77, 84 to 89 and 190 to 197, and at residues 371 to 379 and 448 to 472. All of these features describe the NTD as an intrinsically disordered domain (for a detailed definition of disordered proteins refer to Section 3.4.1). For this reason, a structural characterization of AR NTD by X-ray crystallography is not possible, while its considerable size, comprising almost 600 aminoacids, makes structural studies by NMR particularly challenging.

The only partial information about the structure of this domain comes from biophysical studies, mostly by circular dichroism(212). The NTD is described mainly as disordered, but with 13% of α-helical content in a region called activation function 1 (AF1), described in 3.4.2.

Section 3.4.3 presents an extensive description of the low-complexity regions in the NTD, whereas Section 3.4.4 describes the short linear motifs that are reported to play key roles in AR.

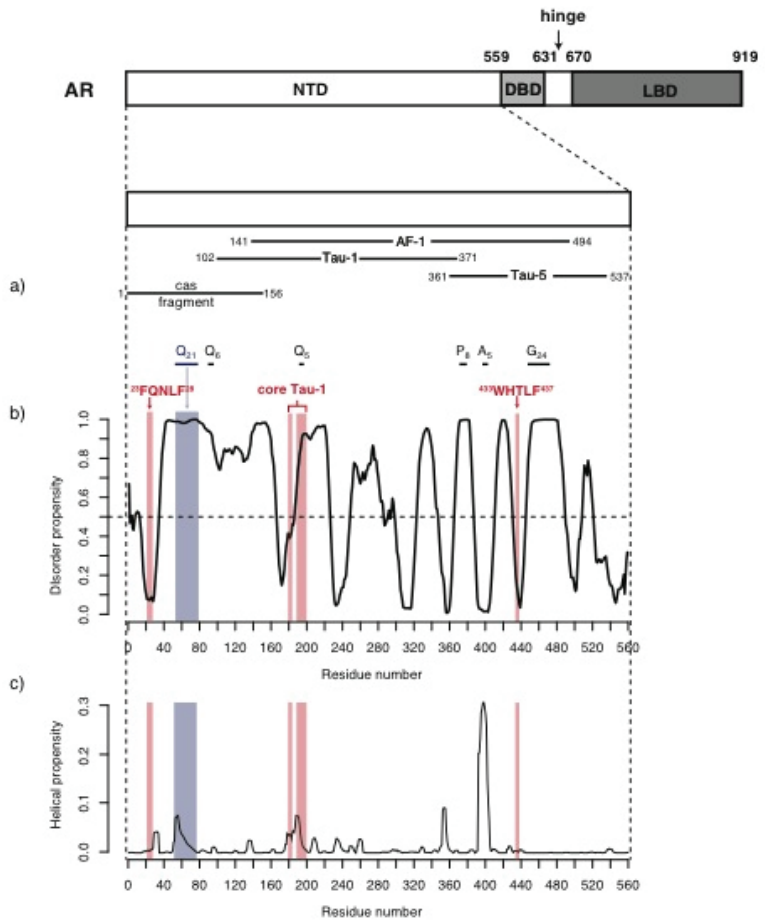


Figure 3.4Sequence features of NTD (a) and relative secondary structure predictions of (b) disorder propensity with PONDR and (c) helical propensity with Agadir. Key motifs are highlighted in red, the polymorphic polyQ is highlighted in purple.Cas fragment is the N-terminal product of the cleavage by caspase-3, discussed in Chapter 2.4.

3.4.1 INTRINSICALLY DISORDERED PROTEINS AND DOMAINS

Intrinsically disordered proteins (IDPs) are a class of proteins that are characterized by a lack of a well-defined three-dimensional fold(213). Eukaryotic proteomes are rich in proteins that present disordered regions of at least 30 residues in length (15-45%), while this feature is less abundant in prokaryotes (10-15%)(214).

As it is not possible to achieve a structural description of these proteins using the conventional structure-function paradigm, a more extensive concept of ensemble of conformations has been put forward(215). Particular emphasis is currently being placed on understanding how these proteins retain the functions in the absence of a specific fold and how these functions are regulated.

Many IDPs participate in signaling and regulation, due to their propensity to fold upon association with partner proteins(214). This process may involve the entire protein, specific regions or even single motifs, while the rest of the polypeptide chain often remains unstructured(216). These complexes of partially folded proteins have also been called fuzzy complexes (217). The hierarchy by which these two events (folding and binding) are associated is a matter of controversy in the field of IDPs. The *conformational selection* model describes the process as the recognition by the partner protein of a specific fold out of all

the possible conformations in the ensemble(218). The *induced fit* or *folding upon binding* model explains that the partner protein, once recognized and bound to the IDP, drives the folding of the protein in the conformation that best fits its pocket(216). Recently, a unified view was hypothesized, whereby one partner binds selectively not to one but to a limited range of conformations of its counterpart and nucleates a folding process once bound(219).

This plasticity in the fold and the ability to undergo different conformations is a key feature of domains associated with protein-protein interactions, because the same amino-acid chain can sample a wide range of conformations, which fit a number of partner proteins(213).

IDPs show relatively low specificity and affinity in their protein-protein interactions(220). It is possible that this relative promiscuity in interaction at low affinities is a critical aspect for their regulation in very crowded environments, like the cell. Phosphorylations and other post-translational modificationsare therefore required to regulate interactions between distinctpartners and to provide specificity(221–223).

3.4.2 ACTIVATION FUNCTION 1 (AF-1)

AF-1 is the main interface for protein-protein interaction in the NTD, and several proteins in the family of co-regulators of expression and transcription factors are reported to interact with it (224, 225). Two regions, called tau-1 and tau-5, are considered key in the transactivation of the AR, and mutations of these regions were found to be related to a form of prostate cancer known as castration-resistant prostate cancer (CRPC)(226).

The binding of AF-1 to the large subunit of TFIIF, RAP74, induces a partial folding with an increase in α-helical content(227, 228). TFIIF is a member of the transcription machinery complex and, when present, causes an increase in the affinity between AF-1 and another cofactor of the initiation complex, SRC-1 (225). The effect of TFIIF on the binding of AF-1 to SRC-1 is a typical example of allosteric regulation of the binding event.

However, it is also possible that proteins belonging to the transcription machinery complex interact with NTD and between each other, establishing interactions characterized by low affinities, but that stabilize each other(229).This feature would allow easy disassembly of the complex by means of post-translational modifications that can change the affinity of the partners.

3.4.3 SHORT LINEAR MOTIFS IN NTD

The NTD contains three linear motifs that play important roles in the transactivation of AR or in the interaction of the NTD with other cofactors. These motifs closely resemble LxxLL, a key motif present in several coactivators of nuclear receptors, such as the family of p160(230). LxxLL forms anα-helix and has been reported to interact with the LBD of AR in a hydrophobic cleft, formed by the second activation function region of the protein, AF-2 (further explained in 3.6).

Similarly, the two motifsFxxLF recognizable in the NTD (^{23}FQNLF^{27} and ^{433}WHTLF^{437}) take part in the most characterized interaction for this protein, namely the binding to AF-2 (231). This event contributes to the N/C interaction, critical for the dimerization of AR and its subsequent translocation to the nucleus(209, 232).

^{23}FQNLF^{27} is located at the very N-terminus of the protein and its sequence is highly conserved across species(233). The binding to the LBD is mediated by the ligand and this interaction involves a folding event of the ^{23}FQNLF^{27} motif, which acquires α-helical conformation. This event is described by an X-ray crystallography study (PDB 1XOW)(234). Isothermal titration calorimetry (ITC) measured an affinity of 1.2 ±0.2 μM and fluorescence polarization of 9.2 ± 0.4 μM by (234, 235).

The N/C interaction mediated by ^{23}FQNLF^{27} is critical not only for the translocation to the nucleus, but also for the manifestation of the SBMA phenotype in cell culture(236). ^{23}FQNLF^{27} is also reported to interact with cyclin D1 and the co-activator melanoma antigen gene protein 11 (MAGE-11), a protein that belongs tothe family of cancer germ-line antigens(237, 238). Furthermore, ^{23}FQNLF^{27} has close sequence similarity to another motif on huntingtin, FQKLL, and both motifs arereported to regulate protein turnover by a proteasome-dependent pathway(239).

The second LxxLL-like motif is ^{179}LKDIL^{183}, located in AF-1. It is critical for the transactivation of AR, as mutations on each hydrophobic residue dramatically decrease the transcriptional efficiency. Furthermore, this motif is reported to interact with AF-2, but with less affinity than ^{23}FQNLF^{27}(226). In spite of this characterization, other AR partners that recognize this motif are still undefined.

The last LxxLL-like motif is ^{433}WHTLF^{437}, which is again involved in an interaction with the LBD that takes place in the absence of hormone(240). In this case, however, it is not clear whether the interaction occurs on the hydrophobic cleft of AF-2 or with another region of LBD, which has not yet been recognized(241, 242).

3.4.4 AMINO-ACID REPEATS IN NTD

The polyQ region beginning in position 59 is the main topic of this thesis, but it is not the only highly repetitive sequence of the protein. The NTD contains at least another two shorter polyQ repeats, in positions 84-89 and 190-197, which are not polymorphic sites in humans, but are polymorphic in other species. For instance, the polyQ region in position 190-197 is polymorphic in *canidae*(243).

The other important highly repetitive and polymorphic region is the polyglycine (polyG) in position 449-472(244). This very flexible region is located in proximity to the beginning of the LBD(Figure 3.3).

The polyQ and the polyG regions might play a role in the transactivation of AR and the length of these two regions might influence the activity of the protein, as both short polyG and short polyQ are associated with higher transcriptional activity(245–247). In contrast, long polyQ regions are associated with hyposensitivity to the hormone, (248),in addition to their association with SBMA. Furthermore, shorter polyQ regions strengthen the N/C interaction and their deletion determines a drastic increase in transactivation *in vitro*(246, 249). Consistent with *in vitro* observations, population studies on the polyQ and polyG polymorphisms associates the presence of short polyQ and polyG tracts with a higher risk of developing prostate cancer(250).

3.5 DNA-BINDING DOMAIN AND HINGE REGION

The DNA-binding domain is the most conserved domain and presents a high degree of conservation across the other nuclear receptors(251). This domain binds to specific regions of the DNA, named androgen response elements (AREs), which are located in couples in the enhancer region of androgen-related genes(252).

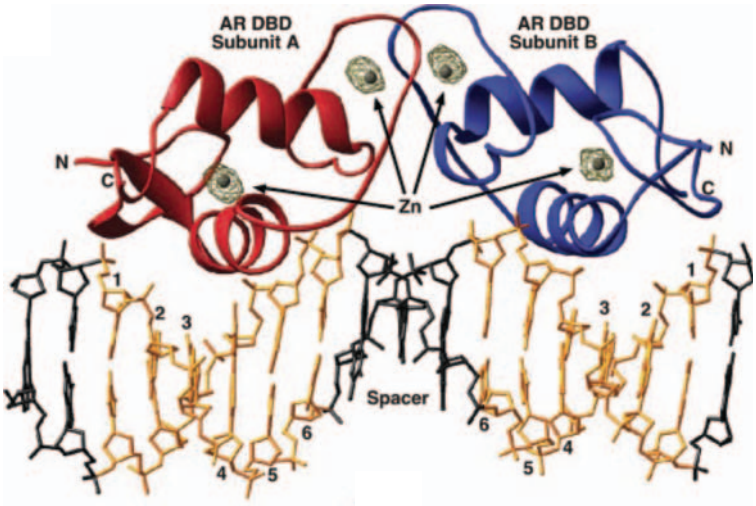


Figure 3.5 Crystal structure of AR DBD dimer bound to DL3 selARE DNA. PDB1R41(253)

The typical organization of the DBD on the NR is the Zn finger, a structure where 4 cysteines are coordinated by a Zn atom, folding the polypeptide chain in four loops(254).

The DBD of AR has two Zn-fingers, the first folds in an α-helix that accommodates in the major groove of the DNA and is responsible for the recognition of the ARE, while the second is associated with dimerization upon DNA-binding (255). The dimerization of the DBD is stabilized by hydrogen bonds between the residues of the two Zn-fingers and by complementarity in their surfaces(209).

A crystal structure of the DBD is available in the presence of DNA (Figure 3.5). The resolved structure of the DBD is that of a dimer with head-to-head orientation, in contrast to other resolved structures of NRs that are head-to-tail oriented(253, 256). This orientation of the DBD allows a higher stability of the dimer, due to a greater number of contacts compared to those established by other NR dimers(253).

The DBD also contains the first part of the bipartite NLS, while the second part is contained in the flexible and poorly conserved hinge region(255, 257). The NLS is formed by two clusters of basic residues separated by a short flexible linker between position 613 and 633 (^{617}RKCYEAGMTLGARKLKK^{633}). The NLS is exposed once the testosterone binds to the LBD and it facilitates the translocation in the nucleus. Mutations of this sequence determine impaired nuclear translocation and reduced activity(244).

The hinge region is also important for the selectivity of the target DNA. Moreover, it modulates the N/C interaction and recruits coactivators and corepressors(258).

The sequence similarity across the LBD of the various NRs is moderate, while considerable structural conservation is retained.

Generally, LBDs are folded in a 12 α-helix bundle, but the one AR lacks helix 2. In spite of the absence of this helix, the numbering remains the same in order to facilitate comparison with other receptors(259, 260).

Ligand-bound AR LBD is organized in a helix bundle, with a central cavity that accommodates the ligand. Testosterone establishes interactions with several residues in the inner part of the cavity and particular importance is attributed to the position of helix 12, which moves at the top of the cavity to lock testosterone in.

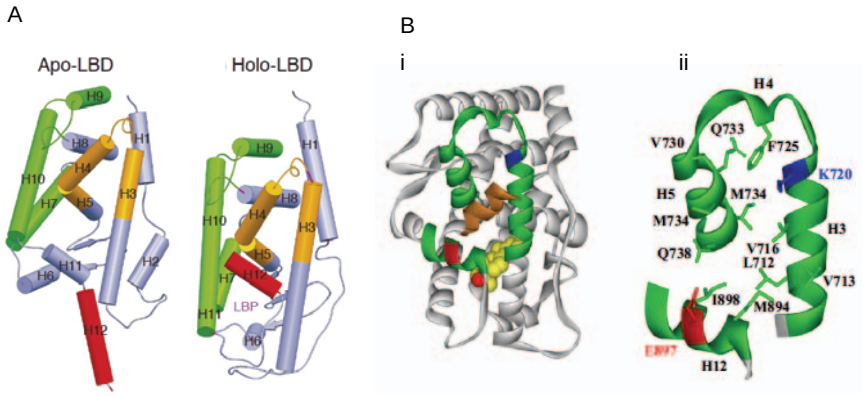


Figure 3.6 3D representations of LBD bound to testosterone. A) Schematic representation of the helix bundle of AR LBD in the unbound (Apo-LBD) and bound state (Holo-LBD). AF-2 is shown in orange, the dimerization region in green, and helix 12 in red(261). *B) Crystal structure of AR LBD (i) bound to hormone R1881 (yellow, oxygen in red) and to a peptide with FQNLF sequence (orange) interacting with AF-2 (green). In ii, AF-2 is isolated and the residues forming the charge clamp are shown in red (E897) and blue (K720)*(262).

As no successful attempt to crystalize unbound LBD has been reported, it is possible that the unbound state shows considerably lower stability. A common view, supported by the fold similarity to other NRs, is that helix 12 in the absence of ligand protrudes away from the rest of the domain and folds back over the cavity upon binding(263).

The repositioning of helix 12 is a crucial event for the dimerization and binding of other coregulators, as it determines the formation of the hydrophobic cleft (AF-2)(264).

In this regard, AF-2 is composed by helices 3,4,5 and 12(234) and is known to interact with the LxxLL motif, which is carried by co-activators like the family of p160, as mentioned in 3.4.2. Residues K720 and E897 establish a charge clamp with the LxxLL motif in helical conformation. As a result of the charged residues, this interaction provides selectivity. However, AR AF-2 is peculiar among the AF-2 of other NRs, as it shows lower affinity for LxxLL motifs and binds preferentially to FxxLF motifs, like ^{433}WHTLF^{437} and ^{23}FQNLF^{27}, as discussed in 4.3.3.

PART II: AIMS AND OBJECTIVES

CHAPTER 4: AIM OF THE RESEARCH AND OBJECTIVES

As is the case for rare diseases, SBMA is relatively little studied and no treatment is currently available.

The main goal of this project is to study how AR self-assembles to form the aggregates from a biophysical perspective. Therefore, before any proper scientific achievement, a suite of techniques had to be adapted from methods already in literature or tailored from scratch, with the aim to obtain reproducible data.

For this reason, the goals of the research are split into two blocks: methodological and scientific.

4.1 METHODOLOGICAL OBJECTIVES

1. Clone the recombinant protein designed from the Cas-3-cleaved fragment of AR.

2. Develop a protocol for the expression and purification of this recombinant protein and obtain yields high enough for performing experiments.

3. Obtain a reproducible starting point for the kinetics measurements in terms of a highly monodisperse and monomeric protein solution and validate this methodology.

4. Develop a set of techniques to monitor the oligomerization, aggregation and deposition of fibrils in a reproducible and most quantitative manner possible.

4.2 SCIENTIFIC OBJECTIVES

1. Characterize the early stage of aggregation of the recombinant protein, in order to understand whether other regions outside the polyQ tract contribute to its aggregation and to what degree.

2. Confirm the correlation between polyQ length and aggregation rate, with special emphasis on studying the role of flanking regions in this process.

3. Study the various steps of aggregation and the forces at play.

4. Characterize the aggregates formed by the recombinant protein and identify the biophysical properties of each species of aggregate, with the final aim of breaking down the process that leads to the formation of fibrils, from the early steps to the last species observed.

The findings from all of these activities will help to fill the gap between the current knowledge on polyQ diseases like HD and that on SBMA, In order to design new therapeutic approaches.

PART III: MATERIALS AND METHODS

CHAPTER 5: MATERIALS AND METHODS

In this chapter, a brief description of each technique applied will be provided, together with the conditions in which the experiments were performed. The results and the outcome of each application will be treated extensively in the results section.

5.1 CLONING AND MOLECULAR BIOLOGY

Molecular biology techniques were used both to produce the first construct for the expression of the wt Cas-3-cleaved fragment of AR (called Nt25Q) and for the introduction of mutations in the oligomerization study.

Challenging mutations were outsourced to the custom-synthesis service of GeneArt. Each of these products (Table 5.1) was codon-optimized for *E.coli* expression and was delivered in a pDONR221 entry vector, provided with AttB sites for the Gateway transcloning system. The coding sequence of the motif recognized by Tobacco Etch Virus (TEV) protease was added at the 5' end of each gene.

Nt51Q	Nt25Q_ALLL	Nt25Q_AAAA
Nt4Q	Nt25Q_LALL	Nt25Q_ΔL
	Nt25Q_LLAL	
	Nt25Q_LLLA	

Table 5.1 Constructs purchased from GeneArt codon optimized for E.coli in pDONR 221 vectors.

5.1.1 GATEWAY CLONING SYSTEM

The recombinant Gateway technology (Life Technologies) is a cloning system that takes advantage of an enzyme mix with the commercial name of Clonase II (Life Technology) to integrate specific fragments of DNA in vectors of the Gateway family.

This system is based on the recombination system used by *Lambda* phage to integrate in the *E.coli*chromosome. Specific sequences, both on the DNA of the target (called attachment sites B, or *att*B) and flanking the DNA to integrate (called *att*P), are recognized by a set of proteins of the Lambda phage with the commercial name of Clonases. These sequences are located at the 5' and 3' of the coding region and are called *att*B1 (or *att*P1) and *att*B2 (or *att*P2),respectively. They are non-synonymous and non-palindromic, to avoid inverted incorporations of the gene of interest (GOI).

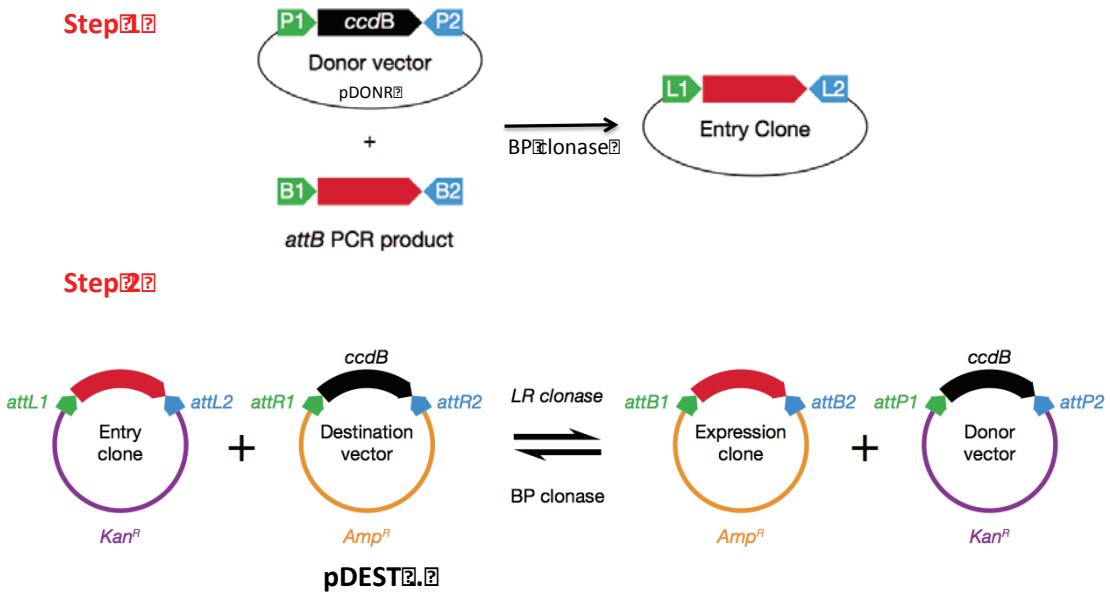


Figure 5.1 Gateway recombination system. Step 1 allows a PCR product provided with attB sites to be integrated in a pDONR by a BP Clonase reaction. Another Clonase mix, LR Clonase, allows the recombination of the GOI from the entry vector to the destination vector (pDEST). The BP Clonase can revert this reaction and provide the GOI in the entry vectoragain. Different antibiotic resistances allow negative selection, while positive selection is provided by the removal of the toxic ccdB gene from the pDONR (in case of BP Clonase) or pDET (in case of LR Clonase)(265).

A first set of enzymes, called BP Clonase, recognizes the *att*B site on the PCR fragment and the *att*P site on the entry vector pDONR 221 and brings them together, exchanging the sequence between *att*B1 and *att*B2 sites with that between *att*P1 and *att*P2 sites. This recombination produces a new flanking sequence for the GOI, called *att*L,which is recognized by another set of enzymes, with the commercial name of LR Clonase.

LR CLonase mediates the recombination between the entry clone loaded with the GOI and the destination vector of the pDESTfamily. These vectors cover a wide range of applications, from bacterial to human expression, and allow a choice between the most applied tags, thus enabling this system to switch rapidly from one expression vector to another.

Another practical advantage is that the flanking sequence produced by a LR reaction is a*att*P site, thereby opening the possibility to move the GOI back to the pDONR again through a BP reaction.

BP Reaction	
PCR product	0.5 µl (63ng)
pDONR 221	1 µl (150ng)
TE Buffer	6.5 µl
LR Reaction	
Loaded pDONR	6.2 µl
pDEST	1 µl
TE Buffer	6.5 µl

*Table 5.3*Amounts applied for the Gateway reactions

The negative selection is provided by a different set of antibiotic resistances, while the *ccdB*gene provides positive selection. The toxic *ccdB*gene is under a constitutive promoter, embedded between the *att*sites of both pDONR and pDEST vectors, so that those vectors that do not participate in the reaction kill the *E.coli* cell in which they are transformed. A successful BP reaction exchanges *ccdB*with the GOI, providing positive selection(265).

For this project, pDONR 221 was used as entry vector for every gene and pDEST17 (Life Technologies) and pDESTHisMBP (Addgene)(266) plasmids were applied for expression.

The amountsof each reagent used for BP and LR reactions are described in Table 5.2. Reactions were carried out at RT for 1-3 h and stopped with 1 µl of Proteinase K (provided with the kit) and 10 minofs incubation at 37°C.

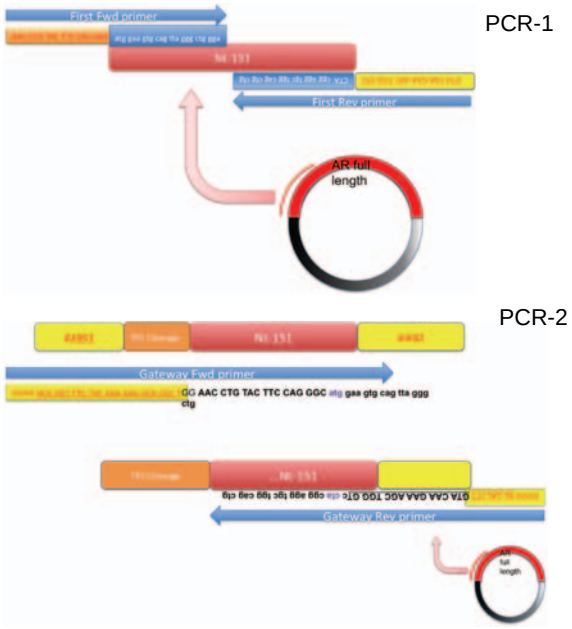


Figure 5.2Schematic representation of the PCR strategy. PCR 1 isolates the fragment (red) from the template plasmid (black) and adds the TEV cleavage site (orange). PCR 2 completes the addition of the attB sites (yellow).

PCR 1	Time (s)	PCR 2	
94ºC	30	94ºC	X 35
56.5ºC	60	59.6ºC	
72ºC	60	72ºC	
72ºC	900	72ºC	
4ºC		4ºC	

Table 5.4 Thermocycler settings used for the two PCRs. The $T_a$values differ due to the use of distinct primers.

An important condition for the success of these reactions was to maintain a 1:1 stoichiometry between the donor DNA molecule (whether the PCR product or the loaded pDONR 221) and the acceptor DNA molecule. The amounts reported are based on this rationale.

The whole reaction mix was used to transform OmniMAX*E.coli* chemically competent cells (Protein Expression and Purification Facility, PECF, IRB Barcelona). The transformed cells were plated on LB agar in presence of, respectively, Kanamycin for pDONR 221 transformed cells and Ampicillin for pDEST transformed cells.

After 16 h of incubation at 37°C, colonies were picked and expanded for DNA extraction (kit by Macherey-Nagel), and the concentration of the purified DNA was measured by a Nanodrop2000 spectrophotometer (Thermofisher). The purity was assessed by evaluating

the ratio of absorbance intensity at 260/280nm (267). The purified plasmids were finally sent for sequencing (GATC Biotech).

5.1.2 POLYMERASE CHAIN REACTION (PCR)

PCR was used to isolate the fragment cleaved by Cas-3 and to add the AttB sites needed for the Gateway cloning system explained above. PCR was also used for some of the mutations (Section 5.1.3). The thermocycler used for every PCR treated is an EppendorfMastercylcerEpgradient S and all the primers were purchased from Sigma Aldrich as a custom synthesis.

The enzyme kit of choice was AccuPrime GC-rich DNA polymerase (Life Technologies), due to its high fidelity and efficiency in replicating highly repetitive and GC-rich templates. The cloning strategy involved two PCR steps, as shown in Figure 5.2. First, the construct was isolated from a plasmid containing the human cDNA of the full-length AR gene (purchased from Addgene)(268), and a DNA sequence coding for the motif ENLYFQG was added at 5'. This motif is recognized by the TEV protease and is required by the purification protocol established (see 5.3). The second step amplifies the DNA fragment and adds the attB sites, *attB1* at 5' and *attB2* at 3' respectively (Figure 5.2).

The primers used for each PCR are listed in Table 5.4, the amounts used were those recommended by the supplier. The PCR conditions are listed in Table 5.5.

Between every PCR reaction, the PCR product was loaded on a freshly prepared 2% agarose gel and run in TAE buffer (40mMTris Acetate, 1mM EDTA) for 1 h at 100mV. 1-2 µl of SYBER-SAFE (Life Technologies) is dissolved in the agarose gel for visualizing the DNA bands with a UV light source.

This procedure allows determination of the length of the PCR fragment and separation from other PCR products, by using a gel extraction commercial kit. The best performance was obtained with an Illustra GFX Gel Band and PCR purification (GE Healthcare), following the manufacturer's instructions.After the BP reaction, the final product was sent for sequencing (GATC Biotech).

PCR 1	T_m*	% GC	nt
Forward AAC CTG TAC TTC CAG GGC atggaagtgcagttagggctg	61.2°C	54	39
Reverse GTACAA GAA AGC TGG GTcctacggaggtgctgg cag ctg	67.3°C	67	39
PCR 2			
Forward GGGG ACA AGT TTG TAC AAA AAA GCA GGC T GG AAC CTG TAC TTC CAG GGC atggaagtg cag ttagggctggga	71.3°C	52	73
Reverse GGGG AC CAC TTT GTA CAA GAA AGC TGG GTcctacggaggtgctgg cag ctg	71.8°C	59	51

Table 5.5 Primer sequences and properties. T_m*, melting temperature, considering only the overlap, %GC, percentage of GC, nt number of bases.

5.1.3 MUTAGENESIS

A Q5 Site-Directed Mutagenesis kit (New England Biolabs), which is designed for multiple mutations in close distance, insertion of relatively long DNA fragments or vast deletions, was used.

This strategy consists of two main steps: a linear PCR amplification and treatment with an enzyme solution. In the former, the mutation is inserted via mismatching primers. The enzyme mix comprises a kinase, a ligase,and the restriction factor DpnI. The kinase and ligase circularize the linear dsDNA, while DpnI recognizes and cleaves only methylated cytosine nucleotides (CpG) and therefore digests only the template DNA (269). Finally, transformation in DH5-αsupercompetent cells is required for maximizing the efficiency.

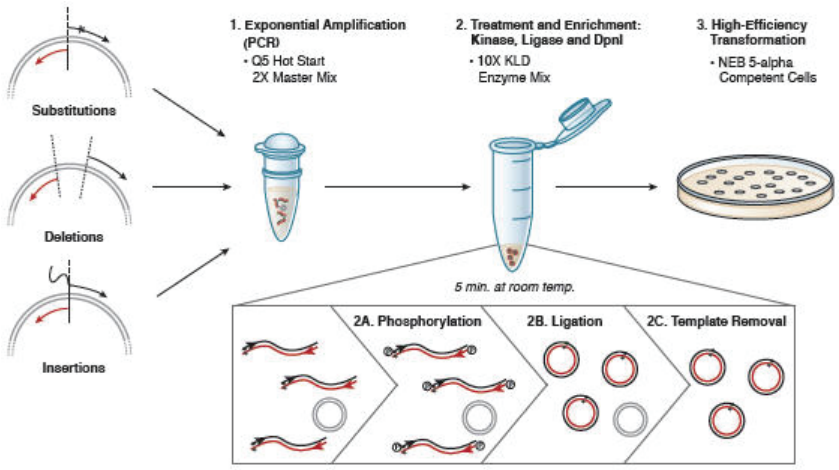


Figure 5.3Q5mutagenesis strategy. The primers are designed for a linear amplification, and the product is circularized and digested to remove the template DNA.

The primers were designed with the help of the NEB-BaseChanger software (http://nebasechanger.neb.com/), and their properties are described in Table 5.6. The PCR reactions were performed with the amounts recommended by the supplier and 1.5 ng of template DNA. The DNA templates for each mutant are specified in Table 5.6.

The times and temperatures of the PCR reactionsused for these mutagenesis were those indicated by the supplier. Particularly critical for the success of the reaction is the annealing temperature (T_a), which, in contrast to canonical PCR conditions, is higher than the melting temperature (T_m)(Table 5.7)(269). The annealing of a parental DNA with the primer is energetically disfavored,; however,when the replication of one fragment is completed, this rare event will produce a PCR fragment with a much higher affinity for the primer and therefore a higher $T_m$than that of the other sequences in the mix. The mutated fragment is therefore enriched at every cycle.

Mutant	T_m	%GC	nt
Nt25Q_AQNAA (template: pDestHisMBPNt25Q)			
Forward tgcggccCAGAGCGTGCGCGAAGTG	67	56	25
Reverse ttctgggcAGCTCCTCGGTAGGTCTTG	64	52	27
Nt25Q_SAREAA (template: pDestHisMBPNt25Q)			
Forward agcggccCAGAACCCGGGCCCCAGG	72	68	25
Reverse tcgcgcgcGCTCTGGAACAGATTCTGGAAAGCTC	67	56	34
Nt25Q_AQNAAQSAREAA (template: pDestHisMBPNt25Q_AQNAA)			
Forward tgcggccCAGAGCGCGCGCGAAGCG	74	80	25
Reverse ttctgggcAGCTCCTCGGTAGGTCTTGGACG	71	61	31

Table 5.6 Primers and relative properties for each mutant performed. The DNA templates are listed below the name of each mutant, the locus of the mutation in capital letters. $T_m$are calculated with the NEB BaseChanger software.

	Temperatures (°C)			
Time (s)	AQNAA	SAREAA	AQNAAQSAREAA	
30	98			
10	98			X25
20	67	70	72	
420	72			
600	72			

Table 5.7 Temperature and times used for the PCR. Times are equal for the three mutations while temperature differs only for the T_a tuned on the primer design. 25 cycles were performed.

1 µl of PCR product was digested and ligated with the mix provided by the kit,following the specifications of the supplier, and transformed in DH5αsupercompetent*E.coli* cells (NEB). From each successful transformation, 10 colonies were selected and expanded for DNA extraction. The UV absorbance at 260 nm and 280 nm was measured for each purified plasmid (Nanodrop 2000) and the DNA of each clone was sent for sequencing. Only one plasmid among those with a confirmed sequence was used for expression.

5.2 EXPRESSION SYSTEM

All expression constructs belong to the pDEST family. pDEST17 (Life Technologies) expresses a 6xHis tag fused N-terminal to the recombinant protein under the control of a pLAC inducible promoter, for expression triggered by the analog of lactose, Isopropyl β-D-1-thiogalactopyranoside (IPTG).

pDESTHisMBP (Addgene) is derived from pDEST 17 and expresses the recombinant protein as a fusion protein C-terminal to a maltose binding protein (MBP) that carries a 6xHis tag at its N-terminus (His-MBP). The MBP was chosen as a fusion protein because of its high solubility and the steric hindrance that opposes direct interactions of other monomers(270). The recombinant protein is therefore expressed fused at the C-terminus of a His-MBP protein.

All constructs were expressed in *E.coli* Rosetta BL21 cells (PECF-IRB), in LB media (Melford) after induction with 0.5 mM IPTG (Sigma). This expression system was adapted from(271). The expressed fusion proteins caried a TEV cleavage site between the MBP and the recombinant protein in order to allow removal of the tag in the last step of purification.

Glycerol stocks of transformed cells were stored at -80°C and retransformed when the expression levels appeared to drop. Expression of Nt51Q was performed onfreshly transformed cells due to drastic decreases in yield from one preparation to one other.

A pipette tip was plunged into the frozen glycerol stock and then into a 200-ml LB pre-culture. The LB pre-culture was grown for 16 h at 37°C and 220 rpm before dilution to 1:20 in the expression flasks containing 4L of LB.

The inoculated cells were grown at 37°C and 220 rpm, until reaching an optical density (OD) of 0.7. The OD was monitored by subsequent measurements with a spectrophotometer Cary 50 Bio (Varian), set at 600 nm. The cultures were induced with 0.5 mM IPTG (Sigma), and expression was carried out at 28°C and 200 rpm for 4 h.

^{15}N- and ^{13}C-labeled samples were produced in MOPS media (Melford) enriched with $^{15}NH_4Cl$ and (when necessary) ^{13}C-glucose (Eurisotope) or non-isotopically labeled glucose (Sigma). The expression protocol was adapted from(272).

The LB pre-culture was grown in LB for 16 h at 37°C and 220 rpm and then diluted 1:20 in a fresh 8 Lof LB culture. The LB culture was grown until OD ≤ 0.7 and pelleted by mild centrifugation at 2000 *g* for 20 min at 4°C. The pellets were rinsed with MOPS buffer and pelleted again with a 5000 *g* centrifugation for 10 min at 4°C. The pellet was then re-suspended in 2 L of MOPS culture medium and incubated at 37°C 200 rpm for 1 h, in order to allow the cells to adapt their metabolism to the new media. It is of critical importance to maintain an OD between 0.6 and 0.7 and to concentrate the cells by a factor of 4 when changing the medium This procedure ensures the maintenance of expression yields(272).

Finally, expression was induced with 0.8 mM IPTG for 4 h at 28°C in shaking conditions at 200 rpm.

5.3 PURIFICATION STRATEGY

Ni buffer A	Ni Buffer B	Size Exclusion Buffer	TEV buffer	RevNi A	RevNi B
20mM NaH_2PO_4	20mM NaH_2PO_4	20mM NaH_2PO_4	20mM NaH_2PO_4	20mM NaH_2PO_4	20mM NaH_2PO_4
500mMNaCl	500mMNaCl	500mMNaCl	100mMNaCl	100mMNaCl	100mMNaCl
5% Glycerol	5% Glycerol	5% Glycerol	0.5mM EDTA	8M urea	8M urea
1mMßMeEtOH	1mMßMeEtOH 500mM Imidazole	1mM DTT	ßMeEtOH	1mM ßMeEtOH	1mMßMeEtOH 500mM Imidazole
pH 8.0	pH 8.0	pH 7.5	pH 8.0	pH 8.0	pH 8.0

*Table 5.*8 Summary of the buffer used in the purification protocol.ßMeEtOH stands for β-mercapto-ethanol andDTT forDithiothreitol.

All constructs were expressed and purified following the same protocol, unless otherwise specified. This protocol was optimizedin order to obtain satisfactory performance with the most difficult construct handled, Nt51Q, and it facilitated an efficient purification all the other more manageable constructs.

Each centrifugation was performed on Beckman centrifuges and each purification in anÄkta Explorer (GE Healthcare) fast protein liquid chromatography (FPLC) instrument, maintained at 4°C.

The cell culture was pelleted by a 20-min centrifugation at 5000 *G* and 4°C and then resuspended in 80 ml of Ni buffer A (Table 5.8). Generally,the protein was expressed in cultures of 4 L. When larger expressions were required, a ratio of 20 ml of buffer per liter of expression was used to resuspend the pellet.

The following were added to the resuspended pellet:

- Lysozyme (Sigma) 1.5 mg/ml in a dilution of 1:500
- Protease inhibitor cocktail (PIC, Sigma) in a dilution of 1:50
- Phenylmethylsulfonyl fluoride (PMSF, Sigma) in a dilution of 1:100

The lysate was incubated at 4°C under mild agitation for 40 min and then sonicated (Vibra-cell, Sonics) for 20 min with pulses of 15 sec every minute. DNAse (Sigma) was added to the lysate in a dilution 1:200 and incubated for 1h in the previous conditions.

The soluble protein was separated from the debris by centrifugation at 25000 G for 30 min at 4°C, and the supernatant was passed through 0.22-mm sterile filters (Millex GP, Millipore) and purified by Ni affinity chromatography using a Ni-NTA HisTrap HP column (GE Healthcare). After passing the lysate through the column, it was extensively washed with 20 mM of imidazole (4% Ni buffer B, Table 5.8) in order to remove unspecific binders, and a gradient from 4% to 100% (500mM imidazole) of Ni buffer B (Table 5.8) was used for elution.

The fractions containing the peak were collected and concentrated to 10 ml with 10 KDa Molecular weight cut off (MWCO) concentrators (Millipore). This volume is small enough to be purified by loaded gel filtration.The protein solution was passed through a 0.22-mm filter

and injected in a Superdex 200 HiLoad 26/60 column (GE Healthcare), previously equilibrated in size-exclusion buffer (Table 5.8), and separated according tosize.

The fractions of the peak associated with the monomeric intact fusion protein were collected and incubated in presence of 1.2-1.8 mg of TEV protease in a dialysis bag with MWCO of 3.5 KDa (SpectraPor) immersed in 5 L of TEV buffer (Table 5.8). Incubation was performed at 4°C for 16 h under gentle stirring.

TEV protease specifically recognizes the sequence ENLYFQG and cleaves it between the glutamine and the glycine, producing a fragment that carries a glycine at its N-terminus, while the His-MBP retains most of the TEV site. The reaction was stopped by dissolving an amount of urea (Sigma) in the protein solution to reach a concentration of 8 M. This step has also the advantage of preventing aggregate formation and disrupting the early aggregates formed after cleavage.

The protein and urea solution was then passed through a 0.22-mm filter and loaded in a HisTrap HP column (GE Healthcare) previously equilibrated with RevNiA buffer (Table 5.8). After complete loading, a linear gradient from 0% to 100% of RevNi B buffer was applied to separate the target protein from the His-MBP protein.

After each purification step, SDS-PAGE electrophoresis (Life Technologies) was performed to check the purity and integrity of the proteins,using precast SDS-PAGE acrylamide gels with a 5-12% gradient (Life Technologies).

Pure protein was concentrated with a 3 KDa MWCO concentrator to 3 ml and dialyzed against 5 L of MilliQ water with a 3-ml dialysis cassette (Thermo Scientific), in order to remove urea. The dialyzed protein was then flash frozen with liquid N_2, lyophilized and stored at -80°C until use.To confirm the MW of each protein, multiple checks with mass spectrometry were performed at the end of the purifications.

5.4 MEASUREMENT OF CONCENTRATION

After size exclusion chromatography, the concentration of pure uncleaved protein was measured by UV absorbance at 280 nm (Varian Cary Eclipse). The extinction coefficient ε for Nt4Q, Nt25Q and Nt51Q fused to His-MBP was 79425, calculated with the ProtParam software (http://web.expasy.org/protparam/).

The concentration of the pure cleaved protein was measured by injecting a 10µl of sample into an Agilent 1200 reverse phase high-pressure liquid chromatography (RP-HPLC), equipped with a C-18 column (Phenomenex) and eluting it with a gradient of 0-90% acetonitrile (Panreac) in 30 ml, monitored at 215 nm.

5.4.1 HIGH-PRESSURE LIQUID CHROMATOGRAPHY

All concentration measurements of pure protein were performed by HPLC as specified in Section 5.4. In order to do this, it was necessary to produce a calibration curve with samples of known concentration. Amino acid analysis of a stable Nt25Q sample was performed as an independent approach to determine the exact concentration of protein. Amino acid analysis

was performed by the Scientific and Technical Services of the University of Barcelona (CCiT-UB).

The sample of known concentration was diluted to obtain a 5 point calibration curve, and each dilution was measured by injecting 10 µl sample into the HPLC chromatograph.The integrated value of intensity at 215 nm (Int A_{215}) corresponding to the peak of the pure protein was reported.

Given the similarity of the proteins studied, we assumed that all of them would have the same residual absorption at 215 nm, thus allowingthe application of this calibration curve to each protein measurement, correcting for the relative molecular weights. Intersperse checks with samples measured by amino acid analysis of each construct were performed over time, to assess the reliability of the measurement.

In a separate experiment, HPLC was also used to run analytical size exclusion chromatography with a Yarra 3000 (Phenomenex) column. In this case, 10 µl of sample wasere injected and eluted in 15 ml, with a flow of 1 ml/min in sodium phosphate buffer (Table 5.10). The signal was recorded at 215 nm.

5.5 SECONDARY STRUCTURE PREDICTION SOFTWARE

A suite of on-line tools has been developed by several groups to obtain preliminary insights into the secondary structure propensities of proteins analyzed. These software packages analyze the amino-acid sequence and produce residue-specific plots, predicting thepropensity of a given region to form α-helices, ß-turns, coil-coils or whether it belongs to a disordered region, depending on the software.

PONDRFIT (www.pondr.com) is a meta-predictor of disordered residues in a polypeptide chain. The sequence is analyzed in windows of 21 aminoacids and their properties are compared with a training library. The software produces as an output a value between 0 and 1 for the amino acid at the center of the window, where 0 means completely ordered and 1 completely disordered. The threshold for considering a residue as located in a disordered region is 0.5(273).

Agadir (http://agadir.crg.es/)is an algorithm used to predict the helical propensity of peptides. This algorithm was developed by performing an empirical analysis of experimental data, to estimate residue-level energy contributions to the stability of α-helical peptides(274). This database was used to train an algorithm that predicts the helical content of a window of residues according to the Helix-coil transition model(275). Properties of the solvent-like ionic strength and pH can be taken in account for the calculation.

Paircoil2 (http://groups.csail.mit.edu/cb/paircoil2/paircoil2.html) is the improved version of Paircoil and is an algorithm that recognizes coil-coil motifs on a protein sequence(276). Both Paircoil and Paircoil2 use pair-wise residue correlations obtained from a database of peptides reported as coil-coil partners to identify new potential motifs that can establish coil-coil interactions and distinguish simple α-helical motifs to helices that can establish this interaction(277).

5.6 PEPTIDE DESIGN AND SYNTHESIS

The peptides used in this project are summarized in Table 5.9 with the name used to refer to them.

Name	Sequence
$KKQ_{25}KK$	KKQQQ QQQQQ QQQQQQQQQQQQQQQ QQKK
$KKL_4Q_{25}KK$	KKLLL LQQQQ QQQQQ QQQQQQQQQQQQQQQ QKK
FQNLF	Ac-YPRPP SKTYR GAFQN LFQSV REVIQ NP-NH_2
C-term	Ac-GPTGY LVLDE EQQPS QPQSA LE-NH_2

Table 5.9 List of peptides and relative amino-acidic sequence

Peptides were purchased from Genscript through custom solid-phase peptide synthesis. The design of $KKQ_{25}KK$ and $KKLLLLQ_{25}KK$ was chosen following the work on the kinetics of polyQ peptides by the Wetzel group(85, 278, 279).

5.6 MONOMERIC SAMPLE PREPARATION

In order to obtain data in a reproducible fashion, a starting point with purely monomeric protein was strictly required.

Chapter 7 of the result section explains in details the process to reach the final protocol and the reasons of the choices done. Here are described the technical details of the two main protocols applied for obtaining monomeric proteins, without explaining the theory.

5.6.1 DISAGGREGATION WITH TFA:HFIP (WETZEL METHOD)

This protocol was developed by the Wetzel group in the university of Pittsburgh, to disaggregate efficiently polyQ peptides and peptides derived from Htt exon 1(85, 278).

Trifluoroacetic acid (TFA) is a very strong organic acid, due to the strong electronegativity of the trifluoromethyl group. Hexafluoro-2-propanol (HFIP) is a fluorinated alcohol, used to dissolve efficiently highly hydrophobic compounds. An extensive use of HFIP in disaggregating Aß fibrils and Htt peptides is described in the literature (280–283).

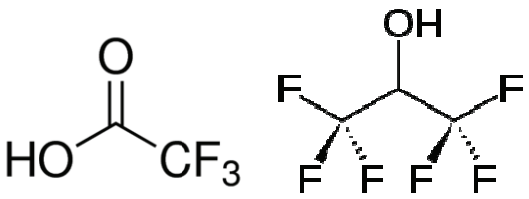


Figure 5.4 Structure of TFA (left) and HFIP (right)

Due to the nature of these two compounds and the amounts used, this procedure is highly hazardous and therefore carried out entirely under a fume hood, with double glove protection.

The protein was dissolved in pure HFIP, to obtain a 1 mg/ml solution in a 20 ml glass flask in the presence of a stirrer bar. An equal amount of TFA was then immediately added, in order to obtain a 1:1 TFA:HFIP solution. The flask was sealed with parafilm and left stirring at RT for 24 h.

The solvents were evaporated with a gentle Ar stream until disappearance of the liquid phase and then the flask was moved into a desiccator for 24 h. This step is required for removing traces of the two solvents, as it has been reported that traces of HFIP can influence aggregation kinetics for amyloid proteins(284) and that this solvent is a strong α-helix inducer(285, 286).

The protein layer was then dissolved in phosphate buffer (20mM NaH_2PO_4, 100mM NaCl, pH 7.5, 0.05% NaN_3) and transferred in aliquots of 200 µl in ultracentrifuge tubes for an ultracentrifugation step of 3 h at 386000 g, 4°C in a Optima TLX or Optima MAX ultracentrifuge (Beckman). The top ¾ of protein solution were taken from each tube and were considered the starting point of the experiment.

The final concentration was measured with HPLC as described in Section 5.4.

5.6.2 DISAGGREGATION WITH GUANIDINE THYOCYANATE 6M (LINSE METHOD)

This method was developed for Aß peptide in the group of Linse in the University of Lund(31) and has been extensively used for measuring aggregation rates of the Aß peptide in several conditions, in order to develop mathematical models of its aggregation(25, 33, 287). The method had to be further implemented, by introducing the final ultracentrifugation step, used in the Wetzel method.

GndSCN 6M buffer	20m mM Na Phosphate buffer
16.2 mM Na_2HPO_4	16.2 mM Na_2HPO_4
3.8 mM NaH_2PO_4	3.8 mM NaH_2PO_4
6 M GndSCN	1 mM TCEP
5 mM TCEP	0.05% NaN_3
pH 7.4	pH 7.4

Table 5.10 buffer composition of disaggregation with Linse Method

A buffer containing Guanidiniumthyocyanate (GndSCN) 6M was freshly produced before each experiment (Table 5.10), filtered with a 0.22 mm filter and used to dissolve the lyophilized protein. The protein solution was kept in shaking at RT for 3.5 h. A 1:5 dilution of the sample was measured by HPLC, to check the concentration and its oxidation state (see results Chapter 7).

A 500 µl aliquot of GndSCN protein solution was loaded in a gel filtration (PD-10) column for buffer exchange (GE Healthcare, MWCO 3.5 KDa), previously equilibrated in sodium phosphate buffer (Table 5.10). The sample was collected by fractioning every 500 µl after the void volume (2.5 ml) and measuring the UV absorbance spectrum with a Nanodrop 2000. The most concentrated fraction was spun for 10 min at 13000 *g* and 4°C before injection in a size exclusion column, equilibrated with phosphate buffer (Table 5.10).

The size exclusion column used is a Superdex 75 10/300 GL and the two 250 µl fractions that correspond to the peak of the monomeric protein are taken, pooled together and their concentration is measured.

The monomeric protein was transferred in aliquots of 200 µl in ultracentrifuge tubes for an ultracentrifugation step of 1 h at 386000 g, 4°C in a Optima TLX or Optima MAX ultracentrifuge (Beckman).

For obtaining monomeric peptides, it was necessary to adapt the protocol, by taking into account their small size. Therefore, the PD-10 gel filtration step was skipped, due to the MWCO (3.5KDa) being too close to the size of the peptides. Instead, just before injection in the size exclusion column, the solution of 6M GndSCN was diluted 2:3 with phosphate buffer, to decrease the concentration of GndSCN. This prevented the GndSCN from crystallizing in the loop, due to the low temperature of the FPLC instrument.

The column applied was a Superdex Peptide 10/300 (GE Healthcare), able to separate efficiently low molecular weight (MW) species, while the ultracentrifugation step was of 3 h at 386000 G, 4°C.

In both cases, the top ¾ of protein solution are taken from each tube and the concentration was measured by HPLC. The beginning of the experiment was considered when the dilution to the desired concentration was applied.

5.7 KINETIC MEASUREMENTS

A set of biophysical tools was adopted to follow kinetics of aggregation of the proteins and the peptides analyzed.

5.7.1 THIOFLAVIN-T BINDING

Thioflavin-T (ThT) is a fluorescent dye commonly used in biophysical studies to identify the presence of amyloid fibrils in an aqueous solution. It is a planar molecule (Figure 5.4) that interacts selectively with amyloid fibrils and, when it does, shows birefringence. The excitation and emission maxima are, respectively, 350 nm and 440 nm in aqueous solutions and with a relatively low quantum yield. ThT shows higher fluorescence when in presence of cross-ß-sheet rich deposits and its excitation and emission maxima shift to 440 nm and 486 nm. The use of a spectrofluorimeter allows quantitative measurements of the build-up of cross-ß-sheet content by measuring the fluorescence intensity at 486 nm with excitation at 440 nm (288).

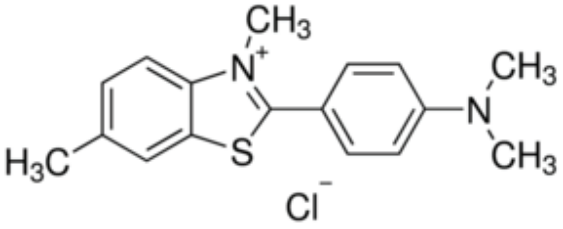


Figure 5.4Structure of Thioflavin-T (Sigma)

The mechanism of action is not entirely understood, but the most accepted theory is that the dye intercalates the fibril, binding to the hydrophobic residues that form the ß-strands.

The increase in rigidity given by the constraint of the two ß-strands determines the shift in emission and excitation maxima (289).

This assay was developed based on the protocol suggested by Dr. Natalia Carulla at IRB Barcelona.

For practical reasons, stock solutions were performed before the beginning of the experiment and conserved at 4°C:

- glycine-NaOH (Sigma) 300mM buffer pH 8.51
- Thioflavin-T (Sigma) 10mM

Aliquots of 25µl were taken from the incubating solution (generally of 20 µM concentration) and mixed with the ThT stock and the glycine-NaOH buffer, to obtain a final 50 µl sample of 10µM protein, 10 µM ThT, 50 mM glycine-NaOH pH 8.51.

The sample was measured by exciting at 450 nm (filter 450/10) and recording the emission at 486 nm (filter 486/20) in a spectrofluorimeterFLuoDia T70 (Photal), with 50 ms of integration time and 10 accumulations at 25°C. Alternatively, a spectrofluorimeter Cary Eclipse (Varian) was also used, by exciting at 440 nm and recording the spectrum of emission from 480 nm to 500 nm with excitation slit at position 5 and emission slit at position 10.

Data were analyzed subtracting the Intensity signal at 486 nm of a blank sample of ThT with buffer instead of protein.

5.7.2 SEDIMENTATION ASSAY BY ULTRACENTRIFUGATION

This assay was developed and extensively used by the Wetzel group of the University of Pittsburgh (22, 278). The main concept stands in measuring the concentration of monomer, or small oligomers, remained in solution, after a sedimentation step by ultracentrifugation.

As long as the monomers assemble forming oligomers that grow in size, the sedimentation coefficient of the particle increases, together with the speed by which the particle forms sediments. However, the sedimentation coefficients of monomer and small oligomers have very small differences, so that a monomer will be hardly separable by preparative ultracentrifugation in aqueous solutions by iteratively changing time the time and speed of ultracentrifugation.

Therefore, we decided to set a centrifugation time and speed based on our experimental data (see Chapter 7.5 in Results) and on data found in literature (278) with the awareness that, in order to obtain a purely monomeric sample, a fraction of the monomer is necessarily sedimented.

For the experiment in stirring conditions, the freshly disaggregated solution was kept in a 1.5 ml sterile glass bottle provided with a stirrer and 2 aliquots of 50 µl were taken from the top part of the solution. For the experiments in quiescent conditions, the freshly disaggregated solution was split in aliquots of 55 µl and sealed with Parafilm. The incubation in each experiment was performed at 37°C.

In both the described experimental conditions, two 50 µl aliquots for time point were taken and ultracentrifuged in a Beckman Optima MAX or Beckman Optima TLX ultracentrifuge for 1 h at 386000 *g* (104000 rpm) at 4°C.

Immediately after the centrifugation, the upper 25 µl were taken and moved to eppendorf tubes and the concentration is measured by HPLC.

Data were reported as mean of two values and they represent the fraction of monomer in solution. The concentration of protein at each time point was normalized for the concentration at t_0, which is the first measurement taken immediately after preparing the sample and before beginning the incubation at 37°C.

5.7.3 DYNAMIC LIGHT SCATTERING

Dynamic Light Scattering (DLS) is a tool extensively used to prove the monodispersity of a sample and, as a consequence, it is very efficient in identifying the presence of oligomers, also in low amounts.

The DLS instrument measures the Brownian motion of particles in solution by a laser beam that explores a window in a quartz cuvette. A digital detector counts the number of photons of scattered light over a defined time window. The speed by which a particle moves by Brownian motion depends on the viscosity of the medium and on the size of the particle. Particles of larger size will tumble more slowly across the window and scatter more light. Differential absorption of the particles in the sample could introduce inhomogeneity in the measurement, but this effect is prevented by using one single wavelength laser source that provides high stability and low dispersion of the beam(290).

The direct measurement of scattered light takes advantage of the Rayleigh theory, which correlates the hydrodynamic radius of a particle to the intensity of scattered light, according to the following equation:

$$I = I_0 16\pi^4 R^6 \frac{\left(\frac{n^2-1}{n^2+2}\right)^2}{(r^2\lambda^4)}$$

Where *I* is the intensity of light scattered and $I_0$the incident light, *R* is the radius of the particle, *n* the relative refractive index of the particle and *r* the distance of the detector from the scattering particle. From this equation is clear a very strong dependence of the scattered light on the radius of the particle, so that large particles risk to overwhelm the signal of other populations also when in low amounts. Another complication is the ratio between the refractive index of the particle and the one of the medium, which is very low, further decreasing the ability to sense populations of both large and small particles.

Instead of measuring directly the scattering of the particle, DLS measures the variation in scattered light by particles in Brownian motion in the medium over a window of time.

A second order autocorrelation function is calculated from the sequence of intensity values recorded:

$$g^2(q;\tau) = \frac{\langle I(t)I(t+\tau)\rangle}{\langle I(t)\rangle^2}$$

An autocorrelation function, in simplicity, is a function generally used to extract repeating patterns from a very noisy signal and it consists in correlating a value with itself at a different point in time. In this second order autocorrelation function g^2, the value of

intensity *I,* associated to a particular wave vector *q,*is correlated with the value of intensity at the time *t+τ*, where τ is the delay.

In the most simple scenario of a sample that is monodisperse, it is possible to produce a first order autocorrelation function and to resolve it as an exponential decay directly associated to the translational coefficient of the particle analyzed, according to these equations:

$$g^1(q;\tau) = Exp\ (-\Gamma\tau)$$

Where Γ is the decay rate:

$$\Gamma = q^2 D_t$$

D_t is the translational coefficient and:

$$q^2 = \frac{4\pi n_0}{\lambda} \sin\left(\frac{\Theta}{2}\right)$$

In this equation, λ is the incident laser wavelength, n_0 the refractive index of the sample and Θ the angle of the detection. DLS instruments have generally fixed angles, therefore is necessary to provide only n_0 to know the translational coefficient of a particle. The translational coefficient is function of the size and the shape of the particle according to the Stokes-Einstein equation:

$$D = \frac{k_B T}{6\pi\eta R}$$

Where k_B is the Boltzmann constant, T the temperature η the viscosity and R the radius of the spherical particle(291).

However, the majority of the samples is in reality a mixture of particles of different size (polydisperse) and the autocorrelation function produced will be the sum of all the exponential decays associated to each particle and the treatment of this calculation is not part of this thesis.

We used DLS mainly as a tool for knowing the size of the monomeric sample and to measure the rate of oligomerization by monitoring a representative value of the polidispersity and of the size of the oligomers produced, which is called z-average.

The z-average value (or cumulants mean) is the average of the sizes of all the particles in solution, weighted by their size. Mathematically, it corresponds to the fit of a polynomial to the log of the g^1 correlation function:

$$Ln(g^1) = a + bt + ct^2 + dt^3 + et^4 + \ \ldots$$

The value *b* is known as the second order cumulant fit or the z-average diffusion coefficient and it is converted in size units, taking in account the viscosity and the fixed angles parametrized by the software of the instrument.

The instrument used for these measurements is a Malvern Zetasizer Nano S equipped with a He-Ne of 633 nm wavelength laser. For each experiment, 20 repetitions of 20 seconds were recorded. Three measurements were performed at each time point.

Before the measurements, samples were spun with a tabletop centrifuge at 4ºC for 10 min at 13,000 rpm and then deposited in a quartz cuvette that was capped and sealed with Parafilm.

Depending on the length of the experiment, the measurements were automatized with the Malvern software, or repeated manually twice a day, keeping always the same settings and the sample at 37°C between incubations.

5.7.4 NATIVE GEL ELECTROPHORESIS

This kind of electrophoresis is used to study behavior of proteins in non-denaturing conditions, preserving the oligomeric state and the native structure of the protein analyzed.

Conceptually, this experiment is very similar to a conventional SDS-PAGE, with the difference that no SDS and no reducing agents are present and that the sample is not denatured before loading the gel. In this case, the sample can migrate according not only to the MW, but also to the charge, the oligomeric state and the degree of compaction at its native state. Oligomers will be visualized as multiple bands on the same lane.

The native gel was prepared for a neutral or acidic protein (Nt25Q, Nt4Q and Nt51Q have pI 6.33) in Native Tris buffer (Life technologies) with 8% of acrylamide (Sigma) and the samples were prepared by adding 4 µl of loading buffer (Life Technologies) in 16 µl of sample.

The gel was run at 4°C for 16 h at 20 mV and then stained with the same Coomassie staining solutions generally used for SDS-PAGE. Destaining procedure is analogue to the one used for canonical SDS-PAGE.

5.7.5 ANALYTICAL SIZE EXCLUSION CHROMAOGRAPHY

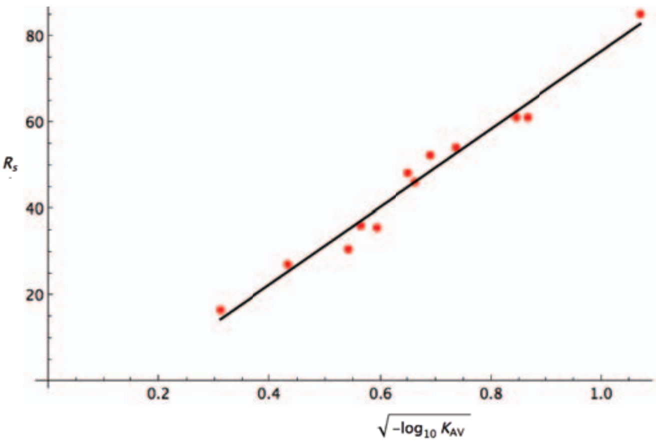
Analytical size exclusion chromatography (SEC) is another method extensively applied to know the size of a protein in solution, by passing the sample through a gel filtration column previously calibrated with a set of protein of known size and MW.

Whereas for folded proteins size and MW are in good approximation proportional, IDPs tend to adopt more extended conformations in solution and they present a larger size than folded proteins with the same MW(292). For this reason, it is necessary to accurately calibrate the SEC according to the size, or hydrodynamic radius (R_s), more than to the MW.

With this purpose a SEC column Superdex 200 10/300 GL (DGE Healthcare) was calibrated with a set of proteins provided in the kits for both low MW and high MW column calibration from GE Healthcare. Accurate measures of void volume V_0 and total volume V_T were performed, obtaining the values V_o = 7.9 ml and V_T = 20.780 ml. These values were used for resolving the following equation:

$$K_{AV} = \frac{V_E - V_0}{V_T - V_0}$$

K_{AV} is the partition coefficient and it is a way to represent with a single normalized number the relative retention volume (V_E) of a protein in the column used. Knowing the hydrodynamic radius of each of the protein used for the calibration, it is possible to develop a calibration curve that correlates the log of K_{AV} with R_s (Figure 5.5).



#	Protein Standards	Rs (Å)
1	*Ribonuclease A*	16.4
2	*Carbonic anhydrase*	27
3	*Ovalbumin*	30.5
4	*BSA*	35.5
5	*Conalbumin*	36
6	*Alcohol dehydrogenase*	46
7	*Aldolase*	48,1
8	*Catalase*	52.2
9	*b-amylase*	54
10	*Ferritin*	61
11	*Apoferritin*	61
12	*Thyroglobulin*	85.

Figure 5.5Calibration curve for R_s according to the elution volume in a Superdex 200 10/300 GL column (R_2=0.97612).

From this linear fitting it was possible to calculate the measured R_s as:

And $R_s$is calculated in Ängstrom (Ä) of radius.

The R_s calculated was then compared with theoretical R_s calculated from the MW by using a set of equations developed by the Uversky group, that produce different values of R_s, taking in account different degrees of disorder of the protein (Table 5.11)(292)

• Native	$\log(R_S^{N}) = -(0.204 \pm 0.023) + (0.357 \pm 0.005) \cdot \log(M)$
• Molten globule (natively **folded** protein on denaturing conditions)	$\log(R_S^{MG}) = -(0.053 \pm 0.094) + (0.334 \pm 0.021) \cdot \log(M)$
• Pre-molten globule (natively **folded** protein on denaturing conditions).	$\log(R_S^{PMG}) = -(0.21 \pm 0.18) + (0.392 \pm 0.041) \cdot \log(M)$
• Completely unfolded protein with urea	$\log(R_S^{U(urea)}) = -(0.649 \pm 0.016) + (0.521 \pm 0.004) \cdot \log(M)$
• Completely unfolded protein with guanidinum hydrochloride	$\log(R_S^{U(GdmCl)}) = -(0.723 \pm 0.033) + (0.543 \pm 0.007) \cdot \log(M)$
• Natively unfolded protein in extended conformation (coil)	$\log(R_S^{NU(coil)}) = -(0.551 \pm 0.032) + (0.493 \pm 0.008) \cdot \log(M)$
• Natively unfolded protein in a collapsed conformation (pre-molten globule)	$\log(R_S^{NU(PMG)}) = -(0.239 \pm 0.055) + (0.403 \pm 0.012) \cdot \log(M)$

Table 5.11Theoretical R_s from MW of proteins in different states of disorder(292)

This equation can only predict with a good degree of accuracy the $R_s$ofmonomeric proteins, as other events of folding and compaction can occur upon oligomerization, increasing the complexity of the behavior. Nevertheless, it was possible to compare the calculated $R_s$with all the theoretic $R_s$for the monomeric protein, to look for confirmation about the entity of the main peak observed and to gain some insight about the level of compaction of the protein.

5.8 STRUCTURAL AND MORPHOLOGICAL CHARACTERIZATION

A suite of biophysical tools was applied to study the structural properties of monomeric and aggregated protein and to characterize the morphology of the aggregates formed.

Tools like circular dichroism (CD), and Analytical Ultracentrifugation (AUC) can analyze samples in monomeric or low-oligomeric state, while techniques like transmission electron microscopy (TEM), atomic force microscopy (AFM) and Fourier transform infrared spectroscopy (FTIR) work better with aggregates.

5.8.1 CIRCULAR DICHROISM (CD)

CD is a technique that measures the difference in absorption of clockwise (R) and anti-clockwise (L) polarized light in an optically active sample. Chiral compounds, like amino-acids, present different absorption patterns of L and R polarized light and proteins analyzed by CD will reveal information on their structure.

The kind of information obtained depends on the window of wavelengths explored. Monitoring the wavelengths from 260 nm to 320 nm (near UV CD), in the spectrum of absorption of the aromatic chains, gives information about tertiary structure, however the result can be of difficult interpretation, due to heterogeneous contributions, like exposure of each aromatic residue, degree of freedom, excitation state.

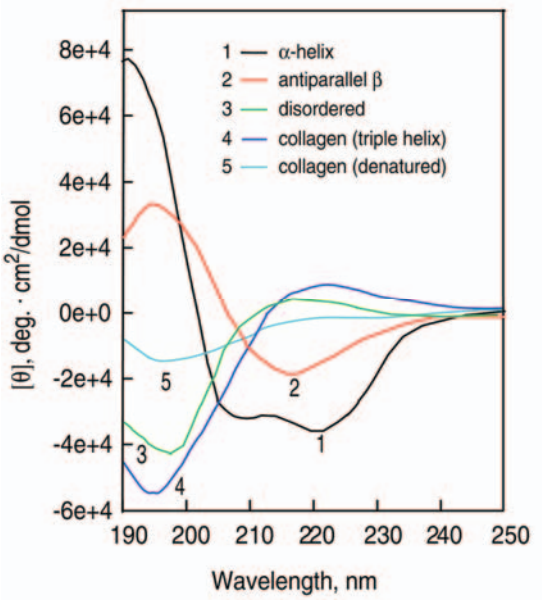


Figure 5.6Examples of CD signals for α-helix (1), ß-sheet (2), random coil (3), collagen triple helix(4)(392)

For this reason, CD is more extensively applied at low wavelengths (far UV CD), between 180 nm and 260nm, in the absorption spectrum of the backbone, to acquire information about the secondary structure(293). In this case, each type of secondary structure is characterized by one specific CD signal and it is possible, at least in theory, to compute what proportion of sample in solution adopts one specific secondary structure by linearly combining the signals for each secondary structure visited by the sample.

The raw output signal of a CD experiment is the difference between the R and L polarized light for each wavelength observed, called ellipticity (θ_λ) and it is expressed in millidegrees (mdeg).

The choice of the cuvette depends on the concentration of the sample and the buffer conditions, and general guidelines for reliable measurements are to maintain the high tension voltage (HT) of the instrument below the value of 600 mV and the absorbance below 1.0, to avoid excessive noise.

For the measurements in this project, proteins were analyzed using a Quartz SUPRASIL 100-QS 1 mm cuvette for proteins ranging from 10 to 30 µM and a Quartz SUPRASIL 100-QS 0.1 mm for proteins ranging from 100 to 150 µM.

The choice of buffer is another critical step for obtaining reliable data. Depending on the absorption properties that each buffer has at far UV and their concentration, they will be more or less applicable to CD spectroscopy. Table 5.12 summarizes their properties and the suitability to CD.

Component	Absorbance (50 mM solution in 0.02 cm pathlength cell)			
	180 nm	190 nm	200 nm	210 nm
NaCl	>0.5	>0.5	0.02	0
NaF	0	0	0	0
$NaClO_4$	0	0	0	0
Boric acid	0	0	0	0
Na borate (pH 9.1)	0.3	0.09	0	0
Na_2HPO_4	>0.5	0.3	0.05	0
NaH_2PO_4	0.15	0.01	0	0
Na acetate	>0.5	>0.5	0.17	0.03
Tris/H_2SO_4 (pH 8.0)	>0.5	0.24	0.13	0.02
HEPES/Na^+ (pH 7.5)	>0.5	>0.5	0.5	0.37
MES/Na^+ (pH 6.0)	>0.5	0.29	0.29	0.07

Table 5.12 Absorption properties of common buffers in the far UV. The higher the Abs at a given wl, the worst less suitable the buffer is(293)

Sodium phosphate buffer (Table 5.10) is reasonably good for measurements until 190 nm and it was applied for all the CD measurements.

Measurements were performed on freshly disaggregated monomeric samples with a JASCO 815 spectropolarimeter and the spectra were acquired at aresolution of 0.2 nm with a scan rate of 50 nm/min and 10 accumulations per spectrum. Two channels were used, for monitoring both ellipticity and HT, in order to have a quality control of the measurement. Data points at wavelengths with HT higher than 600 mV were discarded.

The raw output of ellipticity in mdeg is converted in a normalized value, called mean residue ellipticity (MRE). This conversion takes in account the number of peptide bonds of the protein, the concentration and the path length of the cuvette, to produce normalized spectra that can be compared across proteins. The MRE_λ value is calculated as follows.

$$MRE_\lambda(deg * cm^2 * dmol^{-1}) = \frac{\theta_{\lambda,sample}(mdeg) - \theta_{\lambda,blank}(mdeg)}{(N-1) * d(m) * c(mM)}$$

$\theta_\lambda$is the ellipticity value at the relative wavelength λ for the sample and the blank and it is expressed in mdeg, N is the number of residues of the protein, d is the path length expressed in m and c the concentration of protein, expressed in mM.

Deconvolution of the CD data was performed by the web calculation tool Dichroweb (http://dichroweb.cryst.bbk.ac.uk/html/home.shtml)(294, 295). This software analyzes the CD spectrum provided, by fitting it to a library of CD spectra with known secondary structure. It is possible to choose among a set of fitting algorithm and select the one that provides the best fitting in terms of lowest root mean square deviation (RMSD). The software finally interprets the spectrum as fractional compositions of α-helix, β-sheet and random coils. In this project, the two main algorithms applied were K2D and CONTIN and the results showed in the result section are the best fit, associated with the lowest RMSD from the simulated spectra.

5.8.2 ANALYTICAL ULTRACENTRIFUGATION

Analytical ultracentrifugation (AUC) is the most reliable method to assess the purity of a sample, its monodispersity and to obtain information about its size and oligomeric state, as every term of the equations used can be determined experimentally.

AUC sedimentation experiments are performed with an ultracentrifuge that carries a UV spectrophotometer and an interferometer, which measures the difference between the refractive index of a buffer solution (as a reference) and the same solution in presence of the particle analyzed (Rayleigh interference). The measurement is performed as the ultracentrifuge spins the sample and the variation in absorbance or Rayleigh interference is plotted as a function of the cuvette length. Two types of experiments are possible: sedimentation velocity (SV-AUC) and sedimentation equilibrium (SE-AUC)(292).

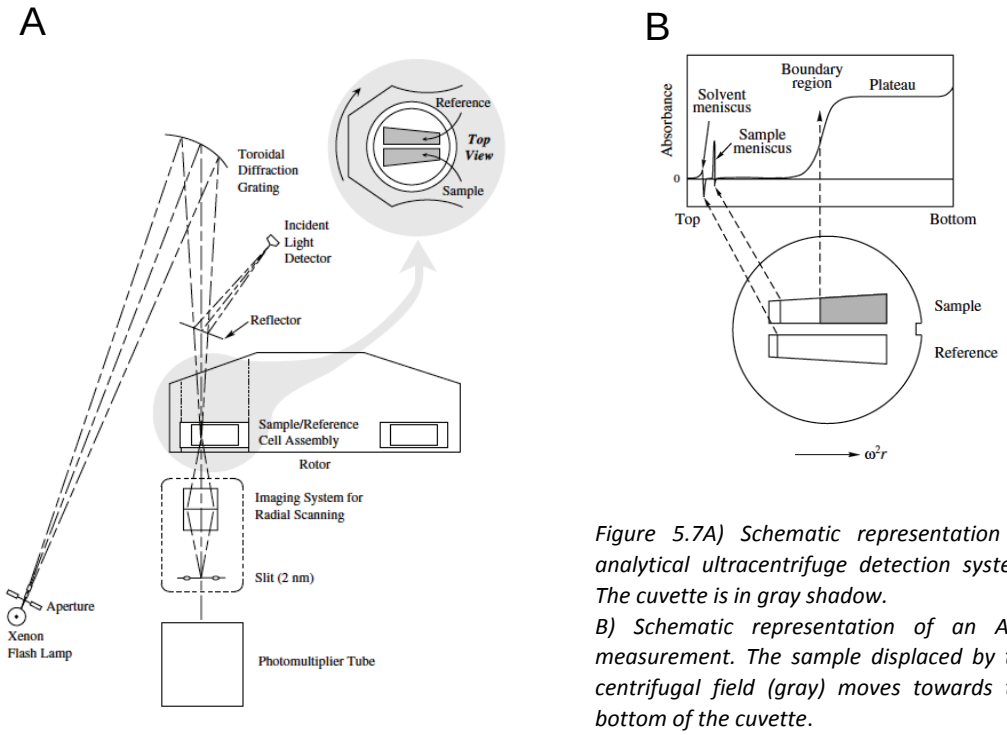


Figure 5.7A) Schematic representation of analytical ultracentrifuge detection system. The cuvette is in gray shadow.
B) Schematic representation of an AUC measurement. The sample displaced by the centrifugal field (gray) moves towards the bottom of the cuvette.

SV-AUC is the more extensively used method and is the one adopted in this thesis. In a SV-AUC experiment, a sample solution is spun in a cuvette at a velocity ω high enough to allow relatively fast sedimentation of solute. The association between variation in concentration *c*, ultracentrifugation time *t*and radial position*r* (which is the position in the cuvette length) is described by the Lamm equation:

$$\frac{\partial c}{\partial t} = -\frac{1}{r} * \frac{\partial}{\partial r} * \left[r * \left(cs\omega^2 r - D * \frac{\partial c}{\partial r} \right) \right]$$

The two values not provided experimentally are *s* and *D* that are respectively the sedimentation coefficient, expressed in Svedberg unit (1 S =10^{-13} s) and diffusion coefficient, already described in 5.7.3. The Svedberg equation describes the relation *s* between the values *D* and the mass *M*:

$$s = M(1 - \rho\bar{v})D/RT$$

The value ρ is the solvent density, while the value $\bar{\nu}$ is the partial-specific volume of the macromolecule. Currently it is possible to obtain from the amino-acid composition and the buffer composition theoretical values of $\bar{\nu}$ and ρ from the software SEDNTERP (http://sednterp.unh.edu/#).

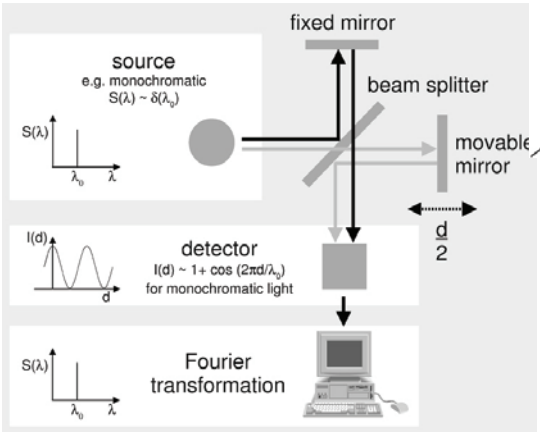
Measurements were performed on Nt25Q and Nt4Q samples freshly disaggregated with an Optima XL-A (Beckman) at 38,000 rpm. Sample displacement profiles were obtained by interferometry and UV absorbance at 280 nm. Data analysis was performed with the program SEDFIT 14.1(296, 297). This program calculates the *s* from the sedimentation profile and provides insights on the oligomeric state by proceeding from some assumptions on the shape of the macromolecule. The best fit among the various simulated values is the one displayed in the analysis.

In case of a monomeric protein, it is possible to calculate the molecular weight by measuring separately SV-AUC and DLS, obtaining *s* and *D*with orthogonal methods. Therefore, the value *s*was obtained from SV-AUC and the value *D* was obtained from a parallel DLS experiment with a DynaPro instrument (ProteinSolutions), while the values ρ=1.000416 g/ml and $\bar{\nu}$=0.7201 ml/g were calculated with SEDNTERP.

By applying the Svedberg equation, it was possible to check whether the MW of the main species observed by DLS and SV-AUC was corresponding to a monomer.

This experiment was performed at the Analytical Ultracentrifugation and Light Scattering Facility at CSIC-CIBio, Madrid.

5.8.3 FOURIER TRANSFORM INFRARED SPECTROSCOPY



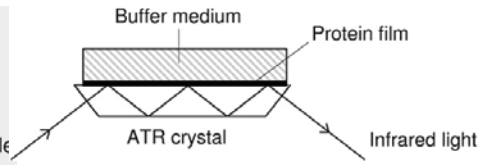


Figure 5.8(Top) schematic representation of a classical ATR setup. The light is reflected repeated times across the crystal and emitted to the detector (Left) schematic representation of an FTIR interferometer. The sample is placed right before the detector, where the two beams (black and grey) collide.

In FTIR spectroscopy, a beam of light containing a wide range of wavelengths is used to analyze the sample, by measuring the absorption of each wavelength. The beam of light is split by a beam splitter and reflects on both a moving mirror and a fixed mirror. The moving mirror determines a difference of optical path with the beam that is reflected by the fixed mirror. The two beams are forced to collide again by a set of mirrors and when they do, they produce interference. This new beam of light is called recombined beam and it is the one that passes through the cuvette containing the sample, before reaching the detector. The interferogram produced is deconvoluted by a Fourier transform in the single components of absorption at each wavelength, represented by a series of peaks in intensity.

The FTIR technique used is the attenuated total reflectance (ATR-FTIR) and it consists in directing the light towards a crystal with an angle of incidence that is entirely reflected from the crystal. The sample is deposited on top of the crystal and absorbs part of the reflected light, so that the crystal emits only attenuated light. The light can be reflected several times on the crystal and when it reaches the detector, it will carry the pattern of absorption of the sample.

ATR-FTIR was applied on aggregated samples of Nt25Q and Nt4Q, after lyophilization. 500 µl samples of Nt4Q and of Nt25Q were incubated at 37°C in sodium phosphate buffer for, respectively, 40 and 50 days and lyophilized.

FTIR measurements were carried out by the technical and scientific service of the university of Barcelona (CCIT-UB), with a scientific NICOLET iZ10 ATR instrument equipped with a diamond detector DTGS.

5.8.4 ATOMIC FORCE MICROSCOPY

AFM is a microscopy technique with a resolution higher than light microscopy, due to its peculiar design. The AFM microscope does not project light directly on the sample, but instead scans its surface with a sharp tip tethered to a cantilever, similarly to a gramophone. Laser light is directed towards the head of the cantilever is reflected with an angle that depends on the position of the cantilever and essentially amplifies the motion done by the cantilever when the tip encounters obstacles or roughness. This system allows appreciating deflections of up to 0.1 nm and it is particularly valuable for studying nanomaterials, like protein aggregates, especially in fibrillar or quasi-crystalline form. The advantage of AFM over other high resolution microscopy technique is its quantitative nature, as the picture formed, in reality, is produced by plotting a series of *xyz* coordinates of the cantilever tip(298).

Aggregated protein samples were prepared of Nt25Q and Nt51Q after respectively 30 and 3 days of incubation in sodium phosphate buffer (Table 5.10) at 37°C.

10 microliters of sample were poured on top of a freshly cleaved mica surface glued on a metallic disc with two-component epoxy cement. The sample was let to absorb for 5 min. After that, the mica was dried under a nitrogen stream.

The method used for the experiment was Peak Force Scan Asyst mode (QNM, Brucker) and it consists in performing the measurement by vibrating the cantilever at frequencies lower than the classical Tapping mode, by applying the force with a sine wave, instead of with a triangular wave. This prevents the cantilever to incur in unwanted lateral contacts (meaning, contacts with the surface while scanning). The ScanAsyst allows to further fine-tune the force applied to the cantilever (299).

The experiments were performed in air with a Multimode 8, Nanoscope V electronics, equipped with an AFM probe SNL-10 (nominal spring constant: 0.35nN/nm, Bruker).

The AFM data collection was performed by the technical and scientific service of the University of Barcelona (CCIT-UB).

5.8.5 TRANSMISSION ELECTRON MICROSCOPY

TEM operates with the same principle of light microscopy, with the only difference that, instead of a light source, a high-energy electron beam (up to 300 kV) passes through the sample and the image formed is due to the thickness and the composition of the material analyzed. The beam projects the image on a CCD or a fluorescent screen.

The advantage of an electron beam, instead of a light beam, is the energy at which the electron can be accelerated, which is orders of magnitude higher than the one of a photon, producing therefore shorter wavelengths and higher resolution. Increases in magnification can be produced by increasing the voltage at which a set of magnets focus the electron beam (300).

As electron beams easily pass through protein layers, the absorption of thin layers of protein aggregates is very low. To increase the electron density of the samples, it is necessary to apply a staining solution that carries heavy atoms. Uranyl acetate is the most commonly used staining solution in biological electron microscopy, as it binds very well to proteins, nucleic acids and membranes (301).

TEM samples were prepared by depositing 6 µl of aggregated protein solution on Cu-carbon coated grids (Agar Scientific) and incubating for 5 min. The excess of sample was removed with filter paper (Whatman). Grids were stained for 5min with 1% (v/v) uranyl acetate for a 5-min incubation.Uranyl acetate excess was removed as for the sample and the grids were washed rapidly with MilliQ water before drying and storage. Images were acquired with a JEOL 1010 electron microscope.

5.8.6 LIMITED PROTEOLYSIS COUPLED TO MASS SPECTROMETRY

The core of a fibril is a very compact environment, protected from the effect of external binders. For this reason, it is possible to obtain information from the residues buried in the core of the aggregate, by degrading it with proteases in a time-controlled fashion. A cleavage site buried in the core is less accessible than one protruding in the medium, especially if the protein is disordered, taking longer to be cleaved. Therefore, by knowing where a specific protease cleaves and exposing the sample to the protease for a range of times, it is possible to rank the cleavage sites for degree of protection.

This experiment is generally applied on folded proteins to identify buried and exposed loops(302, 303). However, by repeating the same experiment with both monomeric protein and fibril, it is possible to know the difference in degree of protection between monomeric and aggregated sample and therefore know what are the buried region(304–306).

The protein solution incubated with the protease is analyzed by mass spectrometry (MS) in order to identify the peptides produced with proteomics software.

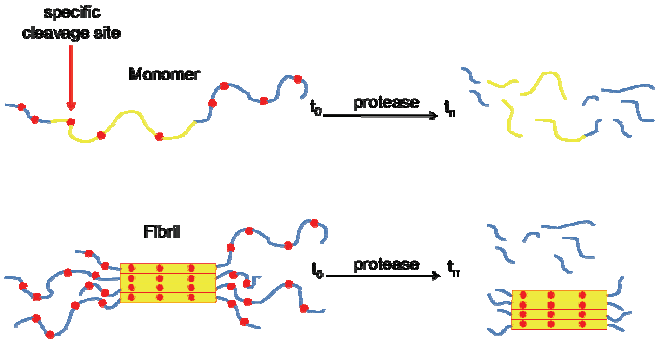


Figure 5.9 Schematic representation of the experimental approach for the limited protease experiment. The cleavage sites not visited by the protease at the time t_n are protected by the aggregate (yellow), those visited both in monomer and aggregate are exposed (blue).

Freshly disaggregated Nt25Q and Nt4Q and a solution of Nt25Q aggregated for 16 h in stirring conditions were split in 50 µl aliquots and incubated with a 1:50 dilution of Trypsin (MS grade, Promega) shaking at 37°C for 0 (t_0), 10, 30 min, 1 h, 3 h and 16 h. The reaction was quenched with addition of 1% formic acid (MS grade, Promega). A separate aliquot was not treated with trypsin (t_{-1}) as a control.

The samples were analyzed by HPLC-MS in the Mass Spectrometry Core Facility at IRB Barcelona in a nanoAcquity UPLC (Waters) instrument, coupled to with a LTQ-FT Ultra mass spectrometer (Thermo Fisher).

MS data were analyzed with Proteome Discoverer v1.3 software (Thermo Scientific) with SEQUEST algorithm, while a prediction of the trypsin cleavage sites was performed with the PeptideCutter software from ExPasy (http://web.expasy.org/peptide_cutter/).

PART IV: RESULTS

CHAPTER 6: PURE RECOMBINANT PROTEIN AND SECONDARY STRUCTURE CHARACTERIZATION

AR is a 100 KDa protein and its dimensions are not compatible with *E.coli*expression. In addition, also a biophysical characterization of the NTD, or parts of it, can be a challenging endeavor, as it is an intrinsically disordered region (Figure 6.1) with a propensity to aggregate.

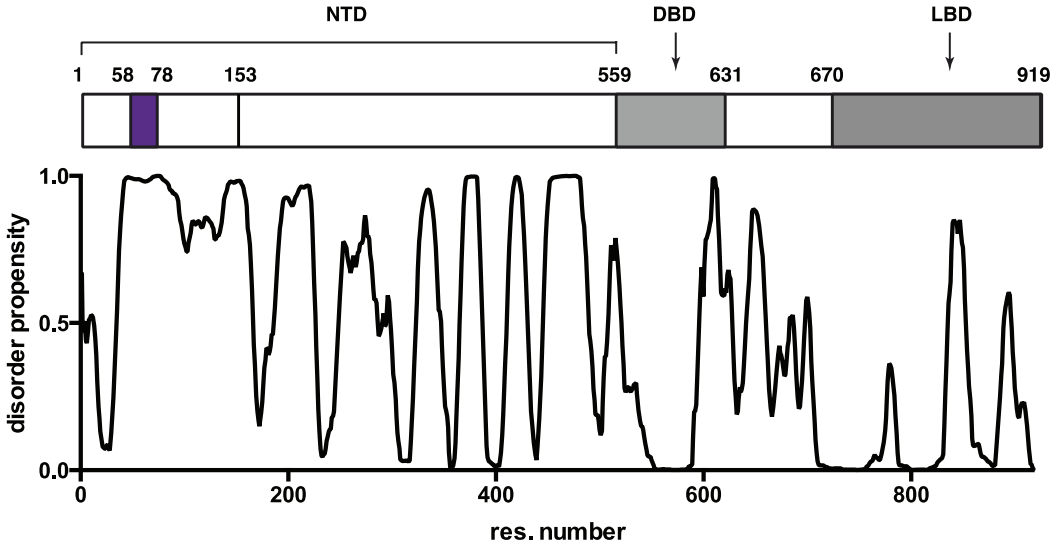


Fig. 6.1 PONDR-FIT prediction of the full length AR with representation of the domains on top. The polymorphic polyQ is shown in purple. The chosen fragment is from fragment 1 to fragments 153 in the PDB sequence. A value of 1 corresponds to completely disordered regions.

The approach we adopted for this project, as for many other polyQ diseases, is to isolate the region that is involved in the oligomerization of AR and to identify the main components that explain this phenomenon.

This chapter describes the process to obtain pure recombinant protein and the characterization of its secondary structure by biophysical and bioinformatics methods.

6.1 CHOICE OF THE FRAGMENT AND DESIGN OF THE CONSTRUCTS

The study of polyQ proteins and their aggregation have been previously carried out using peptides of relatively small size. The well characterized Htt peptide consists of the first 17 residues at the N-terminus of exon 1, followed by the polyQ tract and a polyPro chain of 10 repeats, or Pro_{10}(69), however the huntingtin sequence is far more extended (3142 residues, Uniprot: P42858) and other regions were discovered involved in the initiation of its aggregation(307).

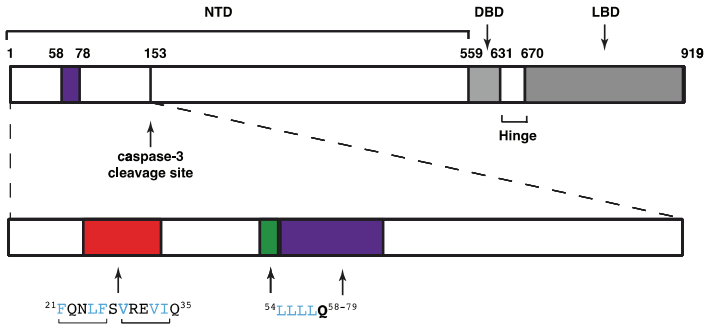
Studies on ataxin-7 have used the first 26 residues at the N-terminus of the polyQ in this protein as a model (86), while others that have focused on the peculiar distribution of His within the polyQ tract of ataxin-1 have used polyQ peptides flanked by Lys(96).

Studies on full length polyQ proteins were performed with ataxin-3, which contains a folded domain, the Josephin domain. This domain is aggregation prone and triggers the polyQ aggregation(99).

In each of these cases, the regions involved in the first oligomerization (whether flanking or far apart from the polyQ) were known, while for AR, little is known about the structural properties of the NTD and about any organization in functional motifs.

It is known that proteolytic events produce fragments of AR in SBMA cells. Caspase-3 has been described to cleave two sites on the AR sequence, the most N-terminal of which is the sequence ^{154}DEDD^{157} and the cleavage is reported between the first aspartate and the glutamate residues. The second cleavage site is less defined and it is located in the hinge region (308). No other cleavage sites are reported for caspases on the NTD to our knowledge and a search with the Expasy software PeptideCutter (http://web.expasy.org/peptide_cutter/) does not identify other motifs recognized by the 10 caspases available.

The fragment cleaved by caspase-3 is the smallest fragment of AR containing a polyQ tract that retains biological relevance, as it has been reported as a cleavage product during the aggregation process(200) and the N-terminal part of the polyQ has been reported in the nuclear aggregates(202).




```
GMEVQLGLGR VYPRPPSKTY RGAFQNLFQS VREVIQNPGP RHPEAASAAP PGASLLLLQQ

        70         80         90        100        110        120
QQQQQQQQQQ QQQQQQQQQQ QQQETSPRQQ QQQQGEDGSP QAHRRGPTGY LVLDEEQQPS

       130        140        150
QPQSALECHP ERGCVPEPGA AVAASKGLPQ QLPAPP
```

Number of amino acids: 156
Molecular weight: 17203.9
Theoretical pI: 6.33

Figure 6.2Top, linear representation of AR full length and the N-terminal cleavage site. The fragment is expanded and the regions of interests are colored. ^{23}FQNLFSVREVIQ^{35} (predicted to be helical and known interacting with LBD in helical conformation(209)*) is in red, the ^{54}LLLL^{58} motif flanking the polyQ is in green, the polyQ in purple. Bottom, primary sequence of the 25Q protein and relative biophysical properties (ProtParam http://web.expasy.org/protparam/)*

During the design of the construct, a TEV cleavage site was introduced (ENLYFQG) at its N-terminus, while the caspase-3 cleavage site was omitted from the sequence. Therefore, the cleaved recombinant protein starts with a Gly residue N-terminal to the canonical Met

residue, as a product of TEV cleavage, and terminates with the residues ^{150}PAPP^{153}, right before the caspase-3 cleavage site.

Once the extension of the construct was decided, it was possible to acquire some preliminary information about its properties and its secondary structure by bioinformatics tools. This would provide initial information on what regions could play a role in the oligomerization of the protein.

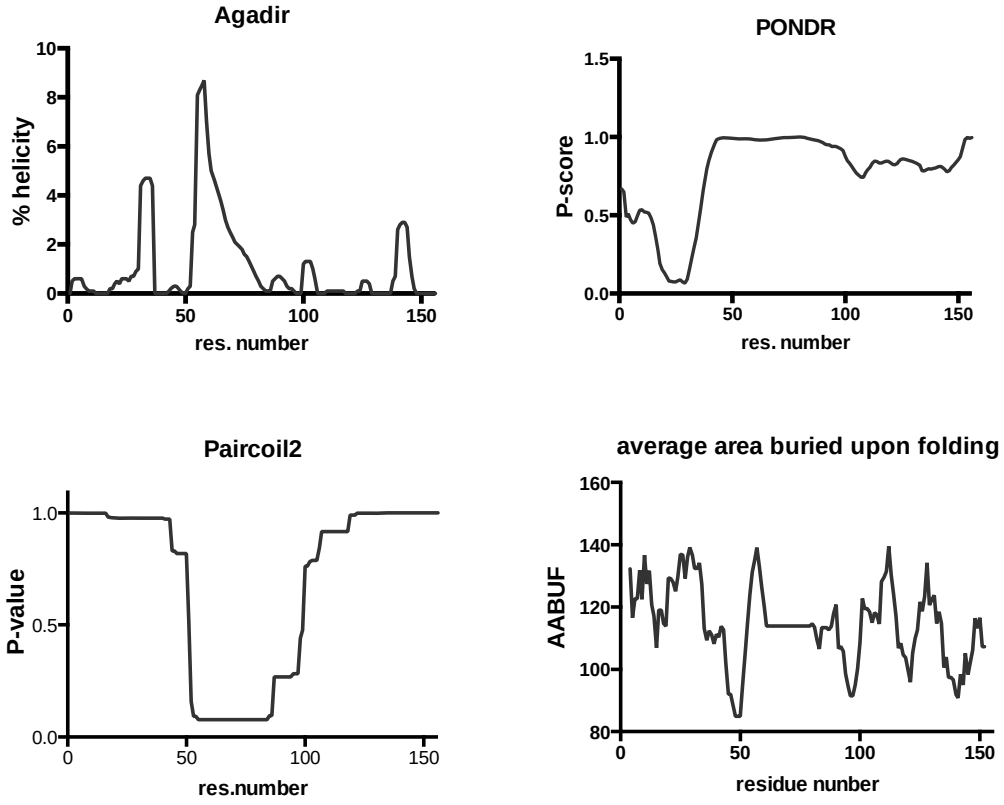


Figure 6.3predictions of AR N-terminal recombinant protein with 25 Gln repeats (25Q). Agadir predicts the helical propensity per residue. PONDR predicts the disorder propensity. A P-score of 1 means completely disordered. Paircoil2 predicts the probability to have each residue in a coil-coil interaction. Low P-values mean higher probability. AABUF calculates the predicted average area buried upon folding (309) *and suggests the composition in hydrophobic residues of the sequence.*

Figure 6.3 shows that the 25Q protein is mainly disordered, except for the region between residues 23 and 27 that correspond to ^{23}FQNLF^{27} motif (mentioned in Chapter 3.4.3), which has been described α-helical conformation when bound to LBDdescribed in a crystal structure(241). This region is flanking at the N-terminus a region (^{28}QSVREVIQ^{35})predicted to be helical, but that is predicted as disorder by PONDR-FIT.

It is therefore possible that this sequence could sample transiently helical conformations, but that this state is not the most stable in solution. The motif ^{23}FQNLF^{27}and its adjacent ^{28}QSVREVIQ^{35} present strong similarities, with hydrophobic residues in position *i, i+3, i+4*, both in ^{23}FQNLF^{27}(Phe_{23},$Leu_{26}Phe_{27}$) and in ^{28}QSVREVIQ^{35}(Val_{30}, Val_{33}, Ile_{34}). Thus, both motifs could adopt a 13-residue long amphipatic helix conformation when folded.

The second region with interesting features is the flanking sequence of the polyQ, ^{54}LLLL^{58}, and the polyQ itself. PONDR predicts this region as disordered, while the prediction of Agadir shows the highest value of helical propensity in the plot across the 4 residues right before the polyQ. This helicity decreases along the polyQ tract.

Consistent with Agadir, Paircoil2 predicts a high probability to form a coil-coil interaction for the region that spans across the first polyQ tract and the second non-polymorphic polyQ tract, formed by 6 Gln residues, separated by the 5 residue sequence ^{78}ETSPR^{83}, determining a predicted coil-coil that covers 1/3 of the entire protein.

6.2 CLONING A POLYQ FRAGMENT

Cloning DNA sequences that contain GC-rich and highly repetitive regions can be a complicated task, due to the number of secondary and tertiary structures that the template DNA can adopt and the high stability that these structures obtain, due to the extra H-bond in the guanosine-cytosine interface(310).

A PCR is the repetition *in vitro* and at larger scale of the DNA replication and highly repetitive regions will reproduce the same behavior that gives rise to genetic instability. We therefore tried to adopt other approaches, in order to minimize the number of PCRsof the wt sequence.

```
Atggaagtgcagttagggctgggaagggtctaccctcggccgccgtccaagacctaccga
 M  E  V  Q  L  G  L  G  R  V  Y  P  R  P  P  S  K  T  Y  R
ggagctttccagaatctgttccagagcgtgcgcgaagtgatccagaacccgggccccagg
 G  A  F  Q  N  L  F  Q  S  V  R  E  V  I  Q  N  P  G  P  R
cacccagaggccgcgagcgcagcacctcccggcgccagtttgctgctgctgcagcagcag
 H  P  E  A  A  S  A  A  P  P  G  A  S  L  L  L  L  Q  Q  Q
cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcaagagact
 Q  Q  Q  Q  Q  Q  Q  Q  Q  Q  Q  Q  Q  Q  Q  Q  Q  E  T
agccccaggcagcagcagcagcagggtgaggatggttctccccaagcccatcgtaga
 S  P  R  Q  Q  Q  Q  Q  Q  G  E  D  G  S  P  Q  A  H  R  R
ggccccacaggctacctggtcctggatgaggaacagcaaccttcacagccgcagtcggcc
 G  P  T  G  Y  L  V  L  D  E  E  Q  Q  P  S  Q  P  Q  S  A
ctggagtgccaccccgagagaggttgcgtcccagagcctggagccgccgtggccgccagc
 L  E  C  H  P  E  R  G  C  V  P  E  P  G  A  A  V  A  A  S
aaggggctgccgcagcagctgccagcacctccg
 K  G  L  P  Q  Q  L  P  A  P  P
```

Figure 6.4 Original AR DNA sequence in the fragment of choice and relative translation. In purple the $(CTG)_3$, in green the polymorphic polyCAG tract. There are 21 repeats.

6.2.1 CLONING AND PCR STRATEGY

After several trials with different polymerase kits, settings and combination of primers, the best result was obtained by applying the AccuPrime GC-rich DNA polymerase, specifically chosen for the high efficiency in replicating GC-rich templates. This kit contains a set of accessory proteins mixed in the polymerase buffer that maintain their activity throughout the amplification reaction and help the polymerase to copy with high fidelity the template DNA. This, in our hands, was the only approach that provided any positive result.

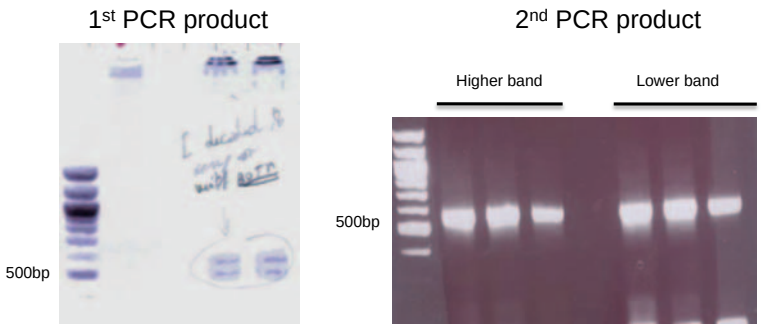


Figure 6.5Agarose gels of the products of the two PCR reactions. All the PCR product was loaded on the gels, split in equivalent lanes. The product of the first, once purified, was used as a template for the second PCR. The purification of the DNA bands from the second PCR was used for the BP reaction for loading the pDONR.

The first PCR was performed using the cDNA of full-length AR gene and two bands were produced in proximity to the 500 bp marker band (Figure 6.5). As there was no possibility to distinguish which of the two bands was the real amplified product, we decided to purify from the gel both bands separately and to use the purified DNA for performing two parallel second PCR. The purified DNA for each reaction was transferred in the entry vector pDONR221 by BP reaction in both cases and finally transformed.

The transformation was successful only for the pDONR221 loaded with the lower band and the 7 colonies obtained were sent for sequencing after gel extraction. Of them, only 4 produced a correct sequencing, however we realized that the number of Gln repeats in each of them was 25.

The template DNA, as well as the sequence deposited in the PDB, counts 21 Gln repeats, meaning that an expansion of 4 trinucleotides occurred during the PCR steps, which can be explained as an episode of simple slippage of the polymerase(60).

However, given the polymorphic nature of the polyCAG, we decided to keep this mutation, as the number of repeats was still non-pathogenic and within the range of common polyQ lengths across the population. The DNA construct and the protein obtained from this procedure were called 25Q.

6.2.2 EXPANSION OF THE POLYQ TRACT

Two methods for expanding a polyQ tract are described in literature: one takes advantage of mismatching primers (311) to expand an already present polyQ tract, while the second inserts a *de novo* synthesized $(CAGCAA)_n$ sequence, produced by the serial amplification of long primers containing repeats of (CAGCAA) units(312).

The first approach would have been more appropriate for our work, but the expansions introduced were producing a too large numbers of repeats. The second approach, conversely, would fit well in terms of size of the expansion to our case, but it was founded on inserting a polyQ tract in a protein that did not have any. The primer used (formed by mixed units of $(CAGCAA)_8$ and relative $(CTGCTT)_8$) were designed to partially overlap one another, without forming tertiary structures, but would anneal extensively on the polyQ tract already present in the sequence of AR, with the risk of producing no expansion.

Our approach was inspired by the one more recently developed involving serial amplifications (312). We designed a protocol to introduce, without the use of PCR, one single long expansion of 30Q, designed as a single forward primer containing $(CAG)_{30}$ and a single reverse primer containing $(CTG)_{29}$. The two strands of DNA would anneal as a double stranded DNA with 4 nucleotides protruding at 5' of the reverse primer and a blunt end on the other side (Figure 6.6).

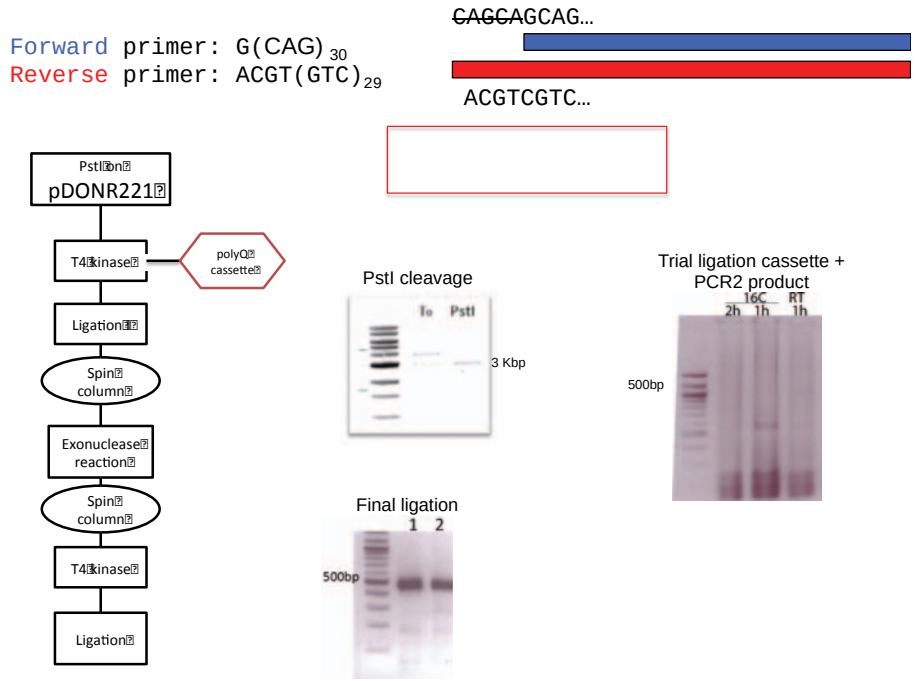


Figure 6.6 Primers for the polyQ insertion, workflow and resultingagarose gels.

The pair of primers for the polyCAG cassette was annealed by decreasing temperature over a long period of time with a PCR machine (98°C to 25°C in 18 h), to minimize the formation of tertiary structures.

The restriction enzyme PstI cleaves the sequence CTGCAG in position 5 (CTGCA'G), producing sticky ends right before the polyCAG (Figure 6.6, upper gel). Therefore we used this enzyme to cleave the pDONR221 plasmid loaded with 25Q, practicing a double strand cleavage in one single position between the codons for Leu_{58} and Gln_{59}.

We introduced the polyCAG cassette and, with a T4 kinase, we added a 5'-phosphate to each strand, in order to allow the ligation by T4 DNA ligase.

In this way, the sticky ends of the plasmid could join those of the cassette on the 5' side, while the blunt end of the cassette is left at 3', with no possibility to be ligated to the other end of the plasmid.

To adapt the 5' sticky end of the plasmid to the 3' blunt end of the cassette, Mung Bean exonuclease was applied, which removes the 5' and 3' protruding nucleotides, producing blunt ends wherever possible.

A new phosphorylation by T4 kinase and a new ligation were performed, to obtain, in principle, a circularized plasmid containing the expanded 65Q gene (Figure 6.6, bottom gel).

In reality, this process was attempted several times, by changing annealing conditions of the pair of primers, temperature and times of ligations, DNA template (parallel trials,showed in Figure 6.6, gel on the right, were performed with the linear dsDNA obtained from PCR2 described in 6.2.1,), exonuclease used (trials were performed with Exonuclease III, that removes selectively the 3' protruding end) and by introducing cleaning steps after every ligation. None of these attempts produced a DNA that, once sequenced, contained the correct expansion.

The reason of this disappointing results lies, very likely, in the fact that the DNA sequence of the polyCAG cassette, due to its highly repetitive sequence, could anneal in too many combinations, producing protruding ends of different lengths, aberrant tertiary structures, or self-anneal. Probably, a design closer to the one proposed by Murphy in the University of Wisconsin(312), after a new round of optimization, would have produced the polyQ expansion.

However, we considered a better investment to purchase a synthesized and codon optimized version of the expanded polyQ recombinant protein with 51 Gln repeats. The advantage of codon optimization for a specific host (*E.coli* in this case) are higher expression yields(313, 314) and, in this case, increased DNA stability, as the polyCAG was substituted by a non-polymorphic polyCAA. This choice resulted advantageous, as the yield of purified 51Q was comparable to those of 25Q, contrary to what could be expected for a non-codon optimized gene.

Given the positive outcome of the codon optimization for 51Q, a second construct was purchased, with a vestigial polyQ tract of 4 Gln (4Q), codon-optimized with the same conditions (Figure 6.7).

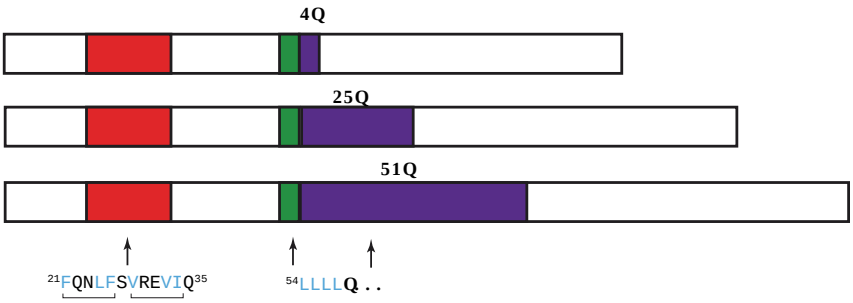


Figure 6.7 The three AR polyQ constructs, namely 4Q (top), 25Q (middle), 51Q (bottom)

6.3 EXPRESSION AND PURIFICATION STRATEGY

Once the challenge of obtaining polyQ constructs with different number of glutamines was overcome, a strategy of purification was necessary. As seen in Figure 6.1 and 6.3, NTD is intrinsically disordered and the NTD constructs under study contain a polyQ tract that makes at least two of them aggregation prone. Therefore, two main complications had to be addressed: fast protein degradation, due to the disordered nature, and aggregation.

A protocol available in literature(271) purifies successfully ataxin-3 by expressing the protein for a short time (3 h) at a relatively low temperature (28°C), whereas ataxin-1 is produced by

expressing the protein fused to the solubility tag glutathione S-transferase and a His_6 tag (His-GST)(315).

In general, expressing difficult proteins fused with folded very soluble polypeptides is a strategy increasingly applied in biotechnology and the mechanisms by which this occurs might change from fusion protein to fusion protein(266, 316–318). In specific, MBP protein has been reported to outperform other solubility tags in maintaining in solution proteins that have the natural tendency to aggregate, like amyloid or polyQ(316).

6.3.1 EXPRESSION VECTOR AND CONDITIONS

The advantage of using a Gateway system is to easily switch from one expression vector to one other, without requiring further molecular biology interventions, by using the BP and LR Clonase kits, as explained in 5.1.1.

Preliminary trials were performed by expressing the 25Q protein fused with a His_6 tag (pDEST17) with a set of expression conditions, by changing temperature (20°C, 25°C and 37°C), concentration of IPTG (0.1 mM, 0.5 mM) and induction time (2 h, 16 h). The SDS PAGE gels relative to these trials are presented in Figure 6.8.

Unsurprisingly, these expression and purification trials had no positive outcome, as the only protein obtained at the end of the purification (last gel, visualized by silver staining) was not of the mass corresponding to 25Q, after a mass spectrometry analysis by HPLC-MS.

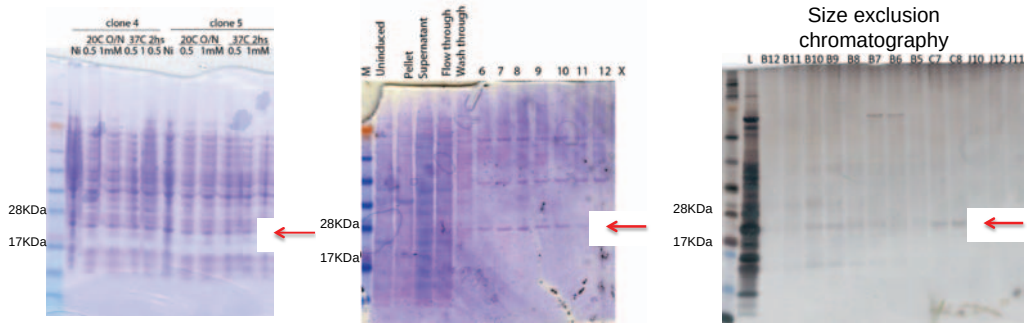
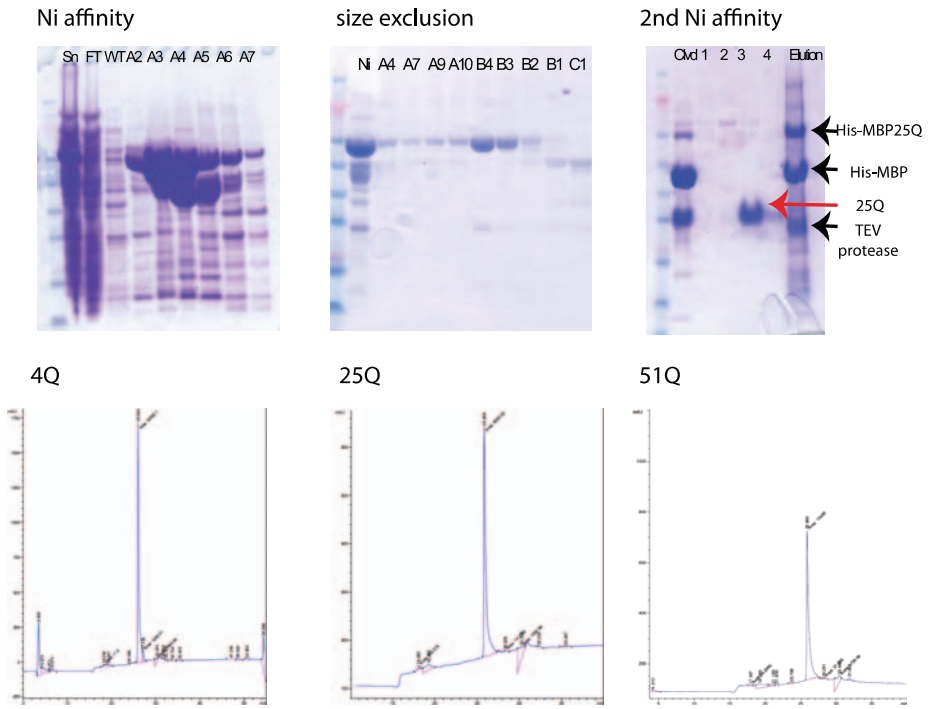


Figure 6.8 Expression (first gel) and purification trials of His-25Q. The second gel is a Nickel chromatography; the third gel is an analytical SEC with a Superdex 75 10/300 GL column as a last step of purification. The band corresponding to the allegedly His-25Q protein is marked with a red arrow.

Our next approach was to express the three constructs 4Q, 25Q and 51Q fused with a His-MBP tag (pDESTHis-MBP) for 4 h at 28°C with shaking. The combination of temperature and time is apparently sub -optimal, as short expressions are generally carried out at 37°C, while long expression (16 h or more) are carried out at lower temperatures than 28°C (generally 18-25°C). Nevertheless, as previously described,(271) this choice is the best tradeoff between high expression rates and low stress for the cells and the consequent release of proteases. Figure 6.9 shows the gel of SDS PAGE electrophoresis for the various steps of expression and purification to obtain pure 25Q.

These conditions provided large amount of protein with substantial purity after the first Ni affinity chromatography. However, a second main band was observed in the first step of purification.

A gel filtration as a second step of purification is generally not advised, due to the low resolution of preparative size exclusion columns, especially compared to a carefully planned gradient in an ion exchange chromatography. However, the ion exchange was not considered suitable for the purposes of this purification, due to the necessity of maintaining the NaCl concentration high (500 mM) and to the presence of glycerol, in order to prevent protein-protein interaction and aggregation. Other approaches led to dramatic decreases in yield.



Protein	Yield per 4 L culture
4Q	900-1200 nmol
25Q	600-900 nmol
51Q	500-600 nmol

Figure 6.9*Top row: SDS PAGE of every step of purification. Left to right: Ni affinity chromatography (the first lane after the marker is the crude supernatant out of expression), preparative SEC, 2nd affinity chromatography in 8 M urea buffer and measurement of final concentration. Central row: HPLC measurements and purity check after the last step of purification 4Q, 25Q, 51Q. Bottom row: average yields per expression (4 L of LB medium).*

The refinement of the purification required maintaining the solubilization tag His-MBP fused to the protein, so the cleavage of the His-MBP was kept to the last step. The pure protein was separated from the tag by applying TEV protease and passed through a final Ni affinity chromatography that retains the His-TEV and the His-MBP, while the flow-through contains pure AR fragment. This step was performed in presence of 8 M urea, in order to prevent aggregation.

Isocyanate ions can be present in urea solutions and can react with N-terminus -NH, Lys and Arg(319) determining chemical modifications of these amino-acids. We took steps to prevent this reaction by working at low temperature and maintaining the protein in urea

exclusively for the time required for the purification. After collecting the pure protein, the solution was dialyzed in a dilution of 1:2000 of MilliQ water before lyophilization and storage.

Therefore, the complete sequence of purification steps consists in a Ni affinity chromatography, a preparative SEC, the tag removal by TEV protease and a final Ni affinity chromatography in denaturing buffer (8 M urea), for separating the pure protein from the tag and the protease.

The quality of each step was checked with SDS PAGE, while HPLC was used to confirm the purity of the final product. HPLC was also used to measure the concentration of the pure protein, as explained in chapter 5.4.

6.4 STRUCTURAL CHARACTERIZATION

The majority of the secondary structure characterization of 25Q and 4Q was carried out in collaboration with BaharehEftekharzadeh and the group of Isabella Felli in CERM, Florence. Circular dichroism (CD) spectroscopy was carried out in IRB Barcelona, while the NMR characterization was performed in CERM.

CD confirmed that 4Q and 25Q proteins are mainly disordered, as described by the minima at ≈200 nm. However, 25Q presents a marked second minimum at 222 nm, which is barely visible in 4Q, and a minimum at ≈200 nm more red-shifted (204 nm) than the one of 4Q (202 nm). These spectra can be compared to the result of the linear combination of a spectrum of a random coil and the one of α-helix, where in 25Q the α-helical component is more represented than in 4Q. Deconvolution of the spectra by using the CONTIN algorithm in the Dichroweb suit calculated the 4Q as 12% helical (NRMSD = 0.11), while 25Q as 32% helical (NRMSD = 0.18)(Figure 6.10).

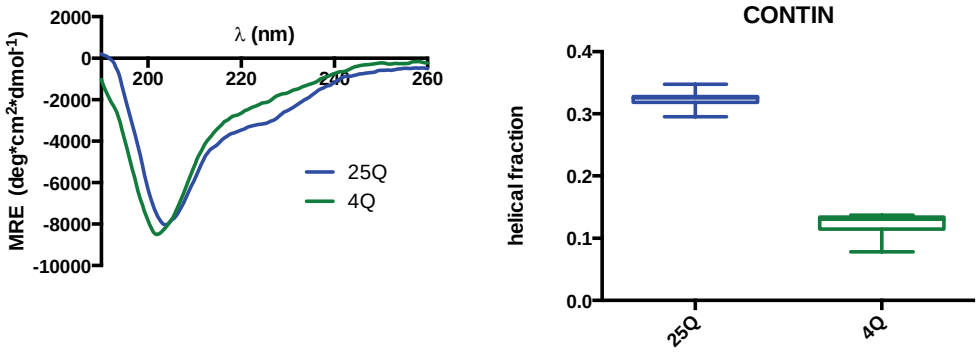


Figure 6.10 CD spectra, and relative deconvolutions, of 25Q 130 µM (blue) and 4Q 140 µM (green).

NMR experiments were performed to acquire higher resolution information about secondary structure at a residue level. ^{15}N- and ^{15}N/^{13}C-labeled proteins were produced and were dialyzed from 6 M urea to Na phosphate buffer. This treatment was adequate in preventing the protein from forming oligomers for a sufficient amount of time to perform the experiments (Eftekharzadeh, Piai et al., under review).

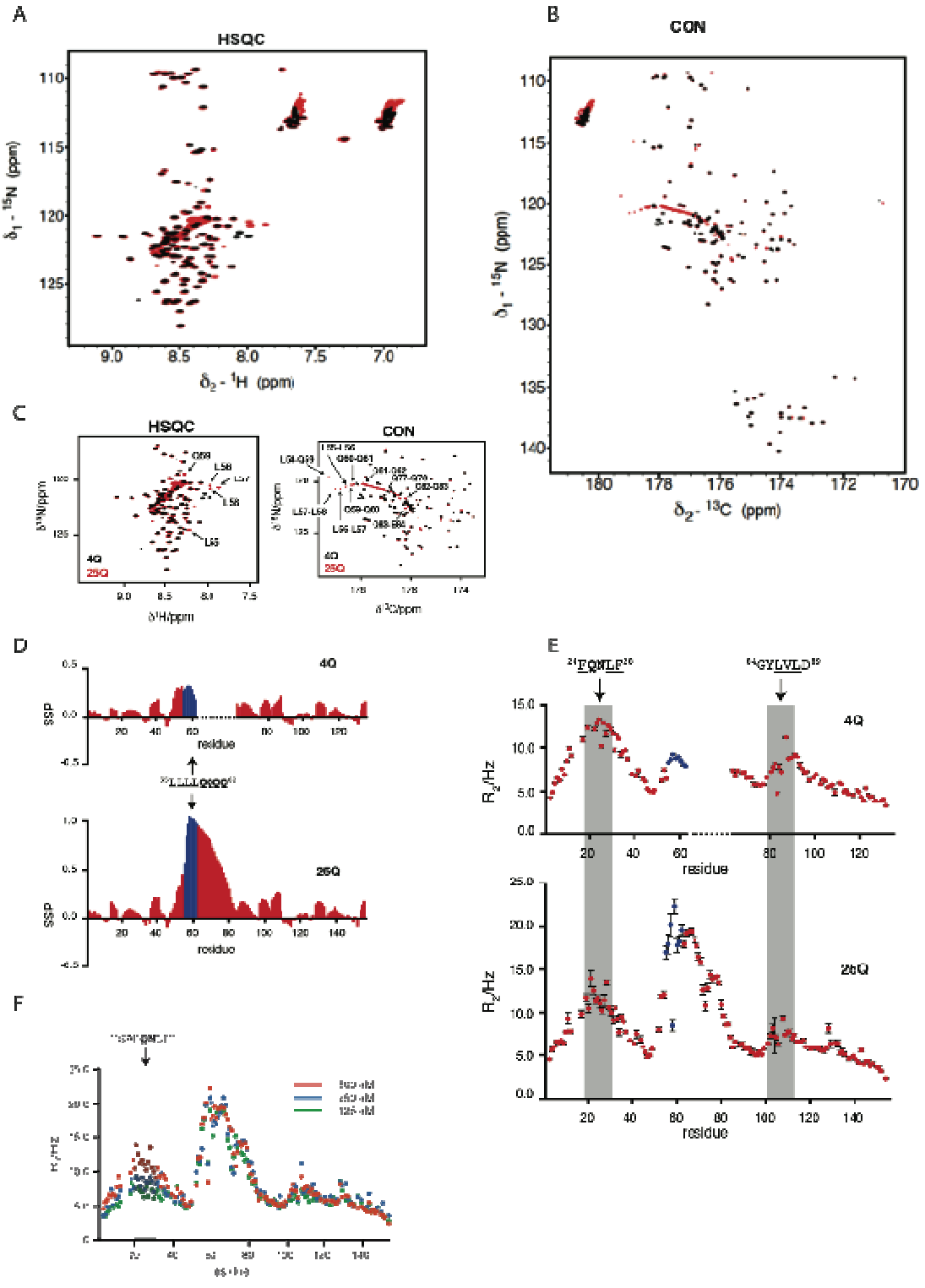


Figure 6.11 NMR experiments with 4Q (black) and 25Q (red). A) HSQC experiments for 4Q and 25Q. B) CON experiments for 25Q and 4Q. For A) and B) full spectrum on the left, highlight of the central part of the spectrum on the right.C) Zoom of the region relative to the polyQ tract both in the HSQC (top) and the CON experiments (bottom)D) SSP analysis of 4Q (top) and 25Q (bottom). E) R^2 relaxation rates reported as a function of residue number for 25Q at 125 µM (green) 250 µM (blue) and 500 µM (red) concentration. In D) 4Q is represented with a gap of 21 residues (corresponding to the polyQ of 25Q) and the region of ^{55}LLLL^{58}, that is more helical in 25Q, is highlighted in blue (EftekharzadehPiai et al., under review).

The assignment of the protein based on the chemical shifts was obtained by performing 4D-transverse relaxation optimized spectroscopy (TROSY) experiments under conditions developed for intrinsically disordered proteins(320).

Figure 6.11-A shows the Heteronuclear Single-Quantum Coherence (HSQC) experiments on 4Q and 25Q, which measure the chemical shifts of ^{1}H and ^{15}N. Figure 6.11-B shows the CON-IPAP experiment, which represents the cross-peaks of the ^{13}CO and the ^{15}N of each peptide bond, to have information about the backbone and to know the positions of the Pro residues, otherwise not visible in an HSQC.

In both HSQC and CON experiments 4Q and 25Q mostly overlap, which made the assignment more straightforward. Only the polyQ, and, interestingly, the N-terminal flanking region presented a marked displacement of the cross-peaks. The peaks associated to the polyQ region of 25Q are organized in a pseudo-diagonal, along which the cross-peaks associated to each Gln residue appear in order, whereas the Leu residues of the ^{55}LLLL^{58} motif were markedly shifted in 25Q respect to those of 4Q.

Analysis of the secondary structure propensity (SSP) for both 25Q and 4Q was performed from their heteronuclear backbone chemical shifts(321). In this analysis, the chemical shifts per residues are computed against the chemical shifts that the same amino acid would adopt as a random coil. Positive values show propensity to form α-helices, negative values show propensity to form ß-sheets.

Again, 25Q and 4Q do not differ one another in secondary structure, both showing motifs with little helical propensity, except for the motif immediately N-terminal to the polyQ, the ^{54}LLLL^{58}. In 4Q this region shows moderate helicity, together with the following 4 Gln residues, while in 25Q the ^{54}LLLL^{58} reaches values compatible with those of a stable α-helix in the region across residues 54 and 62, decreasing towards the C-terminus, to return to values closer to the random coil at the end of the polyQ (Figure 6.11-D).

Direct (R_1), transverse (R_2) relaxation rates and heteronuclear NOE of the backbone amides are quantities that can characterize aspects of the dynamics of a protein backbone. Measuring the R_2 relaxation of amino acids is a common experiment, as it presents larger variations than the other two parameters and is therefore more sensitive in revealing deviations from the random coil. Low values of relaxation are associated to flexibility, while higher values of relaxation are indication of presence of secondary structures, tertiary contacts, local collapse or in general any phenomenon that is associated to an increase in rigidity(322).

Therefore, the backbone dynamics of 4Q and 25Q were studied with the analysis of ^{15}N longitudinal (R_1) and ^{15}N transverse (R_2) relaxation rates. As for the SSP analysis, little difference was found between 4Q and 25Q, except for the values of R_2 relative to the polyQ of the 25Q and its adjacent region ^{55}LLLL^{58} (Figure 6.11-E). Three main regions had high transverse ^{15}N relaxation rates: the motifs ^{21}GAFQNLF^{28}, ^{55}LLLLQQQQ^{62} and ^{87}TGYLVLD^{93} in 4Q, or ^{108}TGYLVLD^{114} in 25Q. Furthermore, the R_2 analysis of 25Q at different concentrations shows progressively higher values of R_2 relaxation with the increase of concentration in the region of ^{22}GAFQNLF^{28} and in the polyQ, which suggests that the 25Q protein forms oligomers by interactions that take place mainly in these two regions (Figure 6.11-F). The ^{21}GAFQNLF^{28} motif also corresponds to a region predicted to have low disorder propensity and that is already known to interact with LBD in AF-2 as an α-helix (209, 241). ^{55}LLLLQQQQ^{62}motif presents higher values of relaxation for 25Q than for 4Q and this

increase can be attributed to the increase in structural propensity observed using SSP analysis in this region. Finally, the third motif corresponds to a region that is relatively abundant of hydrophobic residues and has a modest helical propensity.

In this thesis we focused on studying the role of the motifs ^{23}FQNLF^{28} and ^{55}LLLLQQQQ^{62} in the oligomerization of AR.

6.5 SUMMARY AND CONSIDERATIONS

The lack of knowledge in the mechanism of aggregation of AR in SBMA is, in part, also associated to the technical hurdles involved in the production of recombinant polyQ proteins and the challenges in their characterization. Their coding sequences are difficult to clone and these proteins generally also contain intrinsically disordered regions(54), which, together with the aggregation propensity of the polyQ tract, can provide expression and solubility issues.

Furthermore, the ID nature and the aggregation propensity of these proteins prevent characterization using X-ray crystallography experiments. Therefore, more indirect techniques are required for exploring their conformational space.

Nevertheless, in this first chapter of results, we have characterized the fragment of AR produced by the cleavage of caspase-3. The coding sequence for this fragment has been successfully cloned and recombinant protein of three polyQ lengths were developed. A protocol for expressing and purifying these three proteins with yields high enough to sustain the requirements of the research project was developed and this facilitated the characterization of the secondary structure.

CD experiments suggest a correlation between polyQ length and degree of helicity in the protein, while NMR data, not only supports this hypothesis, but also demonstrate that the polyQ itself adopts a α-helical conformation, together with its N-terminal flanking motif ^{55}LLLL^{58}.

PolyQ peptides have been widely described as random coil in solution both from CD spectroscopy (72, 78, 79, 323), and from NMR(82, 83, 98)while polyQ fibrils are described as ß-sheet structures(81).

In recent years it has been suggested that polyQ polypeptides can adopt secondary structures other than random coil and ß-sheet., This is based on simulation studies(75, 76, 80) and solid state NMR experiments, where the polyQ of Htt peptide is observed in a PPII helix conformation when in the presence of the Pro_{10} motif at its C-terminal(324, 325).

In this case, the ^{55}LLLL^{58} motif, adjacent and N-terminal to the polyQ, suggests to have a critical role in keeping the polyQ in anα-helical conformation. This effect extends at least across the first 10-15 residues of the polyQ, as the tendency to form helices decreases along the tract towards the C-terminus.

The data suggest cooperativity between the ^{55}LLLL^{58} motif and the polyQ tract. A longer number of repeats determines an increase in helicity, not only in the polyQ tract, but also in its N-terminal flanking region ^{55}LLLL^{58} motif.

A well accepted theory describes the polyQ proteins as naturally involved in coil-coil interactions (103, 194, 326) and both the high helicity of the polyQ and the prediction by Paircoil2 of high propensity to form coil-coil interactions fit well in this scenario, however this data do not directly support it.

The second finding, which will be followed up in the next chapters, is the presence in the 25Q protein of another region that shows self-interaction from the R_2 relaxation analysis, the ^{21}GAFQNLF^{28} motif. This region has a prediction for moderate helicity by Agadir, however the SSP analysis of chemical shifts does not highlight presence of stable helices. Nonetheless, this motif is already known to acquire a helical conformation from a crystallographic structure that shows the FQNLF peptide bound to LBD as a helix accommodated in the hydrophobic pocket of AF2(209, 241).

Altogether, these findings pointed the direction of study for characterizing the oligomerization and aggregation of this protein, as the next chapters will explain in details.

CHAPTER 7: MONOMERIC PROTEIN AND METHODOLOGIES FOR KINETIC MEASUREMENT

As in any reaction, differences in aggregation rate can be evaluated only when the starting point is the same. One of the major challenges in measuring kinetics of aggregation is indeed obtaining a starting point that is reproducible and identical over different sample preparations.

The literature is rich in examples of methods tailored to start a kinetic measurement with identical and reproducible conditions. A very straightforward approach is to keep the protein fused to a highly soluble protein, like MBP, and remove the tag by applying TEV protease to the sample at the beginning of the measurement(327). However, this system assumes that the protease would cleave simultaneously all the protein in solution. Most importantly, the MBP-fused protein can potentially interact with other monomers, forming oligomers that would nucleate the aggregation right after the cleavage of the protein, while other proteins are still fused to the MBP. This would produce a complex and non-homogeneous environment, and is highly dependent on the quality of the TEV protease, the amount of TEV protease added, the time of incubation, with the added risk of spurious interaction with the MBP or the TEV and of the unspecific activity of the protease itself.

Therefore, we explored two protocols to obtain monomeric protein after purification. The first uses an organic solvent, hexafluoro-2-propanol (HFIP) and a strong organic acid, trifluoroacetic acid (TFA), to dissolve the protein and shield it with charges(328). The second dissolves the protein in a potent chaotropic agent, guanidiniumthyocyanate(GndSCN), close to saturating concentration (6 M) and then the monomer is purified by gel filtration chromatography (31).

Parallel to the disaggregation, another challenge encountered was the development of reliable assays for measuring kinetics of aggregation. Thioflavin-T (ThT) is a dye that works very well with Aß peptide, but is much less reliable when applied to other systems. We adopted other biophysical tools, like sedimentation assay by ultracentrifugation and dynamic light scattering to characterize the aggregation mechanism of the AR recombinant fragments with different polyQ lengths.

The development of these tools and the development of a disaggregation protocol for the AR protein occurred in parallel,but were closely related. For the sake of clarity, this chapter describes, first, the process to obtain a monomeric sample and how we validated the method developed and, second, the development of the methods for kinetic measurements.

7.1 PRELIMINARY EXPERIMENTS IN ABSENCE OF DISAGGREGATION

A set of preliminary experiments was performed to measure the aggregation rate of 25Q in the absence of any treatment, directly after dialysis. The samples were frozen at the end of the purification in buffer and used directly, without any further modification. The techniques used were ThT binding monitored with a spectrofluorimeter and SDS PAGE. The measurement was performed in stirring for the whole duration of the evolution and the amount of protein required made it possible to perform measurements in triplicate.

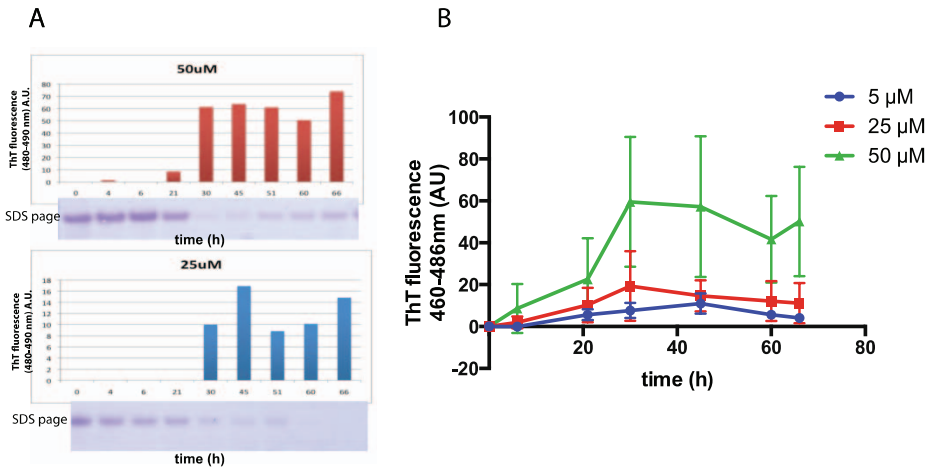


Figure 7.1Preliminary kinetic experiments with 25Q. A) ThT binding against SDS page after tabletop centrifugation of the samples (picture under each table). Each band corresponds to a column in the plot in the ThT binding. B) ThT binding of 25Q at three different concentrations, 5 µM (red), 25 µM (green) and 50 µM (blue).

Whilst promising, these data were biased by the fact that the protein was not in monomeric conditions at the t_0 of the experiment, and thus little could be deduced in terms of mechanistic description. Furthermore, the reproducibility between one set of experiments and another was very low. For all of these reasons, we decided to disaggregate the protein before every experiment.

7.2 DISAGGREGATION PROCEDURE: WETZEL METHOD

This method was developed to efficiently disaggregate polyQ peptides and other highly aggregation prone peptides (huntingtin, ataxin-7)(86, 87, 89, 329). It should be noted that other methods to dissolve amyloid peptides exist, such as using dimethyl sulfoxide (DMSO) solution(330), transient treatment with volatile organic solvents(331), or by ultracentrifugation in aqueous solution(332).

Each of the mentioned methods presents advantages and potential drawbacks and generally the decision of which one to apply depends on the requirements of the peptide and on the experimental conditions. As this project is focused on polyQ protein aggregation, we focused on procedures specifically used for disaggregating polyQ peptides.

TFA is a strong organic acid that extensively protonates the protein, but alone it fails to dissolve polyQ peptide with a higher number of repeats(328). Mixing it with HFIP completely dissolves the peptides, due to the physico-chemical properties of this organic solvent. HFIP is a strong inductor of α-helicity, even stronger than trifluoro-ethanol (TFE)(286) and results as a good solvent for hydrophobic residues than water, surrounding the protein molecule and exposing the fluorine when mixed in aqueous solution.

According to the Wetzel method, 1-2 mg of dry protein were incubated with a 1:1 solution of TFA:HFIP proportional to the amount of protein to obtain a 1mg/ml solution. The time of incubation was adapted to maximize the recovery of protein in solution (Figure 7.3). After incubation, the solution is evaporated with a stream of Ar and incubated in a desiccator for 16 h in order to remove entirely the HFIP(Figure 7.1). The presence of traces of HFIP may

interfere with the outcome of the experiment(284, 285), introducing variability or artifacts, so it is critical in this step to completely remove traces of solvents.

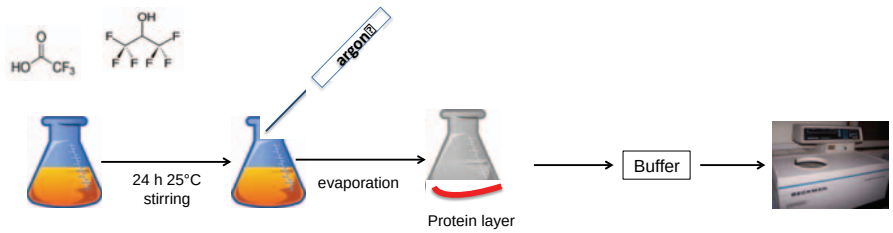


Figure 7.2Representation of the steps in the Wetzel method. The protein is dissolved in a solution of TFA and HFIP, incubated for 24 h at 25°C in stirring and evaporated for 24 h first with a jet of Ar, then with a desiccator. The protein layer is dissolved in buffer and ultracentrifuged for 1 h at 4°C and 386000 g.

The success of the procedure and the integrity of the freshly dissolved protein were monitored with HPLC measurements of the supernatant after an ultracentrifugation step of 3 h (328).

In our hands, the best conditions met were an incubation time in TFA:HFIP of 24 h in stirring with an evaporation time of 24 h, divided between evaporation with the stream of Ar and 16 h incubation in a desiccator. The protein layer was then dissolved in aqueous solution. We realized that aggregation buffer (pH 7.4, described in 5.6.1) worked better than 0.1% TFA water pH 3, as originally described in the Wetzel method.

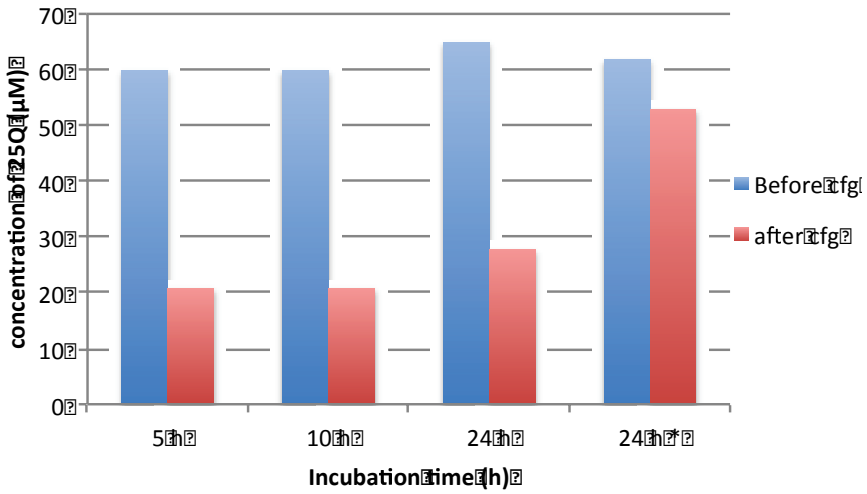


Figure 7.3 Yield of recovery from TFA:HFIP incubation after volatilization of the solvent. HPLC measurements were performed before and after the ultracentrifugation step after different incubation time and the same volatilization time. The samples were dissolved in 0.1% TFA water pH 3.0, except for the sample 24 h, that was dissolved in sodium phosphate buffer.*

7.2.1 VALIDATION OF THE DISAGGREGATION PROTOCOL

To evaluate the efficacy of the Wetzel method, we compared the first set of experiments performed in absence of treatment with a new set of analogous experiments performed on treated protein

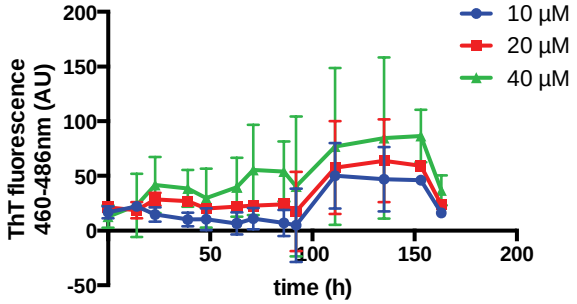


Figure 7.4ThT binding of 25Q at three different concentrations after TFA:HFIP treatment, 10 µM (blue), 20 µM (red) and 40 µM (green).

In ThT binding of freshly disaggregated 25Q, measurements at three different concentrations present a high signal-to-noise ratio and differences among concentrations were difficult to appreciate.

Samples at 10 µM and 20 µM are similar to each other and do not appear to show high intensities in fluorescence. Further, the increase in fluorescence could be also due to particle scattering, than to real birefringence. Later in this chapter, the lack of efficiency in binding to the ThT is discussed (Section 7.4).However, despite the low quality of these preliminary results, the treatment causes a clear delay in aggregation, as shown by qualitatively comparing Figure 7.1-B and 7.4.

7.2.3 PROTEIN DEGRADATION ISSUE

Constant problems with low yields of protein out of the TFA:HFIP incubation led us to closely analyze the protein quality after the treatment. Unsurprisingly, HPLC chromatograms showed more peaks in close proximity of the main peak of the protein, clear sign of protein degradation (Figure 7.5).

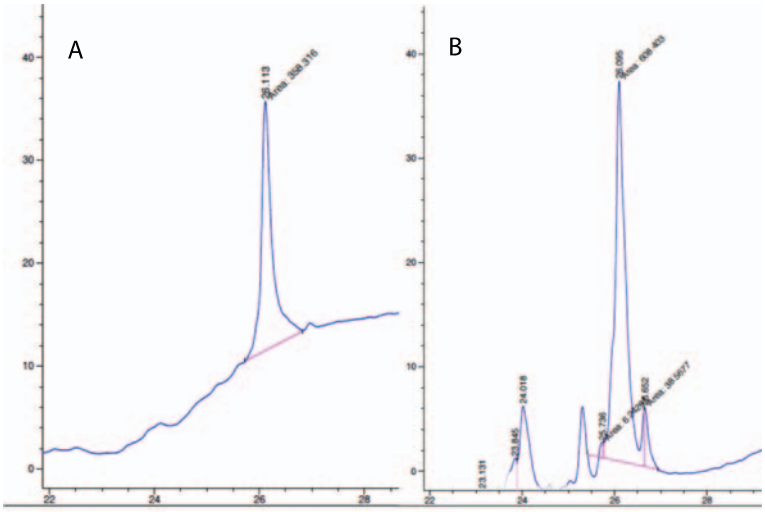


Figure 7.5 A) *25Q 5 µM in H_2O. B) 25Q after lyophilization, incubation in TFA:HFIP and dissolution in sodium phosphate buffer. The satellite peaks are attributable to degradation and make the measurement of concentration more imprecise.*

Lyophilization was discarded as a possible cause of degradation, by checking the quality of the lyophilized protein after dissolving the dry protein in 8 M urea. The total amount of time spent in presence of TFA and HFIP, between incubation and volatilization, is 48 h at RT. In (328) the possibility of peptide degradation is taken in consideration, therefore the long incubation necessary can be the cause of the protein degradation. This phenomenon makes the measurement highly imprecise, both due to the difficulty to find boundaries between the peaks and by introducing uncontrolled amount of different protein fragment. After different attempts of optimization to reduce the protein degradation, we decided to abandon this approach and disaggregate the protein with other methods.

7.3 DISAGGREGATION PROCEDURE: LINSE METHOD

This disaggregation system was successfully applied for measuring the aggregation kinetics of Aß with a plate reader by ThT binding assays(25, 33). The protocol takes advantage of the strong chaotropic agent guanidiniumthyocyanate (gndSCN)(322) to disaggregate and produce an homogeneous solution of monomeric protein(31) To entirely remove the chaotropic agent, a double step of gel filtration is performed, combining a desalting gravity column (PD-10) and a size exclusion chromatography (SEC). This second step allows also separating the oligomers of lower size from the monomer.

In order to adapt this method to 25Q and 51Q, a preliminary choice of denaturing agent was performed, as well as of type of size exclusion chromatography, to better resolve the monomer from the oligomers.

GndHCl successfully dissolved 4Q and 25Q (Figure 7.6-A), but failed to dissolve 51Q (Figure 7.6-B). To address this, the more potent GndSCN(333) was adopted for the disaggregation. The procedure successfully disaggregated 51Q and also revealed that the disaggregation of 25Q with 6 M GndHCl was not complete, as emerges from comparing Figures 7.6A and 7.6C. Therefore, incubation with 6 M GndSCN was applied to every protein (Figure 7.6-C) and the incubation time was decided as described in 7.3.1.

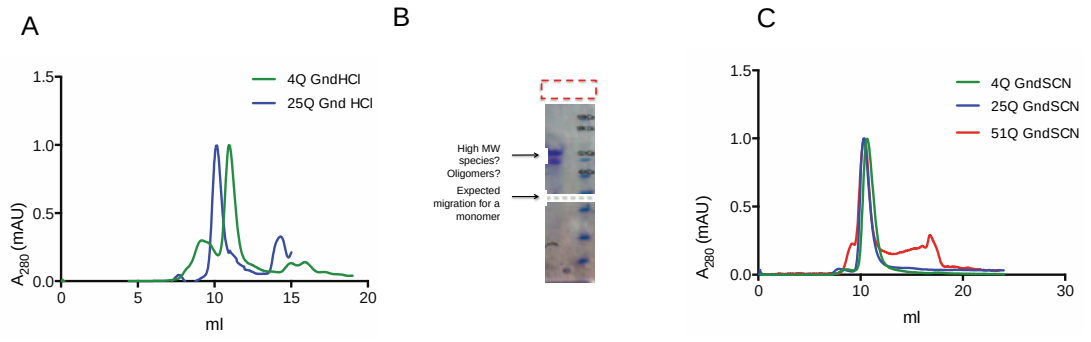


Figure7.6A) SEC samples of 4Q (green) and 25Q (blue) after incubation 3.5 h in 6 M GndHCl and PD-10. B) SDS-PAGE of 51Q incubated in 6 M GndHCl for 3.5 h. The upper bands correspond to high MW oligomers SDS-insoluble of 51Q. The red box highlights the pellet of aggregated 51Q that does not enter the gel. C) SEC samples of 4Q (green), 25Q (blue) and 51Q (red) after incubation 3.5 h in 6 M GndSCN and PD-10. A Superdex 75 10/300 was used for all the SEC and the intensities are normalized by concentration for clarity.

Repeated measurements of the rate of aggregation to validate this process initially failed to show reproducibility, due to the presence of small oligomeric species within the monomer peak. This problem is not encountered in case of studies on Aß, due to the small size of the Aß peptide. Like in our case, the Linse group use in their procedure a Superdex 75 10/300(31), but Aß has a MW of 4.5 KDa, which retention volume (14 ml) is larger than that of the AR fragments, so that a better resolution of the monomer from the oligomers is possible. A direct way to address this would be to change type of Superdex resin; however, trials with a Superdex 200 column did not improve the reproducibility. A critical step in this direction was to implement ultracentrifugation as last step of the protocol, before the beginning of the experiment. This step required a careful characterization, as explained in 7.3.2.

Overall, the whole approach in its final version is summarized in Figure 7.8 and it is composed of a disaggregation in 6 M GndSCN for 3.5 h, a buffer exchange with a gravity column and a SEC. The fractions associated to the monomer undergo ultracentrifugation and their supernatant is the material used for the experiments described in this thesis.

It is important to remark that each of these steps carries a recovery of, in average, 30% of protein; therefore high amounts of material are required for every experiment, of which the purely monomeric protein used for the experiments will count less than 10% of the applied material. A relative advantage is that most part of the discarded material can be recovered, pooled and lyophilized again for a new disaggregation.

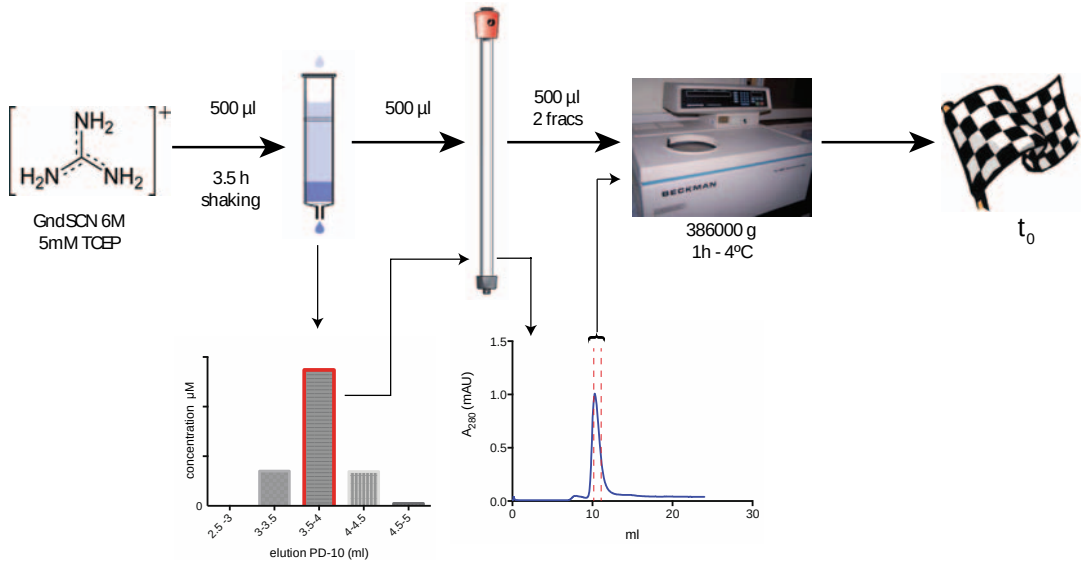


Figure 7.8Linse method adapted to kinetics for AR. The protein is dissolved in sodium phosphate buffer containing GndSCN 6 M and 5mM TCEP and it is incubated until complete reduction (3.5 h). The solution is then transferred into a desalting column and the most concentrated 500 µl sample (measured by Nanodrop) is used for a SEC with a Superdex 10/300 column. The fractions of the peak corresponding to the dashed line are transferred in ultracentrifuge tubes. The ultracentrifugation is of 1 h at 386000 g and 4°C and the upper 2/3 of solution are separated from the pellet and used for the experiment.

7.3.1 INCUBATION TIME IN 6 MGNDSCN

The choice of chaotropic agent has been already discussed, but the conditions of incubation were also critical for a successful disaggregation of the sample. Interestingly, 25Q, as well as 4Q and 51Q, when not entirely reduced, are eluted in the RP-HPLC asmain peak at 26 min of retention time, with a second pair of peaks displaced towards the beginning of the gradient, at retention time 19 and 21 min.

A non-reduced sample was measured by HPLC and then reduced with 5 mM TCEP. Longer incubations in presence of TCEP lead to a complete elimination of the peaks at 19 and 21 min RT.

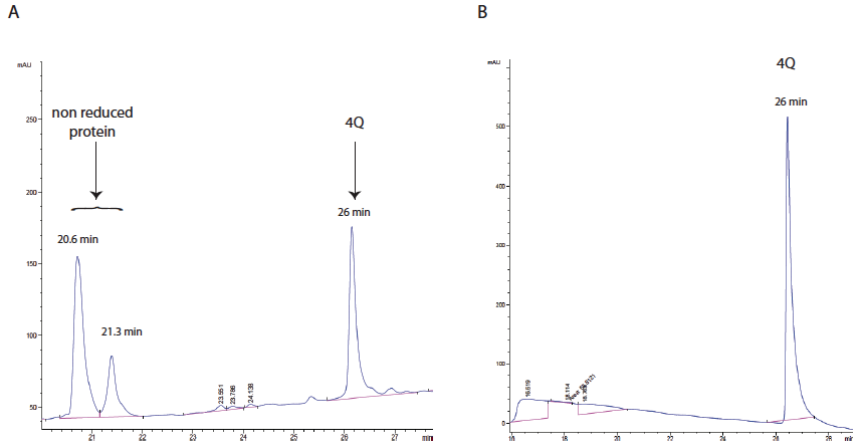


Figure 7.94Q in GndSCN before (A) and after (B) 3.5 h of incubation with 5 mM TCEP. The two peaks at20.6 and 21.3 min disappear in the reduced sample.

We established experimentally that the time for complete disappearance of the peaks at 19 and 21 min was of 3.5 h and we considered this incubation time as sufficient for entirely disaggregating and reducing the protein sample. Incubation times of up to 16 h in shaking did not further enrich in monomer the protein incubated, as SEC performed after 3.5 h and 16 h of incubation present similar profiles.To further purify the monomer an ultracentrifugation step was necessary.

7.3.2 CHOICE OF THE ULTRACENTRIFUGATION TIME

DLS is particularly efficient in detecting even low populations of oligomeric species and it was used for monitoring the efficiency o the ultracentrifugation step in obtaining a solution of pure monomer.

The DLS instrument used provides a set of values produced by the deconvolution of the autocorrelation function. The z-average (D_{hz}) is the weighted mean hydrodynamic size of the ensemble of particles in solution and it is the primary and most stable parameter produced by this technique. The polydispersity index (PDI) is another value that can vary between 0.2 (sample perfectly monodisperse, or a solution of particles of only one size) and 1 (sample highly polydisperse, or characterized by particles of very different sizes), without accounting for the dimension of the particles and it is calculated from a simple 2 parameter fit to the correlation data. Altogether, they can provide a global view of how the sample is pure from oligomers.

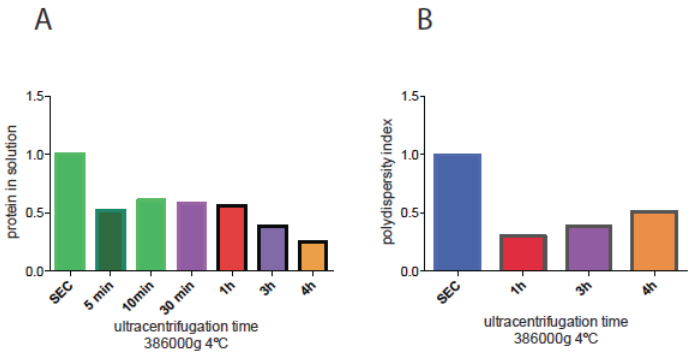


Figure 7.10 *ultracentrifugation tests. A) Protein in solution after ultracentrifugation at different time. B) PDI after ultracentrifugation at different time. Longer centrifugations after 1h do not improve the monodispersity*

However, is also important to take in account that monomeric proteins and small oligomers in aqueous solutions have very similar sedimentation coefficients and therefore a clear separation between low MW assemblies, like small oligomers, and monomeric protein is impossible. Only using solvents of higher or increasing viscosity, like sucrose gradients, would allow this kind of separation, but this approach is unviable for our experimental setup. Nevertheless, it is possible to perform ultracentrifugation steps at a speed that would determine the sedimentation of also part of the monomer, in order to entirely sediment the oligomers and modulate the time of ultracentrifugation to find the best trade-off between purity of the monomer and protein in solution.

Therefore, the concentration of the supernatant after ultracentrifugation was monitored by HPLC, while the oligomeric state by DLS (Figures 7.10-A and B respectively). The best trade-off between polydispersity and protein remained in solution was considered to be 1 h of ultracentrifugation at 386000 *g* and 4°C. This increase in polydispersity can be explained with the fact that a low population of small oligomers is in equilibrium with the monomer, while the monomer keeps being pelleted, so that the overall amount of monomer decreases, while a small amount of oligomer is constantly renewed in solution.

To further prove that the protein obtained was purely monomeric, we performed a series of validating experiments.

7.3.3 VALIDATION OF THE DISAGGREGATION PROTOCOL

Native gels of samples at different stages of the process showed that the band, associated to the low MW oligomer, decreases of intensity from the PD-10 to the SEC step and disappears at the ultracentrifugation step (Figure 7.11-A).

Similarly, the z-average of the sample decreases at every step of purification (Figure 7.11-B). In an ideal monodisperse solution, the z-average and the size of the monomer are equal. Therefore, the progressively lower z-average is demonstration of increased purity of the sample from oligomers. The final step (t_0, in green) has a value (D_{hz}= 13 nm) very close to the one of monomeric 4Q (D_{hz} = 6.7 nm, Figure 7.11-C).

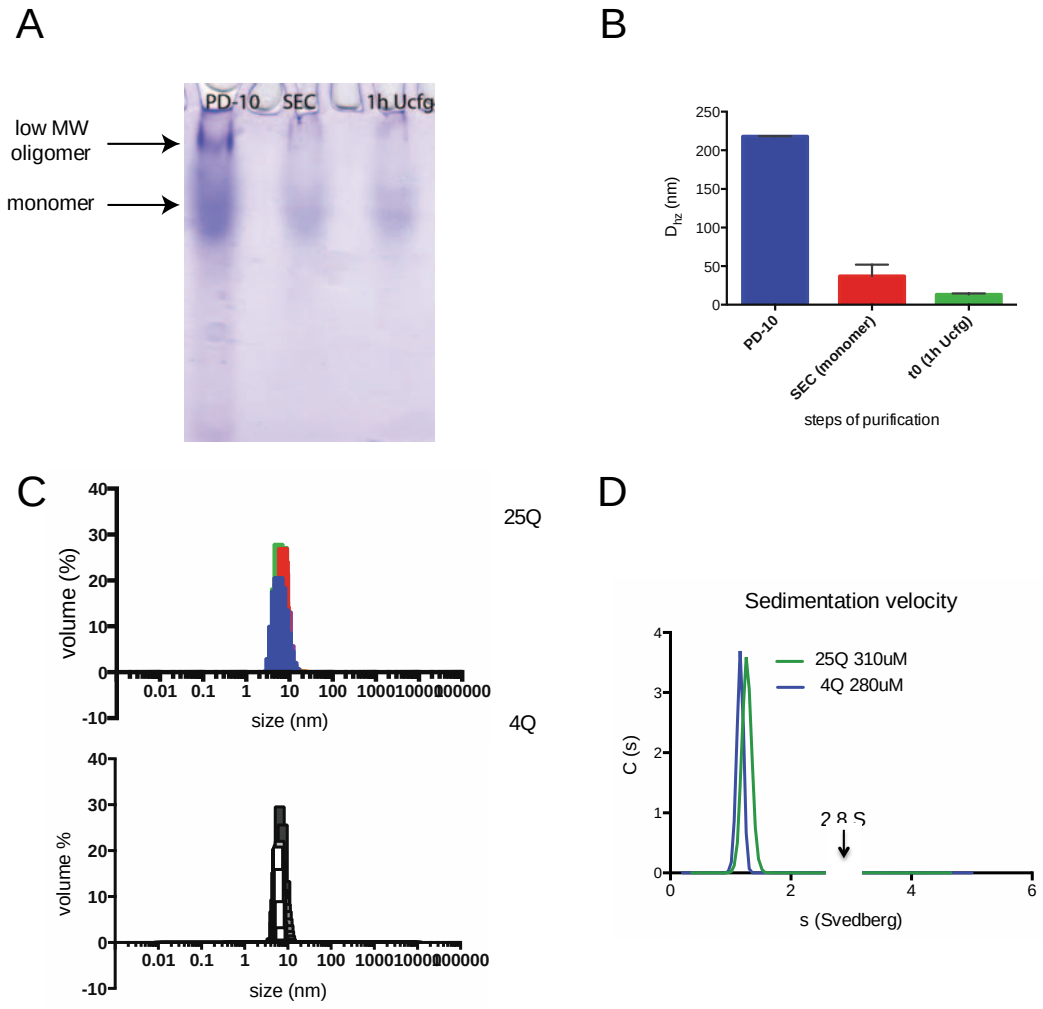


Figure 7.*11 Validation of the disaggregation protocol. A) native gel after the three steps of purification. B) final settings for the preparation of monomeric protein. The z-average decreases at every step, to approach the value of monomeric protein. C) DLS measurement showed in volume performed for 4 different disaggregation for both 25Q and 4Q. D) SV-AUC experiment for 310 µM 25Q (green) and 290 µM 4Q (blue)*

Figure 7.11-C represents a collection of DLS measurements of 4 independent experiments both for 25Q (top) and 4Q (bottom) where the size distribution is visualized by volume, which is the type of deconvolution that better approximates the real distribution in the sample. The size distributions highly overlap across the various experiments and the low populations of oligomers detected are never higher than 1%.

The final evidence that the protocol produced essentially pure monomeric protein is provided by sedimentation velocity (SD) AUC (Figure 7.11-D). This experiment was performed with different concentrations, up to the highest possible for both 25Q and 4Q to test the robustness of the method.

AUC separates species of different sedimentation coefficient (S) measured in svedberg (S) and the fringe analysis identified monomeric 4Q as S= 1.1 S and monomeric 25Q as S = 1.3 S.

S was used in association to the diffusion coefficients produced by DLS data to calculate the MW of the particle analyzed (see materials and methods, Chapter 5.7.2). With this data, it

was possible to univocally determine that the main species observed with SEC, DLS and AUC is actually a monomer.

However, both for 4Q and for 25Q, a low population (2-3%) of oligomeric species was observed, with sedimentation coefficient of respectively 2.4 S and 2.8 S, with no significant difference in the distributions between the two proteins.

By using the software tool SEDFIT, we predicted the size of the oligomer, by simulating different theoretical S, knowing the MW and the S of the monomer. The best fit was for a compacted tetramer. Interestingly, a similar result was found for huntingtin peptide(97), where a freshly disaggregated $htt^{NT}Q_{10}K_2$ (the N-terminal peptide of huntingtin with 10 Gln residues) is analyzed by SD-AUC and, together with a preponderant monomeric species, a main oligomeric species is found, whose S fits well with a compacted tetramer, plus other less abundant species with predicted MW comparable to an octamer and a 12-mer.

Altogether, these data prove that the Linse Method, modified to meet the requirements of the AR recombinant protein, produces a solution of protein highly monomeric in a reproducible way. The complete elimination of the oligomer is impossible, without a drastic decrease of monomer available, which would put even more experimental constraints to the system, and therefore we considered this approach robust enough to be used in repetitive kinetic experiments.

7.4 THIOFLAVIN-T BINDING ASSAY

The most classical system to measure kinetics of aggregation in the amyloid protein field is the ThT binding assay, especially for Aß studies. Therefore, we tentatively applied this method at first, but we soon realized that no ThT binding was occurring for any of the recombinant proteins analyzed, even when the presence of fibrils was confirmed with TEM micrographs (Figure 7.12).

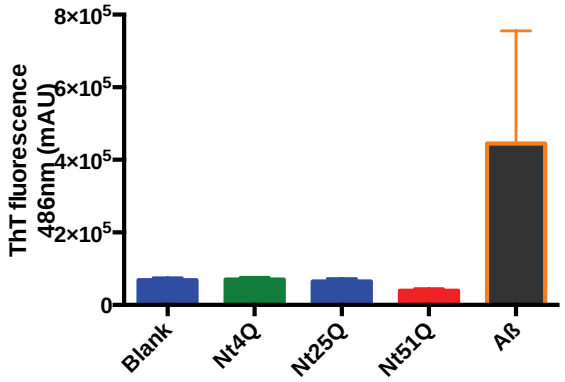


Figure 7.12 *ThT binding assay of 20 µM 4Q (green), 25Q (blue) and 51Q (red), incubated respectively for 2 months (4Q, 25Q) and 10 days (51Q). The presence of fibrils for 25Q and 51Q was confirmed by TEM micrographs. 15µM Aß fibril was used as a control and kindly gifted by the group of Natalia Carulla.*

Difficulty in binding ThT by polyQ proteins has been previously mentioned (102, 271, 334) and it is possible that polyQ proteins form fibrils that do not allow the ThT molecule to intercalate. Conversely, ThT has been observed to bind to fibrils of huntingtin peptide(327) in a quantitative way, while the mentioned observations of low or inefficient binding are

associated to experiments with ataxin-3 protein(334, 335) or with a model system formed by a polyQ tract artificially inserted between two folded domains of a model protein(102).

7.5 SEDIMENTATION BY ULTRACENTRIFUGATION ASSAY (SUA)

To measure the evolution of the AR recombinant proteins, a different approach to ThT binding was considered, based on physical, instead of chemical properties.

The sedimentation by ultracentrifugation assay (SUA) is a method extensively used in the Wetzel group (69) to monitor the fraction of monomeric peptide in solution over the time frame of the experiment. It consists in repetitively taking samples from a solution of peptide or protein in incubation and applying an ultracentrifugation step. The concentration of peptide remained in solution is measured by HPLC, a plot is produced with the change of concentration of supernatant against time(278).

From the characterization of the procedure of disaggregation, we knew exactly what fraction of monomer is present when 386000 g of centrifugal force are applied for 1 h at 4°C. This setting was therefore maintained for the measurement of the concentration of monomer at every time point.

To verify if we could reproduce the quality of the data in literature, experiments with $K_2Q_{25}K_2$ peptide were performed (Figure 7.13-A), which produced an exponential decay, similar to those shown in ref.(86).

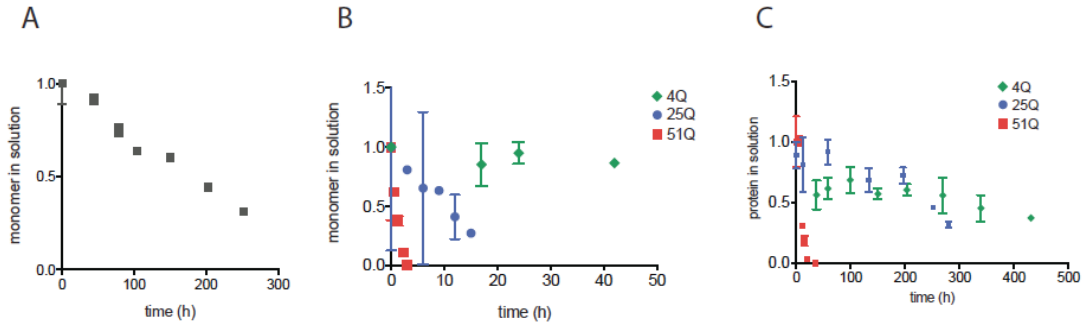


Figure 7.13 Sedimentation assays by ultracentrifugation. A) $K_2Q_{25}K_2$ 43 µM in quiescent condition. B) Sedimentation in stirring condition of 30 µM 4Q (green), 20 µM 25Q (blue), 10 µM 51Q (red). C) Sedimentation in quiescent condition of 4Q (green), 25Q (blue), 51Q (red) all at 20 µM concentration.

Initially, kinetic experiments on the AR recombinant proteins were performed in stirring conditions at 37°C (Figure 7.13-B) and confirmed the polyQ-dependence of the aggregation rate. However, this experimental setting did not allow appreciating the lag phase of aggregation and the rapidity of the reaction did not allow any kind of characterization other than rate measurement.

Changing the experimental conditions to incubation in quiescence determined a longer evolution and a modest improvement in the quality of the measurement that however remained less accurate than kinetic measurements on the $K_2Q_{25}K_2$ peptide (Figure 7.13-C).

In this set of experiments, the polyQ dependence is maintained only between 25Q and 51Q, while the evolution of 4Q and 25Q are hardly distinguishable, pointing that also 4Q, in spite of not having a real polyQ tract, sediments in these conditions (Figure 7.13-C).Whereasin stirring conditions (Figure 7.13-B) the relationship between length of polyQ and aggregation rate is clear, this correlation is less clear in quiescent conditions (Figure 7.13-C).

This observation suggested that a second event of nucleation might occur independentlyfrom the polyQ-driven aggregation and that these two events might follow different rates. This complexity in the reaction could also explain the low accuracy experienced when measuring the AR recombinant proteins. These findings will be described in the next chapter.

7.6 TIME-RESOLVED DYNAMIC LIGHT SCATTERING

Dynamic Light Scattering (DLS) has been already successfully applied to detect low amounts of oligomers in this project. The z-average value provides information on the average size of the particles in solution and we hypothesized that this value changed in time with the increase of the overall size of the aggregates. Therefore we monitored the feasibility of the experiment by performing preliminary tests with a protein known to be stable in solution.

Commercial hen-egg white lysozyme (HEWL) was a convenient standard, as it does not form any higher molecular assemblies as long as its structure is preserved. Partial or complete denaturation by chemicals or heat do result in the formation of fibrils(40, 41, 336), but in physiological conditions the protein is stable and monodisperse. In these conditions, the only type of assembly reported for HEWL in solution at RT is a homo-dimer (337).

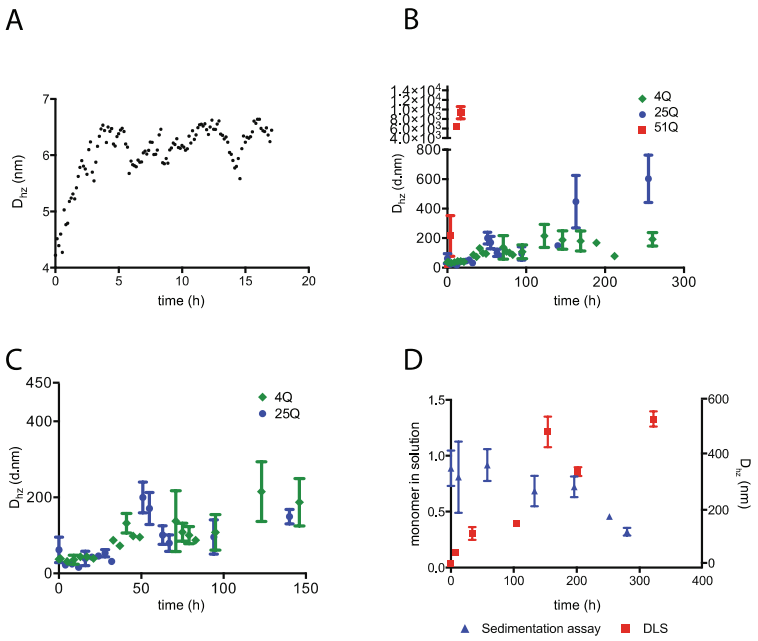


Figure 7.14 Time-resolved DLS. A) DLS experiment with 27 µM HEWL after 1 h ultracentrifugation at 386000 g at 4°C. B) DLS experiment of freshly disaggregated 4Q (green) 25Q (blue) and 51Q (red) at 20 µM. C) Magnification of B, focused on the first 150 h for 4Q and 25Q. Differences in aggregation rate are hardly visible. D) Comparison of the experiments of sedimentation (blue) and DLS (red) for 20 µM 25Q.

The aim was to determine if incubating at 37°C a sample directly in a sealed DLS cuvette would produce noise in the measurement, due to convection or generation of bubbles. Figure 7.14-A shows the preliminary experiment with 26 µM HEWL after an ultracentrifugation step analogous to the one applied to AR recombinant proteins. The protein is monodisperse and monomeric right after ultracentrifugation, but In 4 h at 37°C equilibrates to form a dimer, as the z-average value doubles, and then remains stable with very little fluctuation for the following 16 h.

The experimental settings were then adapted to the AR recombinant proteins and a set of experiments with the same conditions of the sedimentation assays in quiescence was performed to compare the techniques.

Figure 7.14-B shows the time-resolved DLS experiment for 4Q, 25Q and 51Q. The aggregation rates of the three proteins shows polyQ-dependence, with 51Q aggregating much faster than the other two. 25Q and 4Q show initially similar rates of aggregation that later diverge towards 150 h (Figure 7.14-C).

We hypothesized that this difference cannot be observed by sedimentation assay, because the two proteins oligomerize with a similar rate and the oligomers formed are equally depleted from solution so that the second slower process that emerges at late stage, which is polyQ-dependent, is not observable, as suggested by Figure 7.14-D.

Nevertheless, this system resulted the more reliable in our hands in revealing differences among proteins studied and we applied alternatively sedimentation by ultracentrifugation and time-resolved DLS along the study.

7.8 SUMMARY AND CONSIDERATIONS

Achieving exact and reproducible measurements of the kinetics of protein aggregation is a major hurdle in aggregation studies. This obstacle is overcome by various approaches, both from the perspective of sample preparation and that of measurement.

The two latest and most reliable methods for obtaining monomeric protein were developed by Ronald Wetzel's teamat the University of Pittsburgh(328) and by Sara Linse's group at the University of Lund(31).

The method described by Wetzel et al. is the latest in a series of chemical approaches to disaggregate amyloid fibrils. Such methods use organic solvents, like DMSO(330) or HFIP(331), or combinations of HFIP and TFA(338) and combine this method with a purification of the monomer by ultracentrifugation(332). The resulting protocol is divided into a disaggregation step, volatilization of the solvent, and an ultracentrifugation step.

Adapting the Wetzel method to the AR fragments caused a number of complications, the most important of which was considerable protein degradation. Consequently, we abandoned it in favor of the Linse Method.

This other approach takes advantage of the chemical properties of the strongest known denaturant(339), GndSCN, at saturation.Despite the lack of knowledge about the

mechanism of actionused by Gnd^+, this molecule is thought to interact preferentially with the protein backbone instead of water surrounding the protein molecule, whereas the SCN^- ion forms hydrogen bonds with the amides(333). A size exclusion chromatography step is used to purify the monomer. This protocol has been successfully applied to measure the kinetics of amyloid peptides, especially Aß(34, 340, 341).

We adapted this method to disaggregate the AR fragments, but the relatively low resolution of the chromatographic step did not fully separate the monomer from smaller oligomers and a further ultracentrifugation step was implemented.

The efficiency of the entire protocol was tested with DLS, native gels and AUC experiments. We confirmed that the species finally purified is a monomer and that it coexists with a not eliminable minor species (< 3%) that we recognized as a compact tetramer.

This disaggregation protocol,at the expense of a considerable loss of material, allowed us to perform kinetic experiments with a set of tools that had to be in turn optimized.

The most extensively used technique for measuring the formation of amyloid fibrils is the Thioflavin-T binding and this system has been used very efficiently, especially with the Aß peptide(34). Nevertheless, we observed that 4Q, 25Q and 51Q bind very weakly ThT, at any state of the aggregation. The observation that polyQ proteins bind weakly ThT is reported also in other works(102, 334, 335). In these studies,the polyQ tract is hosted by proteins with a large region that do not take part of the fibril and might interfere with the binding. It is also possible that the combination of amorphous aggregates and fibrils (as will be described in the next chapter) would decrease the available binder for the ThT.

Other experimental approaches were adopted. The first one was directly taken from the Wetzel group, as it is extensively applied in measurement of kinetics of peptide with different polyQ lengths(278). It is referred to as sedimentation by ultracentrifugation assay (SUA) and measures the concentration of monomer after an ultracentrifugation step in different points during the aggregation.

We optimized this method to obtain in the supernatant only monomeric protein and we observed a clear polyQ-dependence of the aggregation rate between 25Q and 51Q. We also found that 4Q tends to aggregate despite the short number of Gln and the rate of aggregation was hardly distinguishable from that of 25Q.

In parallel, we developed a time-resolved DLS measurement method, which is very efficient in revealing even low populations of oligomer. Also, this method has been effectively used in different studies, both on polyQ proteins(327, 334) and on other amyloid or aggregation prone proteins(342, 343). In this method, the buildup of the macromolecular assembly is reported as the weighted average size of all the particles in solution, or z-average (D_{hz}).

With this system, we confirmed the polyQ dependence over the three AR recombinant proteins, however the curves for 4Q and 25Q diverge only after 150 h. This data explains why these two proteins have similar aggregation rates according to the sedimentation assay, as they have very similar rates of formation of lower-size oligomers, whereas higher assemblies are formed at different rates.

All of these observations suggest the presence of two phenomena of nucleation, only one of which is polyQ dependent. Interestingly, also the difference in kinetics experiments carried out in stirring and quiescent conditions can be explained in light of the independent

nucleation events. It is possible that the non-polyQ oligomerization forms assemblies that are easily disrupted in stirring conditions, allowing only the polyQ aggregation to occur.

The characterization of these two nucleation processes and the relationships between them will be discussed in the next two chapters.

CHAPTER 8: INDEPENDENT PATHWAYS OF AGGREGATION IN AR

As previously described in Chapter 7, the aggregation rates of 4Q and 25Q were hardly distinguishable in the sedimentation by ultracentrifugation experiments. Also in the time-resolved DLS experiments, the early evolution of 4Q and 25Q was similar, but the z-average value of 4Q reached a plateau after 150 h, while the z-average value of 25Q increased for the whole time span analyzed. These observations suggested that another aggregation process occurs in 25Q and 4Q, independent of the aggregation caused by the polyQ tract.

We hypothesized that the second region involved in the oligomerization is ^{24}FQNLFQSVREVIQ^{35}. This hypothesis was based on the higher transverse relaxation rates obtained in the concentration-dependent NMR experiments (See Figure 6.11-F) and on the Agadir and PONDR predictions (see Figure 6.3). According to the predictors, this region is predicted to have both helical propensity and propensity to form an ordered structure. In addition, this region is characterized by a pattern of hydrophobic residues that recalls that of an amphipaticα-helix, like those involved in coil-coil interactions (typically characterized by hydrophobic residues at positions *i, i+3,i+4,i+7*).However, the SSP data do not show high helical propensity in this motif.

Altogether, this data suggest that the ^{24}FQNLFQSVREVIQ^{35} region can acquire α-helical conformation and nucleate oligomerization by an interaction between amphipathic helices, resembling the coil-coil, but less organized.

The polyQ-dependence of the aggregation rate is well documented in literature(78, 86, 344), as well as the contribution of flanking regions(54, 97, 100). However, less well characterized is the contribution of regions away from the polyQ. Ataxin-3, for instance, is described to contain a folded domain (the Josephin domain) with marked aggregation propensity and capable of forming amyloid fibrils also in absence of the polyQ tract(98).

The experiments presented in this chapter are aimed to characterize these two independent aggregation processes, to understand what forces are driving them and what are the key residues involved.

8.1 AR FORMS AMYLOID-LIKE FIBRILS WITH A POLYQ DEPENDENT RATE

Transmission Electron Microscopy (TEM) micrographs taken at different times showed that 4Q, 25Q and 51Q aggregate by forming amorphous assemblies that evolve into fibrillar species with a rate that depends on the length of the polyQ tract (Figure 8.1). 4Q does not evolve into fibrils for the entire time-span, 25Q forms fibrils only after a long incubation (41 days), whereas 51Q is entirely fibrillar within 6 days and fibrils are already visible within the first day of aggregation, immediately upon completion of the sedimentation by ultracentrifugation experiment. Nevertheless, also in 51Q is possible to observe amorphous aggregates in the TEM micrograph relative to the first day, while no amorphous aggregates are found at 6 days.

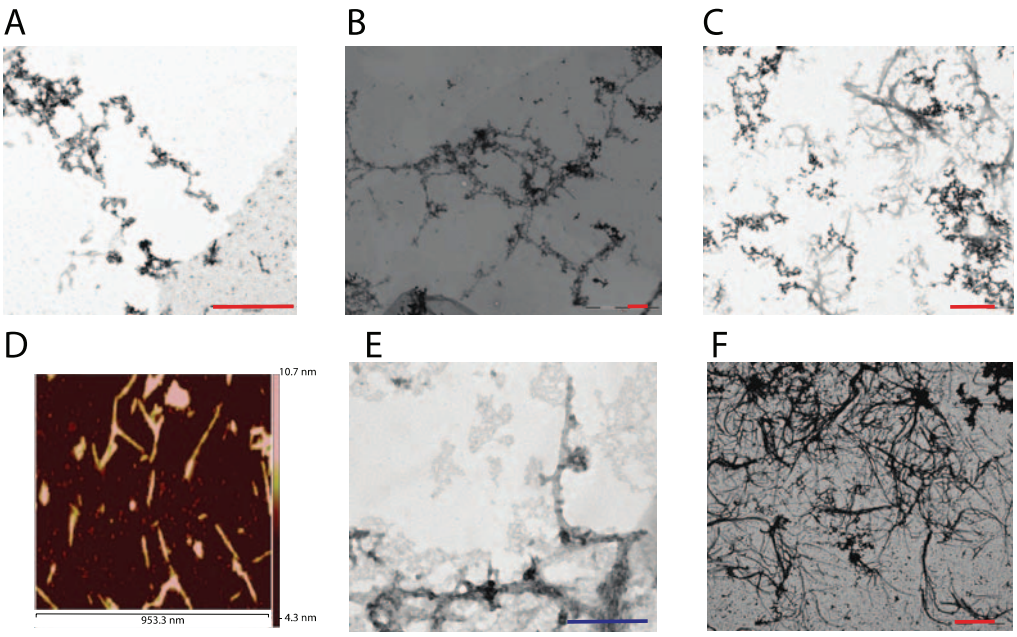


Figure 8.1 TEM micrographs of 4Q after 60 days (A), 25Q after 14 (B) and 41 days (E) and 51Q after 0.5 (C) and 6 days (F) and AFM image of 51Q after 6 days (D). The red line correspond to 500 nm, the blue line correspond to 100 nm.

The quality of the fibrils obtained by the incubation of 51Q allowed clear imaging by AFM. Fibrils obtained were of 3-500 nm long and 5-6 nm thick in average, which are dimensions compatible with those of classical amyloid fibrils(345). The AFM images obtained match well, both for morphology and for size, also with others obtained from polyQ huntingtin peptide(346).

Interestingly, both the TEM and the AFM images report morphologies that have been found in fibrils extracted from mouse tissues, where the amorphous aggregates are found in non-phenotypic animals, while fibrils are found extensively in degenerated muscles and spinal cord of animals of an SBMA mouse model (Eftekharzadeh, et al. unpublished data).

8.1.1 FTIR ANALYSIS OF LATE STAGE AGGREGATES

Aggregates of 4Q and 25Q were incubated for40 and 50 days respectively and TEM micrographs were taken to confirm the presence of fibrils in 25Q and of aggregates in 4Q. The aggregating solution was lyophilized and sent for FTIR analysis to the CCIT-UB spectroscopic facility.

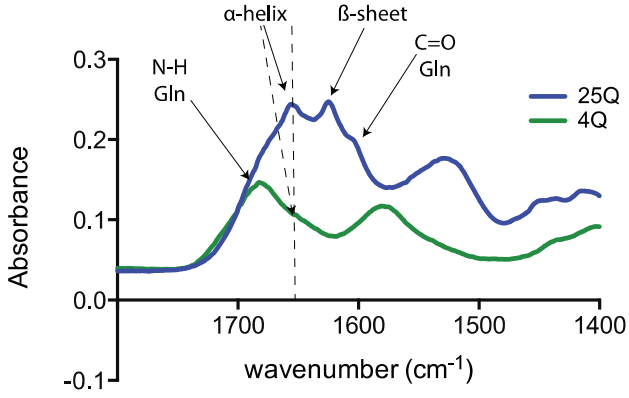


Figure 8.2 FTIR second derivative for 4Q (green) and 25Q (blue). The signal associated to secondary structure properties and those associated to the side chain of glutamine are marked with arrows.

Study of the bands obtained from the second derivative of 4Q and 25Q spectra was complicated by the strong signal associated to the N-H and to the C=O bonds of glutamine residues (Figure 8.2), but is possible to identify clear differences between the two spectra (other than the signals related to glutamines). 25Q presents higher α-helical content than 4Q, but also a strong signal associated with ß-sheets. The bands were assigned according to ref.(347). The presence of a band at wavenumbers relative to ß–sheet for 25Q and their absence in the 4Q sample, suggests that the 25Q fibrils are in ß-sheet. However, the presence of a band at a wavenumber relative to α-helix for both samples suggests that not the whole 25Q protein is part of the fibril.

8.1.2 ONLY THE POLYQ AND ITS CLOSE SORROUNDINGS ARE PART OF THE FIBRIL

Limited proteolysis was performed in 25Q to understand which regions of the protein are more buried in the fibril and what parts are exposed. Trypsin cleaves with high specificity at the C-terminal of Lys and Arg residues and, with lower efficiency, when these residues are followed by a Pro (Figure 8.3-A).

However, proteases cannot access regions that are sterically hindered, like regions in the core of a protein or when they are part of aggregates or fibrils and thus are protected from enzymatic cleavage. Therefore, a protected region will be cleaved less efficiently by the protease. If the cleaved sample is studied by mass spectrometry, is possible to identify what fragments are produced by the proteolysis and reconstruct which regions were exposed to the effect of the protease. With increasingly longer times of incubations, also regions initially difficult to cleave will be affected. In this way it is possible to establish a degree of protection, expressed as the time required for observing peptides produced from a cleavage event on this site.

This system is generally applied in structural studies of protein that are hard to crystalize, to obtain low-resolution information on buried loops of the protein(302), and has been used for mapping the regions of protein buried in the core of amyloid fibrils(304–306, 348).

In this case, 25Q was aggregated in stirring conditions for 24 h, to produce a homogeneous solution of fibrils and digested with trypsin with a range of incubation times spanning from 5

min to 3 h. A monomeric 25Q was treated in the same conditions, as a reference and the degrees of protection were compared. Peptides were recognized by mass spectrometry, matching the masses encountered in the analysis with the theoretical masses of the cleavage products (Figure 8.3-B). The degree of protection (expressed in minutes of incubation with trypsin), is defined as the time of the first observation of a fragment produced by the cleavage site analyzed.

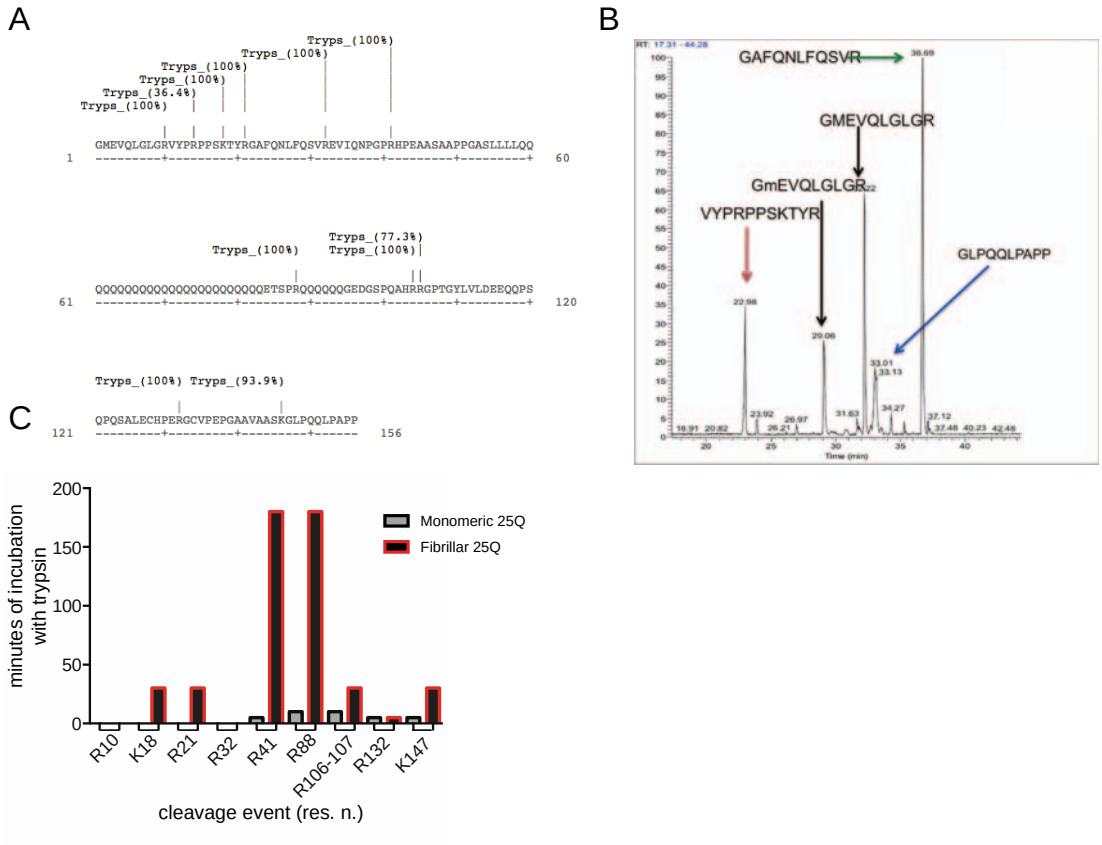


Figure 8.3Limited proteolysis experiment on 25Q. A) Map of the cleavage sites for trypsin (web.expasy.org/peptide_cutter/); B) Example of MS spectra with the relative abundance of each peptide found and the assignment of each MW to a specific peptide. C) Resistance to cleavage for each site (categories) for monomeric 25Q (grey, black borders) and fibrillar 25Q (black, red border). The fibrillar 25Q shows higher resistance to cleavage in the sites in proximity of the polyQ and moderate resistance in those in proximity of ^{23}FQNLFQSVREVIQ^{35}.

Figure 8.3-C represents the degree of protection from trypsin digestion per each cleavage site, expressed in minutes of incubation. Longer incubation times are associated to higher resistance to cleavage, which means higher protection of the region where the cleavage site is located. The monomeric 25Q is entirely accessible to the protease and its maximum protection time is of 10 min.

The sites in close proximity of the polyQ tract (position 41) and, interestingly, the cleavage site between the polyQ and the Gln_6 non-expanded repeat (position 88) are highly protected and are cleaved only after an incubation of 3 h (180 min). Also, the regions in proximity of *^{24}FQNLFQSVREVIQ^{36} show a moderate protection respect to the monomeric 25Q.*

This experiment shows that most part of the protein is not involved in the formation of the fibril, but is probably extended outside the fibril axis. The ^{24}FQNLFQSVREVIQ^{36} sequence (positions 18 and 21) is more protected in the fibril than the monomer, but far less than the region surrounding the polyQ. It is possible that the interaction on ^{24}FQNLFQSVREVIQ^{36} is maintained in the aggregates, but the aggregates formed by this region oppose less resistance to the protease, suggesting lower stability.

8.2 ATTEMPT TO BLOCK THE EARLY OLIGOMERIZATION

Experiments in literature showed that it is possible to block the oligomerization of the huntingtin peptide by incubating it in presence of a smaller peptide that mimics the N-terminal flanking region, which is the one that nucleates the oligomerization of the Htt peptide(349).

For this reason, we incubated the 25Q protein in presence of a peptide spanning over the ^{23}FQNLF^{27} region (Ac-YPRPPSKTYRGAFQNLFQSVREVIQNP-Am, called FQNLF peptide) and with another peptide designed always from the sequence of 25Q, but based on a sequence that shows no high R_2 relaxation, nor decrease in intensity in the time-resolved NMR experiment (Figure 8.4-A). The control AR peptide was incubated also alone to make sure that it does not form any aggregate (Figure 8.4-B).

For this experiment, 20 µM of 25Q was aggregated in presence of 6 equivalents (E) of the FQNLF peptide or the AR control peptide and the evolution was monitored with the sedimentation by ultracentrifugation technique, in order to avoid a misinterpretation of the data, due to oligomerization of the protein with the peptide.

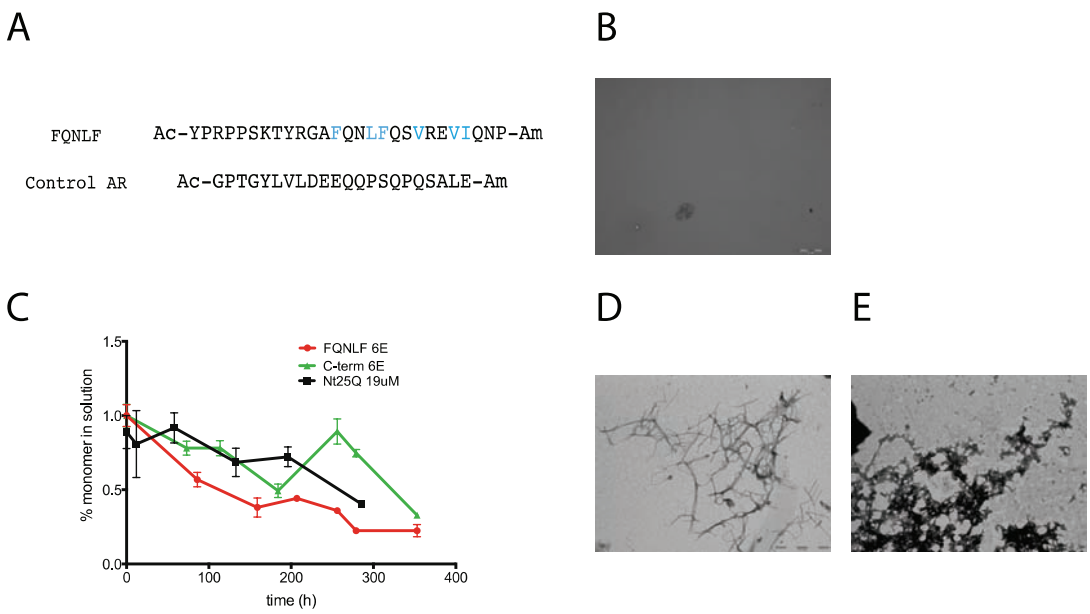


Figure 8.4 Aggregation of 25Q in presence of peptides. A) Description of the peptides. B) AR control peptides 120 µM incubated for 9 days. No formation of aggregates. C) Sedimentation experiment by ultracentrifugation of 20 µM 25Q without peptides (black), in presence of 6E of FQNLF peptide (red) and 6E of AR control peptide (green). The % monomer in solution refers to 25Q monomer. D) 25Q incubated in presence of 6E FQNLF after 9 days of incubation. The entire field shows fibrils. E) 25Q incubated with control peptide for 30 days. Aggregates are formed as for the 25Q alone.

Figure 8.4-C shows that, instead of blocking or delaying the aggregation, the FQNLF peptide induces a slightly faster aggregation to 25Q. More strikingly, we found every field analyzed for 25Q in presence of 6E FQNLF entirely covered with fibrils already after 9 days of incubation, meaning that this peptide favors the fibrillization of 25Q.

These data show that it is not possible to block the aggregation of AR by using peptides targeted against the region that nucleates the oligomerization. To find a possible explanation, we further studied the behavior of the FQNLF alone.

8.3 THE FQNLF PEPTIDE AGGREGATES WITHOUT FORMING FIBRILS

To prove that the region ^{24}FQNLFQSVREVIQ^{36} aggregates also when isolated from the protein, we produced a synthetic peptide spanning over this region (and we monitored its aggregation, compared to the same concentration of $K_2Q_{25}K_2$ peptide (Figure 8.5).

The FQNLF peptide aggregates faster than $K_2Q_{25}K_2$ (Figure 8.5-B), but produces a different species of aggregates. $K_2Q_{25}K_2$ forms amyloid-like fibrils and no amorphous aggregate is observed (figure 8.5-C), while FQNLF peptide forms amorphous aggregates that never evolve to fibrils for the time span analyzed (Figure 8.5-D,E).

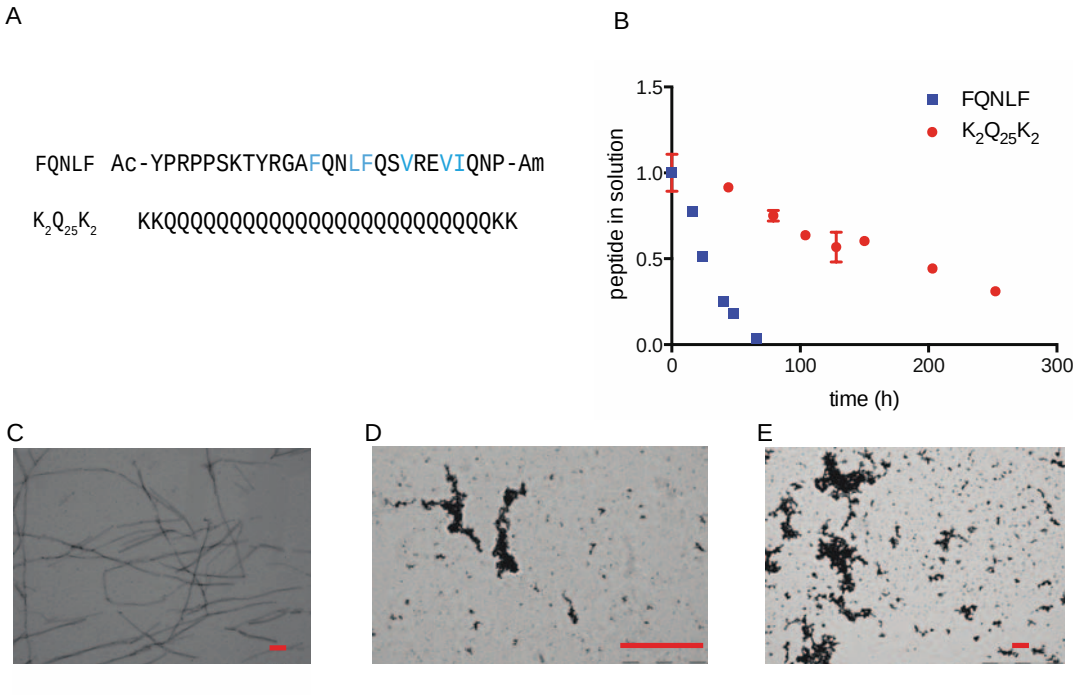


Figure 8.5Aggregation experiment on FQNLF peptide. A) Description of the peptides used. B) sedimentation by ultracentrifugation assay of 20 µM FQNLF (blue) and $K_2Q_{25}K_2$ (red). C) Fibrils of $K_2Q_{25}K_2$ after 11 days of incubation. D) Aggregates of FQNLF peptide after 23 days and (E) 52 days. In C,D,E, the red bar represents 100 nm.

This data clearly demonstrate that also the ^{24}FQNLFQSVREVIQ^{36} region forms aggregates and with a faster rate than the polyQ. However, only the polyQ forms fibrils, meaning that the

first amorphous aggregate observed in the AR recombinant proteins are due to the self-assembly of the FQNLF region, while the fibril formation is a later event, as is suggested by the different kinetics of the two peptides.

8.4 OLIGOMERIZATION OF ^{24}FQNLF^{29} PRECEDES THAT OF POLYQ IN AR

The slow oligomerization time experienced by the 4Q and 25Q proteins opened the possibility to investigate changes in secondary structure along the oligomerization process.

From DLS data in the previous chapter, we knew that oligomeric species are visible immediately after the first day of incubation after disaggregation, however the decrease in monomer concentration is not apparent from the sedimentation by ultracentrifugation assay, so that most of the protein is soluble for a time long enough to perform other assays.

We initially studied 30 μM of 25Q with circular dichroism at different time points along the oligomerization time (Figure 8.5-A). No clear change in secondary structure was visible, but this could be due to the small fraction of protein that is undergoing conformational changes,relative to the rest of the protein that does not change secondary structure.

We repeated this experiment by studying 25Q by NMR, in collaboration with Dr. BaharehEftekharzadeh and Dr. Jesus Garcia from IRB, performing a series of ^{1}H-^{15}N HSQC experiments at 18.8 T 8800 MHz) on 50 μM freshly disaggregated ^{15}N-25Q in a time-resolved fashion.

Unlike what observed for other faster aggregating proteins, like Aß(350), it was possible to distinguish specific cross-peaks decreasing in intensity or shifting, while the most part of the resonances did not experience changes (Figure 8.5 B).

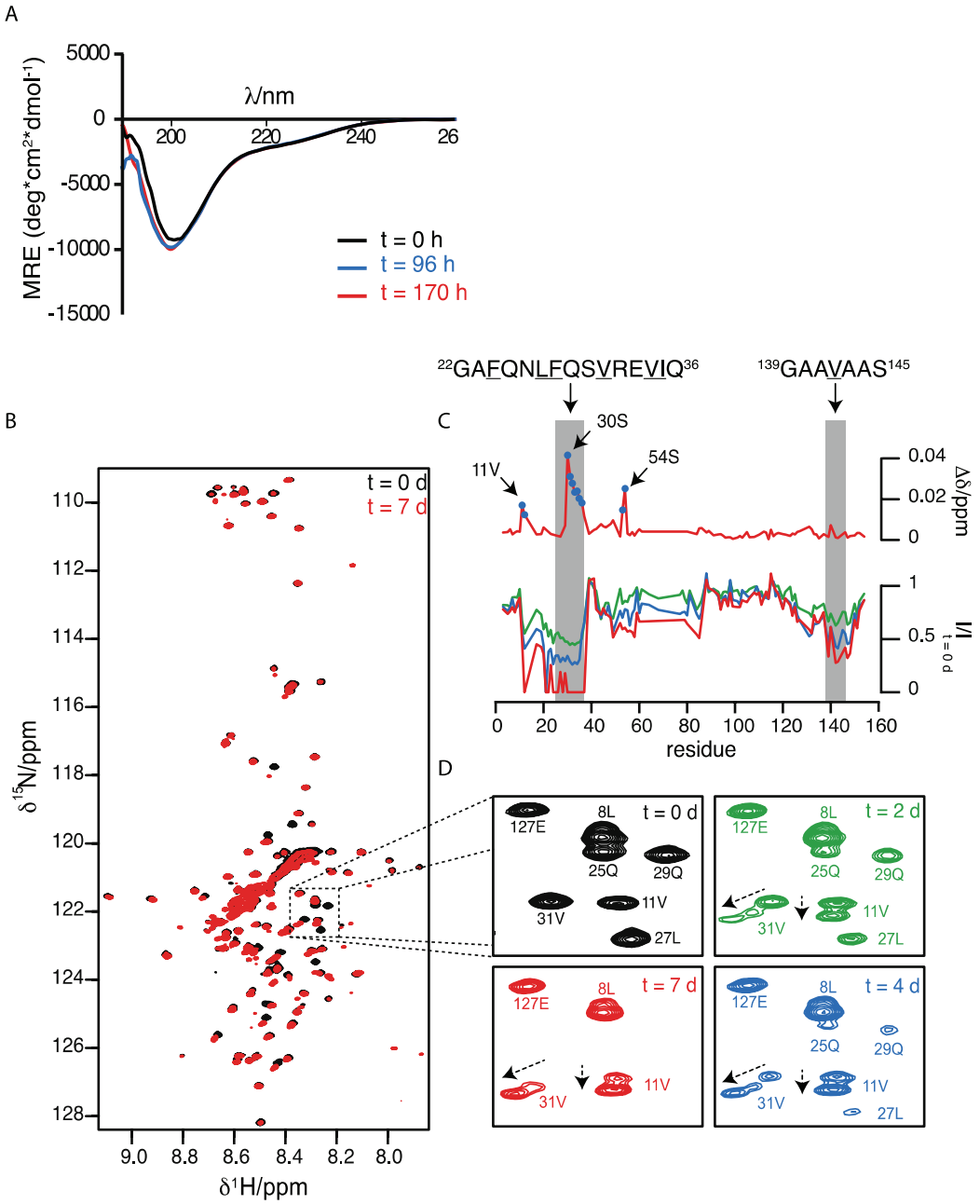


Figure 8.6Analysis of oligomerization of 25Q by CD and NMR. A) CD experiment with 30 µM 25Q, measured after disaggregation (0 h, black), 96 h (blue) and 170 h (red). B) ^{1}H-^{15}N HSQC spectrum of 50 µM freshly disaggregated sample of 15N-labeled 25Q (black) and of the same sample after 7 days of incubation (red). C) Top: chemical shift perturbations, where $\Delta\partial=sqrt[\Delta\partial_H^2+(\Delta\partial_N/5)^2]$, observed after 7 days of incubation reported as a function of residue number. Blue dots correspond to residues, which peaks experience splitting. Bottom: differences in intensities compared to the HSQC at day 0 ($I/I_{t=0d}$) reported as a function of residue number for 2 (green), 4 (blue) and 7 (red) days. D) Details of the spectral changes observed in a specific region of the ^{1}H-^{15}N HSQC spectrum immediately after disaggregation (black), and after 2 (green), 4 (blue) and 7 days (red) of incubation. In the motif mentioned in the figure, hydrophobic residues are underlined and Q and N residues are shown in bold (EftekharzadehPiai 2015).

From Figure 8.6-C is clear how a set of resonances associated to a specific region in the 25Q protein is decreasing in intensity. Decreases in intensity and shifts in the resonances, like in Figure 8.6-D, are often associated with protein-protein interactions.

This scenario is compatible with a slow exchange between the monomer and a species too large for being observed by NMR (an oligomer), which reflects in a set of residues that experience a decrease in intensity, as they are in slow exchange with this second state.

The decreases in intensity are focused mainly on the regions of ^{22}GAFQNLFQSVREVIQ^{36} and ^{139}GAAVAAS^{145}, while polyQ experiences it only towards day 7. The region of ^{29}QSVREVIQ^{36}, is associated to the resonances that decrease most rapidly in intensity. This region also experiences a splitting of the cross-peak, showing that these residue visit at least an intermediate state still visible by NMR.

Altogether, these data suggest that the region ^{24}FQNLFQSVREVIQ^{36} drives the nucleation of 25Q oligomers, while the interaction at the level of the polyQ should occur only at a later stage. However, these experiments could not address what forces are driving this interaction or how the oligomer is formed. To address this, is necessary to perturb the system, introducing mutations in the protein and evaluating their effect on the aggregation mechanism.

8.5 HELICITY IN ^{24}FQNLFSVREVIQ^{36} IS IMPORTANT FOR OLIGOMERIZATION

We studied in closer details the sequence of ^{24}FQNLFQSVREVIQ^{36}, and we observed that hydrophobic residues are located in positions that closely recall those required for forming an amphipathic α-helix, typical of a coil-coil interaction.

In protein engineering, a widely used method to evaluate the contribution of key residues in cases of simple two-state kinetics of folding is the Φ-value analysis. The stability of a protein is expressed as the difference between the energy at its transition state (the state at higher potential energy along the reaction coordinates, which is rate limiting) and the energy at its denatured state (ΔG^{TS-D}) and the Φ-value is the ratio between the ΔG^{TS-D} of the wild type and the one of the mutant protein, where the mutation is generally a mutation to alanine or another small side chain instead of the residue analyzed. A Φ-value close to 0 means that the interaction between the residue analyzed and its surroundings is poorly formed in the transition state, which means that the transition state closely resembles the denatured state. Conversely, values close to 1 mean that the transition state and the folded state are very similar one another in the surroundings of the mutated residue(351).

This system works elegantly with simple kinetics of folding, which can be explained by a two-state model, while becomes less straightforward when the energy landscape presents several local minima, or, even worse, when it is very shallow, as in the case of IDPs.

Nevertheless, an approach with directed mutations could shed light on what sequence properties are important to stabilize the oligomer, by mutating the regions that might be involved in the interaction or in the key secondary structure changes.

We identified helicity as a main requirement for the formation of the oligomer. Alanine is a very helical-prone and a poorly hydrophobic residue, such that substitutions of hydrophobic residues, like Phe, Leu, Val or Ile would increase the helical content of the ^{23}FQNLFQSVREVIQ^{36} motif(352).

Therefore, we designed a series of mutations to Ala affecting the hydrophobic residues of each motif (Figure 8.7-A). To monitor how the secondary structure would change as a consequence of Ala substitutions, we performed a set of secondary structure predictions with Agadir and of disorder predictions with PONDR (Figure 8.7-B,C), whereas the effect of the mutation was monitored by time-resolved DLS (Figure 8.7 D).

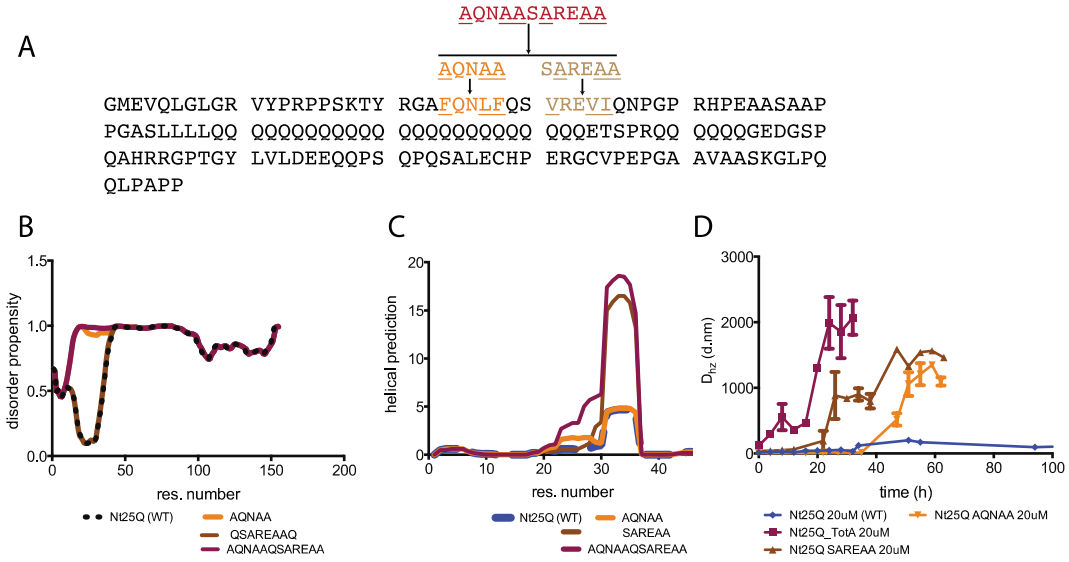


Figure 8.7Experiment with mutant in ^{24}FQNLFQSVREVIQ^{36}. A) Design of the mutants and legend: orange for AQNAA mutant, brown for SAREAA mutant and red for AQNAASAREAA mutant. B) PONDR prediction for the three mutants and wild type 25Q. AQNAA abolishes the predicted ordered region. C) Agadir prediction for the three mutants and wild type 25Q D) Time-resolved DLS for the three mutants and wild type. In B, C, D 25Q is reported as dotted black, AQNAA as orange, SAREAA as brown, AQNAASAREAA as red.

Comparison of Figure 8.7-C and Figure 8.7-D show a correlation between the predicted helicity of this motif and its aggregation rate. Higher helical content determines higher oligomerization rate. Conversely, hydrophobicity of residues in positions *i, i+3, i+4, i+7* do not seem to play a role in this interaction, so that probably the residues that form the contacts that stabilize the tetramer are not the ones mutated. Examples of the importance of helicity in the oligomerization process are found in aggregation studies of islet amyloid polypeptide (IAPP)(353, 354) and α-synuclein(46). In both cases, an α-intermediate was found on-pathway for the formation of the amyloid fibril and was reported to form the first oligomeric assembly(46, 353).

In addition, the assembly of early oligomers is mediated by α-helical intermediates, which do not necessarily change conformation once in the aggregate. TEM data already showed in this chapter and limited proteolysis data demonstrate that the fibril is composed essentially by the polyQ tract, while the rest of the protein is involved in an amorphous aggregation, which could be called α-aggregation so that it can be distinguished from ß-aggregation.

Nevertheless, this data do not show what are the residues responsible for the contact, but at least point those that are not responsible, as mutating them does not decrease but instead increases the aggregation rate. We therefore investigated the residues Gln_{24}, Asn_{25}, and $Gln_{28,}$ as well as Ser_{29}, Arg_{31} andGlu_{32}.

A close look at the sequence of the ^{23}FQNLFQSVREVIQ^{36} motif reveals that there are two faces: a hydrophobic side, which is the one described in 8.5 and a polar and charged side, represented by the residues in position 24, 25, 28 and 38. This organization would produce a polar spine opposite to the hydrophobic phase.

Interestingly enough, the hydrophobic side is involved in the N/C interaction, as it is binding to the hydrophobic cleft of AF-2(241).

Therefore, the explanation of the results obtained in 8.5 could be that the stabilization of the helix by more helical prone residues forces the motif in α-helical conformation, so that the polar residues are organized in a spine and can interact.

Figure 8.8 model of the heptad wheel of the α-helix for the ^{23}FQNLFQSVREVIQ^{36} region. Hydrophobic residues are colored in green, polar residues in purple, the positively charged residue in red and the negatively charged residue in blue.

In order to identify which residues were key for the interaction, we decided to produce peptides derived from the FQNLF peptide, to exploit its fast kinetics of aggregation, to better appreciate reductions of the aggregation rate. A set of peptides analogous to the FQNLF peptide was synthesized carrying a mutation to Ala or to Glu in positions Gln24 and Gln28. We chose these two specific residues because mutations to Ala, or to the analog charged residue Glu, do not change its helicity and that these two residues are highly conserved throughout the phylogenetic tree in the species that bear the ^{23}FQNLFQSVREVIQ^{36} motif.

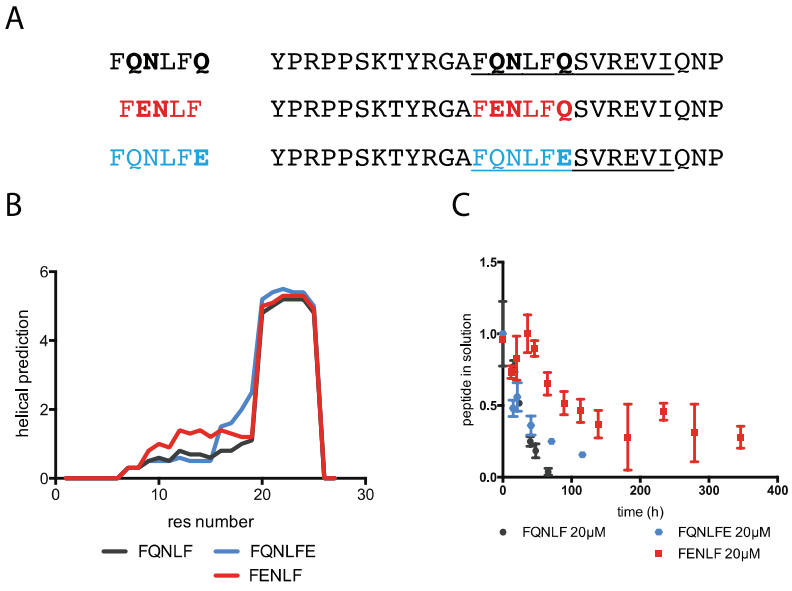


Figure 8.9 Kinetic experiment on mutation in position 24 and 28. A) Peptides used in the experiment and relative names and color code. B) Agadir prediction of the effect of the mutation on helicity. In black the FQNLF wt peptide, in blue FQNLFE, in red FENLF. C) Kinetics experiment by sedimentation by ultracentrifugation. The FQNLF (black) peptide aggregate faster than the two mutants and FENLF (red) aggregates slower. The effect of mutation in FQNLFE (blue) is lower, but it still aggregates less rapidly than FQNLF.

Figure 8.9 shows how the mutations in position Gln_{24} and Gln_{28} affect the aggregation rate of the FQNLF peptide. More specifically, mutation to Glu of Gln_{24} dramatically decreases the aggregation rate of the peptide, whereas mutation on the second site is less effective, but with a still noticeable decrease in aggregation.

This data defines that the residues key in the interaction between ^{23}FQNLFQSVREVIQ^{36} motifs are precisely Gln_{24} and Gln_{28} as exchanging Gln for a negatively charged Glu decreases the aggregation rate of the peptide. Mutations to Ala determine an analogous effect. All together, thesedata strongly suggest that the 'polar spine' comprised of residues 24, 25 and 28 mediates the first oligomerization.

8.7 SUMMARY AND CONSIDERATIONS

The set of techniques developed in Chapter 7 were tested using the current knowledge in the polyQ field and we indeed confirmed the polyQ-dependence of the aggregation rate of the AR recombinant proteins, as well as the capacity to form amyloid fibrils with a polyQ-dependent rate. However, we realized that also the recombinant protein with the least number of repeats (4Q) was aggregation-prone and oligomerized with the same initial rate as 25Q.

To make clarity, we had to formulate a distinction between oligomers and aggregates, as in literature there is little consistence in the terminology, also due to the fact that there is real no separation between very broad oligomeric species and aggregates(355). We decided to call oligomer all of those species with relatively low z-average value (<500 nm) in the time-resolved DLS, that do not decrease the intensity of the CD spectrum and that are still visible by NMR spectroscopy (Figure 8.5). Aggregates were those assemblies that did not fulfill each of these requirements and that were also observable by TEM.

TEM micrographs revealed at least two different species of aggregates: an amorphous aggregate and a more fibrillar aggregate. We confirmed that a polyQ peptide ($K_2Q_{25}K_2$) incubated in the same conditions forms only fibrils, whereas we found that 25Q and 51Q evolve in fibrils with different rates, with 25Q forming fibrils only after 40 days of incubation, while heterogeneous populations of fibrils and amorphous aggregates are observed already within the first day of incubation of 51Q.

The differences in morphology between late stage 25Q (mostly fibrillar) and 4Q (amorphous) is reflected in a marked difference of FTIR spectra, where 25Q presents bands associated to ß-sheets that are not present in the 4Q sample. 51Q fibrils are markedly similar to those isolated by other members of our group from SBMA mice (Eftekharzadeh, unpublished data).

Moreover, the experiments of limited proteolysis show that the fibrils formed by 25Q leave most part of the protein accessible to the protease, with the exception of the region in proximity of ^{23}FQNLF^{27}, which shows a moderate resistance in the fibril sample compared to the monomeric sample, and the region in close proximity to the polyQ, which is almost completely precluded to the protease.

At the level of the oligomers, the time-resolved NMR study shows that the region surrounding ^{24}FQNLF^{28}, (^{24}FQNLFQSVREVIQ^{36}) is in slow exchange with at least two other oligomeric states, as the resonances associated to this region experience both a decrease in intensity and a shift in the position of the cross-peak chemical shifts. This phenomenon occurs before that any change in chemical shift occur to the resonances of the polyQ region, meaning that the two processes have two clearly different rates, with the ^{24}FQNLFQSVREVIQ^{36} region aggregating faster than the polyQ tract.

As a further confirmation, a peptide designed over this region aggregates with a higher rate than the $K_2Q_{25}K_2$, but it does not evolve into fibrils, as for the 4Q protein.

Altogether, this information points towards a scenario where 25Q forms oligomers by interactions in the ^{24}FQNLFQSVREVIQ^{36} region. This macromolecular assembly sediments as an amorphous aggregate that evolves in fibrils with a polyQ-dependent rate

To understand what residues in ^{24}FQNLFQSVREVIQ^{36} are interacting and with what rules, we performed mutation analysis of this region, comparing the aggregation rates of different. Mutants with increasingly higher helical propensity were found to aggregate with increasingly higher rate.

We hypothesized that, if stabilizing the α-helix shifts the equilibrium towards the oligomer formation, it is very likely that the ^{24}FQNLFQSVREVIQ^{36} region exists both in extended and in α-helical conformation and that only the α-helical intermediate can form oligomers. IAPP and α-synuclein are documented to experience a meta-stable oligomer formed by a bundle of α-helices, that latter evolving into fibrils through a α-to-ß transition(46, 353, 354, 356).

In this case, it is improbable that the α-helical intermediate undergoes a further change of conformation towards the ß-sheet, as we show by limited proteolysis experiments that the region in proximity of ^{23}FQNLF^{27} does not take part of the fibril.

We then observed that residues Gln_{24}, Asn_{25}, Gln_{28} and Gln_{36} are located as *i+3,i+4,i+7 and i'+7* in the heptad repeat representation of an α-helix, forming a polar surface We hypothesized that this 'polar spine' could form a loose polar zipper and a charge clamp with the charged residues Glu_{32} and Arg_{31} of a second molecule.

We tested this hypothesis by performing mutations on the highly conserved residues Gln_{24} and $Gln_{28}$and we reduced the propensity of the ^{24}FQNLFQSVREVIQ^{36} region to aggregate, thus confirming our hypothesis.

CHAPTER 9: POLYQ FLANKING REGION AND HELICITY

From the previous chapters, we can begin to propose a scenario where AR aggregates via a complex mechanism, by which two regions self-associate with different rates. Chapter 8 was dedicated to describing the region of nucleation away from the polyQ, while we will discuss in this chapter the N-terminal flanking sequences to the polyQ region.

In Chapter 6, the NMR study of the 25Q and 4Q protein revealed that the polyQ in both constructs has helical propensity, which is stronger in 25Q. Interestingly, the same sequence ^{55}LLLLQQQQ^{63} in the two constructs is more helical in 25Q than 4Q, suggesting cooperativity of the polyQ in folding.

Nevertheless, there is wide description in literature of polyQ as a random coil. CD spectra of polyQ peptides show the typical signal of random coils, however the slight red shift of the main minimum (around 200 nm), together with the moderate second minimum at 220 nm, suggest that these peptides present small populations of other secondary structures (295, 323, 344). Also FTIR data show polyQ peptides as random coils, which convert in ß-sheet with a polyQ-dependent rate, while aggregating(81).

HSQC NMR experiments both of ataxin-3 and of a model system (where a polyQ is fused to GST protein) both indicate that the polyQ region have the typical collapsed organization of a random coil(82, 83). However, ssNMR experiments show the polyQ as adopting a series of different secondary structures, among which the PP-II helix when in presence of a polyPro helix (Pro_{10}) at its C-terminal(325). Interestingly, the Pro_{10} flanking region is described to block the aggregation of the huntingtin peptide only when C-terminal of the polyQ, while does not produce any effect when located at its N-terminus(357).

We identified the ^{55}LLLL^{59} motif N-terminal to the polyQ as the N-terminal flanking region of this protein and this chapter provides the evidence that this motif directly affects the secondary structure of the polyQ tract. Furthermore, this chapter discusses the role of ^{55}LLLL^{59} in respect to the aggregation propensity of the protein

9.1 A POLYQ PEPTIDE WITH A ^{55}LLLL^{59} MOTIF IS HELICAL

The first approach for evaluating the effect of the ^{55}LLLL^{59} motif is to compare a model of polyQ peptide, already described in literature(344), with a similar version containing this repeat. We already characterized the aggregation of $K_2Q_{25}K_2$ and we used this peptide to design the $K_2L_4Q_{25}K_2$ peptide (Figure 9.1-A), meaning a peptide in everything similar to $K_2Q_{25}K_2$, with the insertion of 4 Leu residues between the N-terminal Lys and the polyQ. Freshly disaggregated peptides of $K_2Q_{25}K_2$and $K_2L_4Q_{25}K_2$ were studied by CD at comparable concentrations (113 µM and 140 µMrespectively).

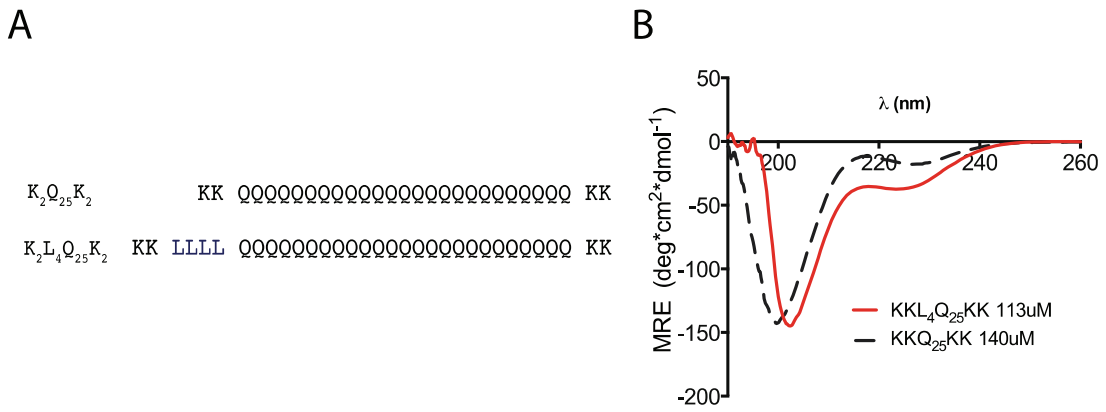


Figure 9.1Comparison of $K_2Q_{25}K_2$ and $K_2L_4Q_{25}K_2$. A) Description of the peptides. B) CD spectrum of 113 μM $K_2L_4Q_{25}K_2$ (red) and of 140 μM $K_2Q_{25}K_2$ (black streaked).

Figure 9.1-B shows the CD spectra of $K_2Q_{25}K_2$ and $K_2L_4Q_{25}K_2$. The signal related to $K_2Q_{25}K_2$ peptide is mainly that of a random coil, with a small component of helicity, in very good agreement with the current literature(72, 323, 344). Conversely, the marked second minimum at 222 nm of the spectrum related to $K_2L_4Q_{25}K_2$and the red shift of the minimum from 200 nm to 204 nm are clear signs of increased helicity.

Nevertheless, to a close comparison with the NMR data, the $K_2L_4Q_{25}K_2$ peptide presents a degree of helicity lower than what expected from the SSP analysis described in 6.4. The lower level of helicity in the $K_2L_4Q_{25}K_2$peptide can be attributed to the presence of two Lys residues at C-and N-terminus. Lys is a relatively good C-capping residue, but a very bad N-capping residue and the presence of Lys both at N- and C-terminus destabilizes the helix (358).

A α-helix is stabilized by H-bonds between the backbone –NH in position *i* and the C=O in position *i+4*. To preserve this interaction also at the extremities of the helix, a charged residue is often present at each side with different charge, each of which establishes an H-bond with, respectively, the N-H or the C=O on the backbone. In practice, placing positively charged residues on both sides of the helix destabilizes the dipole and decreases the stability of the helix. Probably, a better approach for capping the helix would have been to use pairs of Asp or Glu residues at the N-terminus and a Lys residue at the C-terminus, but we were limited by the requirements of solubility that a peptide containing exclusively Gln and Leu would imply, as well as for their extended use in literature, which has been already discussed.

We concluded that this experiment, however strongly suggestive, does not unequivocally demonstrate that ^{55}LLLL^{59} is the only cause of the strong helical propensity of the polyQ in AR.

9.2 THE ^{55}LLLL^{59} MOTIF PREVENTS FIBRIL FORMATION

The $K_2L_4Q_{25}K_2$ peptide was also used to monitor the effect of ^{55}LLLL^{59} on the aggregation of a polyQ peptide and on the eventual formation of fibrils.

A freshly disaggregated solution of 20 µM $K_2L_4Q_{25}K_2$ was incubated and monitored by sedimentation by ultracentrifugation and TEM micrographs were performed at different stages of the aggregation.

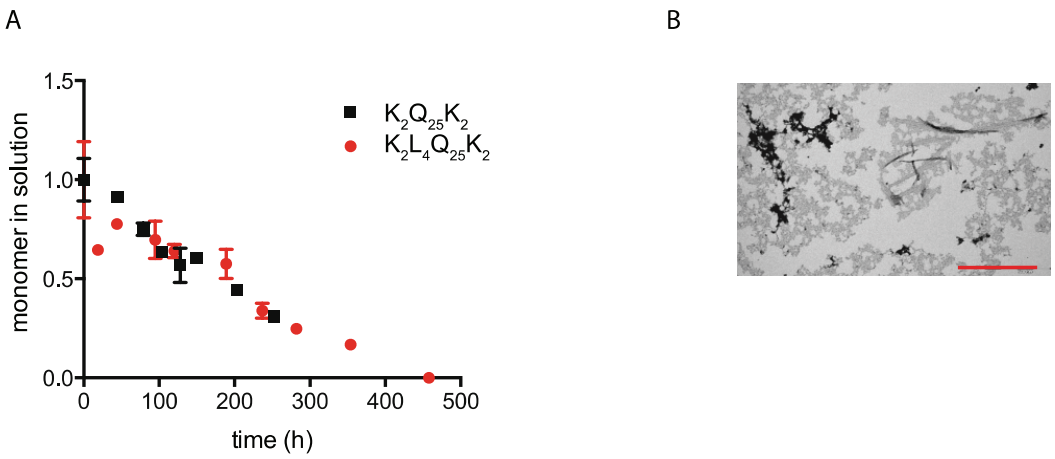


Figure 9.2 Aggregation experiment of 20 µM $K_2L_4Q_{25}K_2$ and $K_2Q_{25}K_2$. A) Sedimentation by ultracentrifugation. 20µM $K_2L_4Q_{25}K_2$ B)TEM micrograph of $K_2L_4Q_{25}K_2$ after 16 days of incubation. Mixed field of amorphous aggregates and fibrils. The red bar represents 500 nm

The sedimentation by ultracentrifugation assay does not show any distinguishable difference in the kinetics of the two peptides (Figure 9.2-A). However, TEM micrographs of the $K_2L_4Q_{25}K_2$ peptide after 16 days show a mixed population of amorphous aggregates and fully formed fibrils, where fibrils and aggregates are never in the same bundle, but instead seem to cluster separately (9.2-B).

All together, these data suggest that $K_2L_4Q_{25}K_2$ aggregates with a behavior similar to 25Q, where a preliminary amorphous aggregate is formed, which then reorganizes in fibril. This reorganization seems to be consequence of a nucleation episode, as aggregates that are separate in space evolve independently.

9.3 MUTANTS IN THE LEU$_{55\text{-}59}$ REGION

Working on peptides provided good insights in the function of ^{55}LLLL^{59}, but it did not provide a definitive explanation of its relationship with the aggregation process.

We therefore approached the same issue by performing, in collaboration with Soranalftemi, a set of mutations on 25Q, first substituting each Leu residue in the four positions with a less hydrophobic Ala, then the whole region with an ^{55}AAAA^{59} motif (called AAAA) (Figure 9.3-A).

Agadir predictions suggested an overall increase in helicity of the polyQ tract and its flanking region in each mutation to Ala (figure 9.3-B). Despite the predictions, only a modest decrease in helicity was observed with CD (figure 9.3-D).

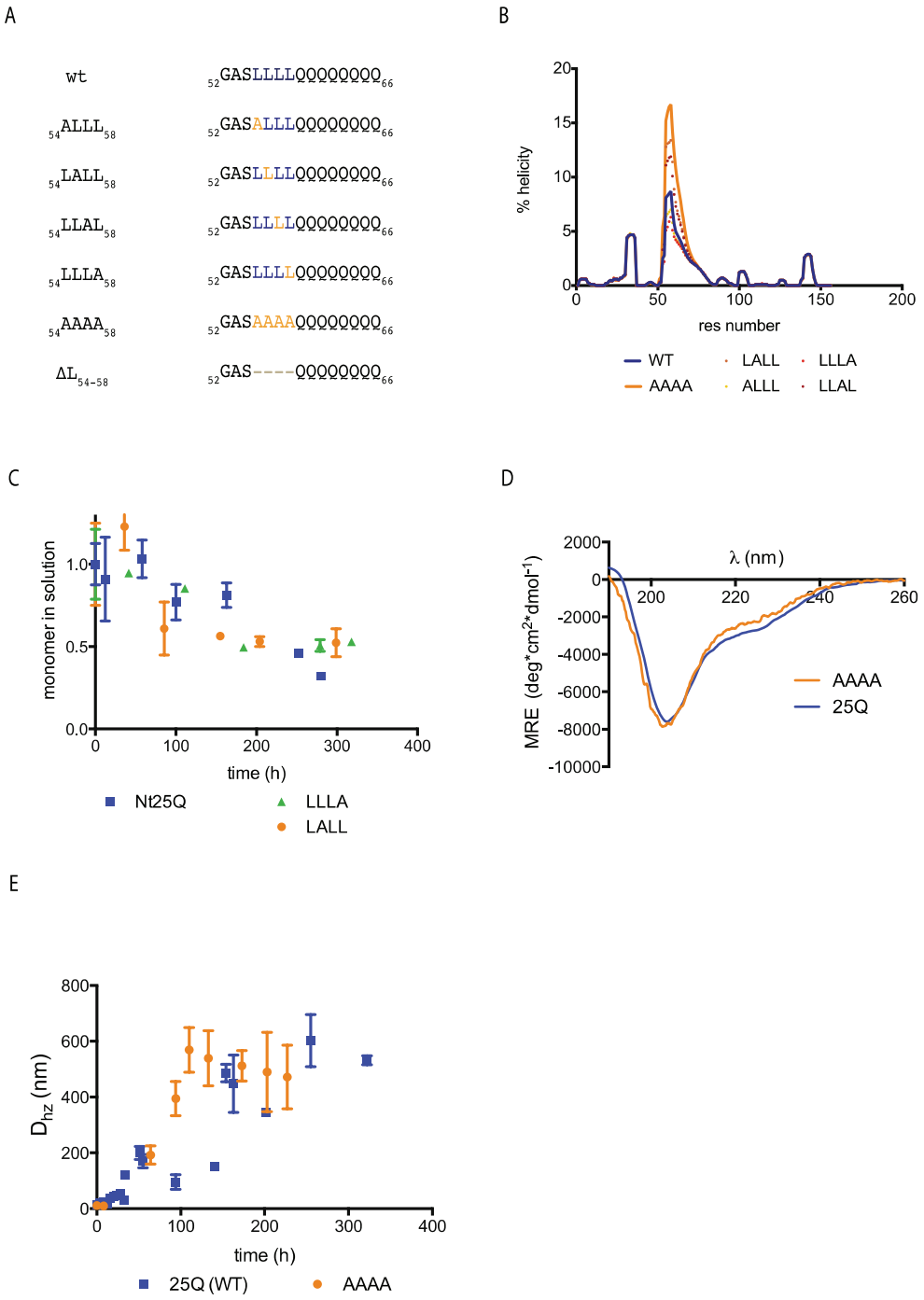


Figure 9.3 Experiments on Leu→Ala mutants. A) Name and description of each mutant, the mutation to Ala is marked in orange, the deletion with grey lines. B) Agadir prediction of each Leu→Ala mutant. 25Q (wt) is the blue straight line, AAAA the orange straight line, point mutations are dotted lines. C) Sedimentation assay of 20 µM 25Q (blue), 20 µM LALL (orange), 20µM LLLA (green). D) CD spectra of 25Q (blue) and AAAA (orange). E) Time-resolved DLS of 25Q (blue) and AAAA (orange)´.

In addition, the aggregation experiments did not provide clear results, as both sedimentation by ultracentrifugation and time-resolved DLS failed to find differences in aggregation or oligomerization rate among the mutants (Figure 9.3-C,E).

Given these inconclusive results, a new mutant of 25Q was designed, which carried a deletion of the ^{55}LLLL^{59} motif. Again, CD spectra were recorded and aggregation assays performed, trying to understand the correlation between any change in secondary structure and the aggregation rate. This approach also had the advantage of being complementary to the experiments performed on peptides.

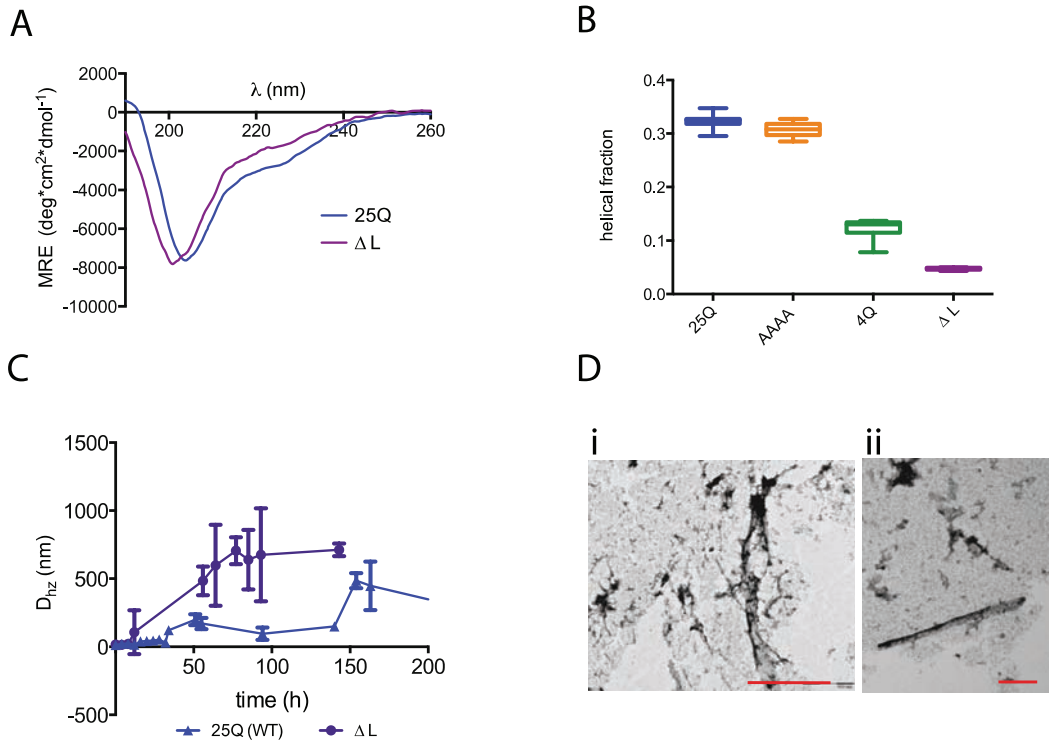


Figure 9.4 Experiments with ΔL_{54-58}. A) CD spectra of 130 µM 25Q (blue) and 130 µM ΔL_{54-58} (purple) and (B) relative deconvolution by CONTIN. AAAA mutant is in orange and 4Q in green. (NRMSD of each simulation is annotated in the same color on top of the boxplot) C) Time-resolved DLS experiment of 20 µM 25Q (blue) and 20µM ΔL_{54-58} (purple). D) TEM micrographs of ΔL_{54-58} after 10 (i) and 20 days (ii) of incubation. The red bar represents 100 nm.

These experiments show how ^{55}LLLL^{59} affects secondary structure and aggregation propensity in 25Q (Figure 9.4). The CD spectra of ΔL_{55-59} and 25Q are markedly different, as that of ΔL_{55-59} shows with much less pronounced helicity (figure 9.4-A). A deconvolution of the spectra with the algorithm CONTIN in Dichroweb allowed a straight comparison of the helical content of 25Q, 4Q, ΔL_{55-59} and AAAA mutant. 25Q and AAAA had no real difference in helicity, while ΔL_{55-59} presented a helical content even lower than 4Q. This demonstrates that the removal of the ^{55}LLLL^{59} motif, but not of the polyQ, completely abolishes helicity in this protein.

Time-resolved DLS experiments show a higher aggregation rate of the ΔL_{55-59} protein than 25Q (figure 9.4-C), whereas TEM micrographs taken at different time points show the presence of fibrils far earlier than for 25Q wt (10 days against the 41 days of 25Q) and in 20

days it is already possible to observe perfectly formed fibrils, similar to those of 51Q or the late stage 25Q (figure 9.4-D).

Altogether, these data strongly support the hypothesis that ^{55}LLLL^{59}nucleates the α-helix formation in the polyQ tract and stabilizes this conformation, preventing the conversion to ß-sheet and the consequent formation of fibrils. Conversely, lacking this motif, the polyQ adopts random coil conformation, which has a lower entropic cost of a conversion to ß-sheet, leading to a prompt fibril formation. This mechanism finds strong similarities with the effect of the Pro_{10} repeat C-terminal to the huntingtin peptide(91, 359).

9.4 POLYQ IS UNSTRUCTURED IN THE ΔL_{55-59} MUTANT PROTEIN

The Δ_{55-59} protein was studied by NMR to know at a residue level the effect of ^{55}LLLL^{59}.An HSQC spectrum was recorded of freshly disaggregated 50 µM ΔL_{55-59} ^{15}N-labeled protein, in collaboration with Daniele Mungianu and Dr. Jesus Garcia and then compared to an HSQC performed on an analogous preparation of 25Q.

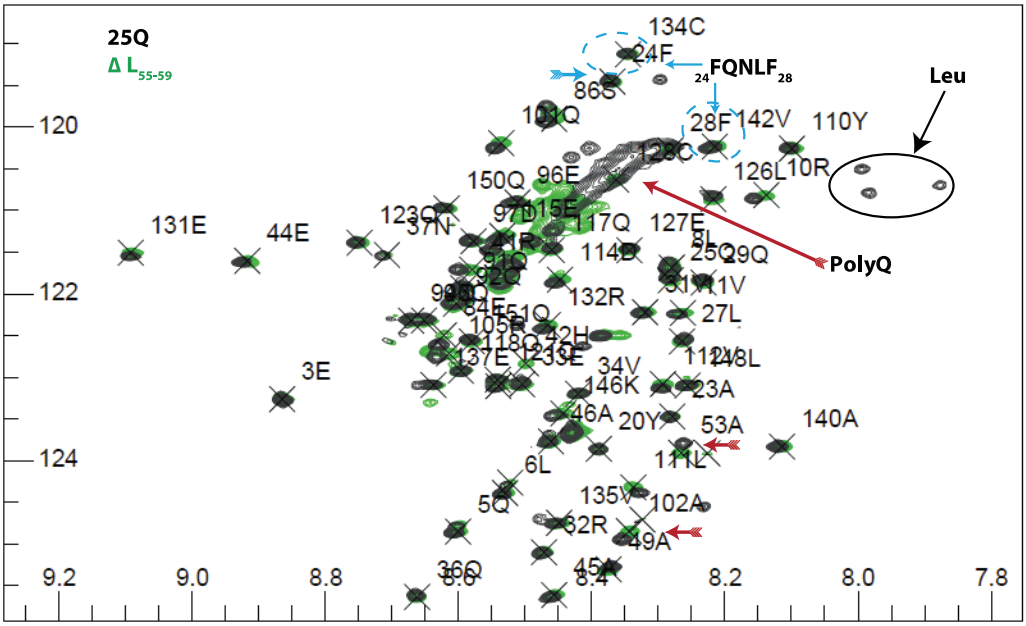


Figure 9.5 Overlay of the HSQC experiment for 25Q (black) and ΔL_{55-59} (green). The Leu residues missing in $\Delta L_{55-59}$are circled in black. Examples of regions that do not change their chemical shifts are marked with blue arrows, region that experience shift in the cross-peaks are pointed with red arrows.

The transfer of the assignment was facilitated by the overlay of most part of the spectra, however it was not possible to reassign the peaks associated to the polyQ, due to the broad shift of the cross-peaks in the central region of the spectrum. Furthermore, the region flanking ^{55}LLLL^{59} experiences a large change in the positions of the chemical shifts: whereas it was possible to reassign the peaks associated to Ala_{53} and Ala_{49}, the positions of residues like, among others, Ser_{54} and Gly_{51} were impossible to reconstruct.

Interestingly, no shift was observed of the peaks associated to the ^{24}FQNLF^{28} motif, as well as to the region right C-terminal to the polyQ (^{84}ETSPR^{88}).

The strong collapse of the peaks associated to the polyQ towards the center of the spectrum means that, in $\Delta L_{55\text{-}59}$, every Gln residue in the tract explores very similar chemical spaces, a scenario compatible with a random coil. It is also interesting that the cross-peaks associated to the ^{84}ETSPR^{88} sequence, between the polyQ and the second polyQ (Gln_6in position 88-94) do not experience shifts, meaning that the effect of ^{55}LLLL^{59} does not extend on the whole polyQ, but decreases along the tract, as already suggested by the SSP analysis of 25Q (Section 6.4).

9.5 POTENTIAL COEVOLUTION OF THE ^{55}LLLL^{59} WITH THE POLYQ

We performed a sequence alignment with the T-Coffee Server (www.tcoffee.crg.cat) of a set of 60 AR sequences, to check the degree of conservation of ^{55}LLLL^{59} across *chordata (*figure 9.6). NTD is a poorly conserved domain, partly due to the presence of highly repetitive species (two polyQ and a polyG), so that every conserved sequence has to play a crucial role in the physiology of the protein.

It is clear from figure 9.6-A that both the polyQ length and ^{55}LLLL^{59} are poorly conserved. The polyQ, as expected for polymorphic regions, presents a very wide variety across species, but interestingly the last Leu residue before the polyQ is conserved in every species where the polyQ tract is present.

It is also possible to recognize a correlation between the lengths of the repeat and the phylogenetic distance to human. This association has been documented already for huntingtin, where a phylogenetic analysis of this gene dates the first appearance of the polyQ in sea urchin, the first representative of deuterostomes according to evolutionary distance. In the work, is reported a progressive increase in the polyQ length across evolution(360).

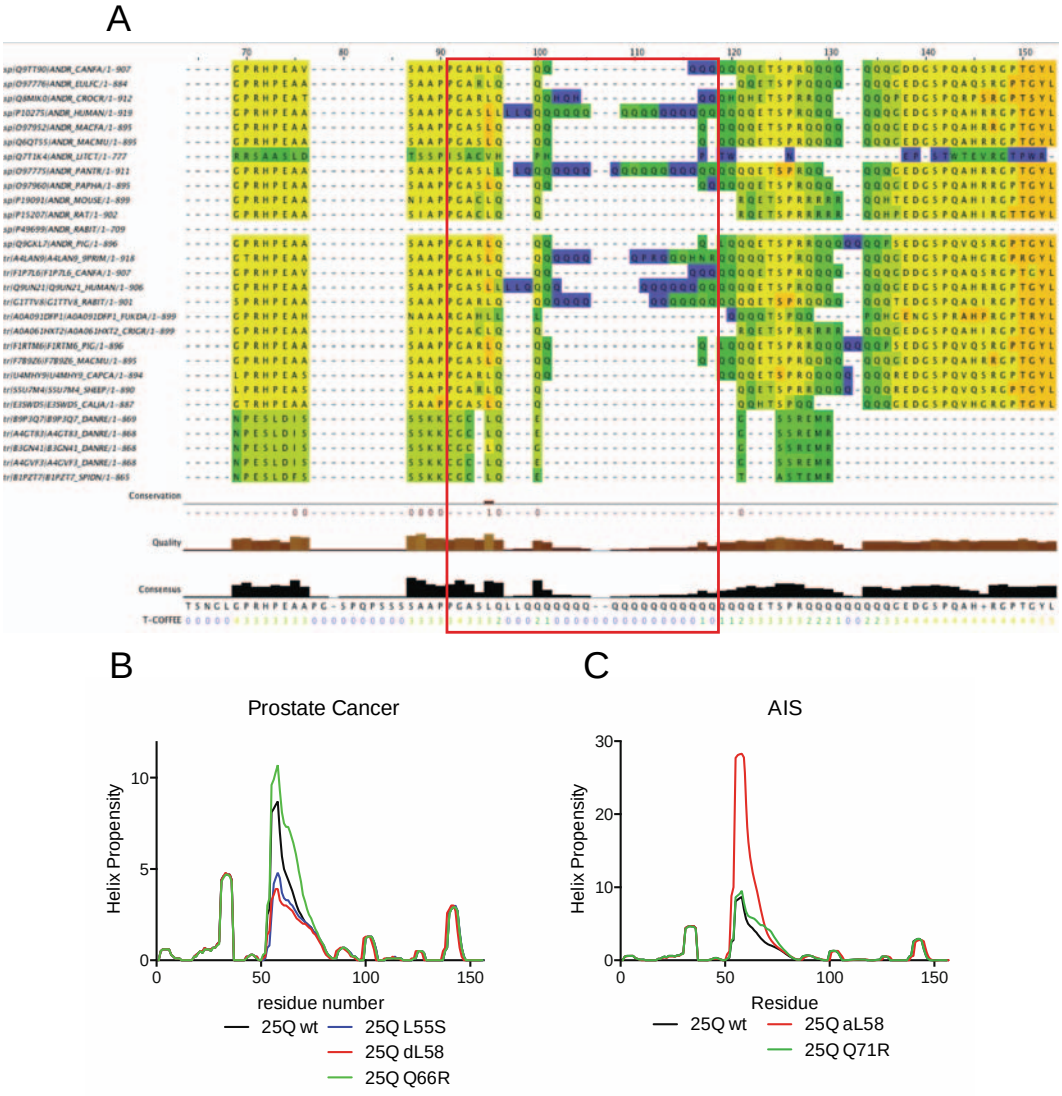


Figure 9.6 A) representation of the best-conserved sequences of an alignment of 60 different sequences of AR belonging to different species. The polyQ and its Leu-rich flanking region are framed in red). B) Agadir prediction of helicity for mutations in the polyQ flanking regions that are found in prostate cancer cases C) Agadir prediction of helicity mutations in the flanking region of polyQ that are associated in androgen insensitivity syndrome (AIS). As a general observation, cancer-related mutations, with the exception of Q_{66}→R, tend to decrease the predicted helicity of the tract, whereas mutations associated to AIS have higher predicted helicity on the polyQ.

We finally noticed a loose correlation between polyQ length and the length of Leu-rich motifs (only in human 4 Leu residues are present): short polyQ tracts are almost never preceded by Leu-rich motifs, except for the Damara mole-rat (*fukomysdamarensis*), unique species found in our search containing a Leu-rich motif of 3 repeats associated to a polyQ of only 4 repeats. Conversely, the only Leu-rich motifs N-terminal to the polyQ, other than the human ^{55}LLLL^{59}, is found in the common chimpanzee (*Pan troglodytes*), with 3 repeats.

This particularly low conservation is typical of regions flanking a highly repetitive polymorphic regionand in general of all the low-complexity regions (LCR)(361, 362) and it is reported that regions surrounding LCRs have a probability of substitution that inversely correlates with the distance from the polymorphic site: residues in closer positions have high

variability, which decreases steeply with distance(363). In this work, extracting a pattern from the highly variable sequences studied was possible only by using the very high number of sequences of higher primates and humans, deposited in the 1000 Genome Project(364). The high variability of these regions requires sampling from very evolutionary close species or within human populations. Therefore, the dataset here analyzed is too limited to draw any conclusion on the evolutionary pressure on ^{55}LLLL^{59}, but the hypothesis remains suggestive.

In parallel, from an overview of the McGill University database on known mutations of AR associated to diseases(190), emerges that deletion of Leu_{59} are found in cases of prostate cancer, while addition of a further Leu in position 60 (with a total of 5 Leu residues) is associated to mild androgen insensitivity syndrome and, more in general, mutations that increase the helicity of this tract are associated to androgen insensitivity (AIS, figure 9.6-B,C), suggesting a function of helicity of polyQ in the correct functioning of AR.

9.6 SUMMARY AND CONSIDERATIONS

The NMR studies discussed in Chapter 6 describe the polyQ tract as moderately helical in 4Q, but highly helical in 25Q, at least in the N-terminal part, and progressively less helical towards the C-terminal Gln residues. The highest values of SSP in both 4Q and 25Q were observed between Leu_{55} and Gln_{63} and the ^{55}LLLL^{59} of 25Q presents values associable to a fully formed α-helix.

From these observations, we hypothesized that ^{55}LLLL^{59} was responsible of the helicity in the polyQ tract and to prove it we adopted first a bottom-up approach, by designing a polyQ peptide with a ^{55}LLLL^{59}, and then a top-down approach, where this motif was deleted from the 25Q protein.

In the bottom-up approach, we designed a peptide ($K_2L_4Q_{25}K_2$) that mimicked the polyQ of AR. $K_2L_4Q_{25}K_2$ showed markedly more helical content than its control peptide $K_2Q_{25}K_2$, as reported by CD.

The degree of helicity of $K_2L_4Q_{25}K_2$, however, did not fully match with the NMR data on 25Q, which suggest an even higher helical content in the polyQ tract of the protein. This discrepancy between CD and NMR data can be due to the design of the peptides, where the presence of positively charged residues in both sides of the polyQ destabilizes the helix dipole(358).

Nevertheless, other valuable information was obtained from the study of the aggregation of $K_2L_4Q_{25}K_2$ peptide. Despite the difficulty to distinguish the oligomerization of this peptide from the one of $K_2Q_{25}K_2$, the aggregates formed present strong differences.

From Chapter 8, it is shown that $K_2Q_{25}K_2$ peptide deposits as fibrils in the TEM grid (see figure 8.4), whereas a sample taken from $K_2L_4Q_{25}K_2$ peptide aggregation after a similar incubation time (16 days) presents a mixed population of amorphous aggregates and fibrils. This suggests that the conversion to ß-sheet is slower for $K_2L_4Q_{25}K_2$ than that of $K_2Q_{25}K_2$ and potentially it occurs when the aggregates are already deposited. In general, $K_2L_4Q_{25}K_2$ forms aggregates and fibrils according to a similar mechanism as 25Q.

These data were further confirmed by the top-down approach, where deletion of the ^{55}LLLL^{59} in 25Q completely abolishes helicity in the polyQ, as showed by the CD experiment on $\Delta_{55\text{-}59}$. This finding is also in good agreement with the NMR data from the HSQC experiment on ^{15}N-labeled $\Delta_{55\text{-}59}$, where the chemical shifts associated to the polyQ are positioned with the typical collapsed organization of the random coil.

Interestingly, aggregation of $\Delta_{55\text{-}59}$ is faster and the aggregates reorganize in fibrils earlier than for 25Q. The fibrils were analogous to those formed by 51Q, while the amorphous aggregates were similar to those of 4Q and 25Q. This data suggest that the presence of ^{55}LLLL^{59} prevents the formation of fibrils, by increasing the energy barrier for the transition to ß-sheet. The α-to-ß transition has a high cost, due to the double transition required, from α-helix, to random coil, which has to overcome a first high-energy barrier, and then from random coil to ß-sheet(365), so that the polyQ in 25Q has to overcome a higher barrier than that of $\Delta_{55\text{-}59}$.

A very similar mechanism is observed for the huntingtin exon 1, where the Pro_{10} motif C-terminal to the polyQ hinders the aggregation by locking the C-terminal part of the polyQ tract in PP-II helix conformation(91, 325).

However, this mutation does not seem to alter the first mechanism of aggregation via ^{23}FQNLFQSVREVIQ^{36}, as the increase in aggregation rate can be explained as the additive effect of the two mechanisms.

PART V: DISCUSSION AND CONCLUSIONS

10.1 A MODEL FOR POLYQ-NTD AGGREGATION

The data discussed in this project point towards a complex mechanism of aggregation of the polyQ NTD, where more aggregation prone regions assemble by different mechanisms at different rates.

The main players identified with the AR recombinant proteins used were the N-terminal motif ^{23}FQNLFQSVREVIQ^{36} and the polyQ tract beginning in position 59, controlled by the ^{55}LLLL^{58}motif N-terminal to it, with a modest participation of a ^{139}GAAVAAS^{145} motif that needs yet to be addressed.

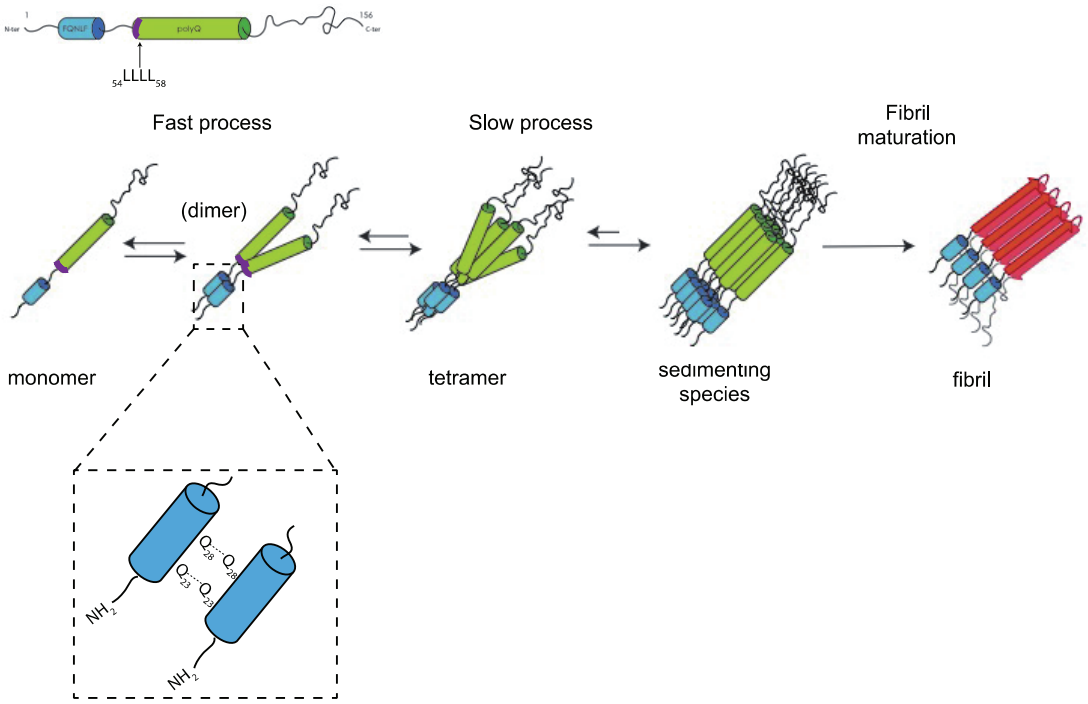


Figure 10.1 Model of aggregation of the Caspase-3 cleaved fragment of AR. The monomer is in fast exchange with an unstable dimer, through an interaction via ^{23}FQNLFQSVREVIQ^{36}. The dimer evolves quickly into a tetramer, which is the first stable species (visible by AUC) and the oligomerization becomes a slow process. Oligomers are stabilized by amphipatichelix interactions and they grow in size up to sedimentation. Sedimented species evolve in fibril, after anα-to-ß transition with a polyQ-dependent rate. This transition is an independent process respect to the $_{23}$FQNLFQSVREVIQ$_{36}$ aggregation.

In this model, the first event of oligomerization occurs in the ^{23}FQNLFQSVREVIQ^{36} region, is a process in fast exchange and very unstable. In solution, ^{23}FQNLFQSVREVIQ^{36}exists in equilibrium between a random coil and a low population in α-helical conformation and the formation of the first oligomer is mediated by the α-helical intermediate that is stabilized by a set of interactions that recalls the coil-coil but are less specific. We referred to them as amphipatic helix interactions.

The stabilization of the first oligomer is a slower process and it is the one that can be followed by NMR and DLS. The low specificity of the amphipatic helix interaction brings together polyQ tracts of different monomers, which can in turn interact, further stabilizing and enlarging the oligomer, that finally sediments as amorphous aggregates.

This kind of aggregation has been described in the literature as α-aggregation(46, 366) or coarse coil(367) and the proteins described undergo generally a slower α-to-ß transition that allows the formation of fibrils(368).

Despite the amorphous nature of the early aggregates, due to the spurious interactions that drives them, the polyQ tracts present in the oligomer can rearrange to form fibrils, with a rate that is polyQ-dependent In non-pathogenic proteins like 25Q, this process is slow and can take weeks, while expanded polyQ proteins, like 51Q, can produce fibrils in a few hours. This process and the α-aggregation are independent one another and an expanded polyQ is very likely to overcome the aggregation via ^{23}FQNLFQSVREVIQ^{36} and to progress rapidly towards the fibril formation.

10.2 ROLE OF ^{55}LLLL^{58}AND CHAMALEONIC PROPERTIES OF POLYQ

PolyQ tracts are known to be very flexible and in random coil conformations, however more recently polyQ tracts have been described as folded in secondary structures, such as PP-II helices(325), when in the presence of the C-terminal Pro_{10} motif of huntingtin. In this case, the Pro-rich motif stabilizes the polyQ residues in this conformation, preventing its conversion to a ß-sheet(91).

Similarly, the ^{55}LLLL^{58} motif nucleates a α-helix from the N-terminus of the polyQ with an effect that goes dissipating along the tract. Locking the polyQ in α-helical conformation imposes a high-energy barrier to the conversion in ß-sheet and prevents the reorganization in fibrils(365).

From an evolutionary point of view, it is known that regions flanking polymorphic sites have variability increasingly higher in the positions next to the polymorphism(363) and it is therefore possible that this Leu-rich motif coevolved rapidly in parallel to the elongation of the polyQ tract, to cope with its propensity to acquire ß-sheet conformation.

Moreover, another reported correlation between the polyQ sequences and their flanking regions describes a sequence bias towards an over-representation of a subset of residues (Pro, Leu and His) and an under-representation of others (Asp, Cys, Gly)(104) The over-represented residues are strong inductor of α-helix (His and, even more, Leu) or PP-II helix (Pro), both conformations that prevent the conversion to ß-sheet, but their codons are also not distant from CAG (CTG for Leu, CCG for Pro and CAC for His), so that point mutations can switch a Gln residue to each of these residues. This provides genetic stability by breaking the codon repeat (369). However, at the protein level, these same mutations induce the formation of a secondary structure and can prevent the aggregation of the polyQ tract.

An elegant example is the polyQ of ataxin-1, which contains a His mutation that breaks the CAG repeat and has been shown to increase the genetic stability of this tract(370–372). Interestingly, a polyQ peptide with a His residue interrupting the homopolypeptide chain aggregates slower than the same polyQ peptide without interruption (96). It is also possible

that these two functions are consequence of convergent evolution, where the same mutation produces both genetic stability and increased protein stability.

The mechanisms described allow polyQ proteins to bear a long polyQ tract, without incurring in the formation of nuclear aggregates with toxic properties. This means that the length of the AR polyQ is critical for its functions, however it remains unclear how.

It is known that short polyQ in AR are a risk factor for prostate cancer(373), which suggests that polyQ might work as a spacer between regions, to reduce the ability of transactivation. It is also known that cases of prostate cancer were found with a deletion of a codon for Leu in the ^{55}LLLL^{58} motif, whereas an addition of one Leu codon is associated to mild androgen insensitivity syndrome(190). It is enticing to draw the conclusion that the number of Leu residues N-terminal to the polyQ determines the degree of stability of the helix, or otherwise tunes the dissipation of their nucleating effect of ^{55}LLLL^{58}.

Altogether, these observations are suggestive that the polyQ needs to be long and with some degree of flexibility to comply with its physiological functions, but to prevent the aggregation, is necessary to impose structural constraints on the tract and this is performed by the Leu-rich motif in AR and by the Pro-rich motif in huntingtin.

Finally, the effect of ^{55}LLLL^{59} can potentially explain the difference between pathogenic and non-pathogenic polyQ tracts in SBMA. The overlay of the HSQC spectra of 25Q and $\Delta_{55\text{-}59}$, suggests that the last Gln residues are in random coil conformation, as the cross-peaks for the ^{84}ETSPR^{88} region do not change position, while the SSP analysis of 25Q shows that the effect of helix nucleation of ^{55}LLLL^{59} dissipates along the polyQ. Therefore, it is possible that AR with longer polyQ tracts has a broader portion of the tract prone to acquire random coil conformation. The random coil part of the polyQ can nucleate the conversion to ß-sheet and convert the aggregates in fibrils and a longer polyQ in random coil conformation favors this transition.

Therefore, longer polyQ tracts are less restrained by their flanking regions and it is possible that more portion of the tract remains as random coil, which can favorably explore the ß-conformation and nucleate the formation of a ß-sheet. This could be a mechanistic explanation of the threshold in toxicity, typical of polyQ diseases.

10.3 ^{23}FQNLFQSVREVIQ^{36} AND THE N/C INTERACTION

The interaction of ^{23}FQNLF^{28} with LBD has been characterized and crystalized(209, 241) and the crystal structure shows that ^{23}FQNLF^{28} is accommodated in the hydrophobic cleft of AF-II with the hydrophobic Phe_{23}, Leu_{27} and Phe_{28} pointing towards the inner part of the pocket and probably stabilizing the interaction with LBD, while the residues Gln_{24} and Asn_{25} are pointing outside the complex and are exposed to the solvent. (241).

From our study, we know that an increase in helicity in ^{23}FQNLFQSVREVIQ^{36} significantly favors the oligomerization of 25Q, suggesting that the α-helical intermediate is important both for the N/C interaction and for the oligomerization of the AR fragment. We also demonstrated that the interaction that stabilizes the oligomer do not involve the same residues than the ones responsible for the binding to AF-II, which means that both interactions could take place in the same moment.

The oligomerization of AR is still matter of controversy, as it is not clear whether the two monomers form the dimer with a head-to-head or a head-to-tail geometry. The first scenario is suggested by the only crystal structure of the AR DBD bound to DNA, while the second one is considered the most probable scenario that takes in account the N/C interaction (236, 253). In the first case, the two monomers are forming a parallel dimer, in the second, an antiparallel dimer. This expands the controversy to the nature of the N/C interaction, whether it involves domains of the same protein (a *cis*-interaction) or of different monomers (a *trans*-interaction).

Our data suggests an enticing hypothesis whereby the N/C interaction that is required for dimerization is actually a *cis*-interaction that leads to the formation of a parallel oligomer. In this case, the interaction is established all along the NTD and is triggered by the interaction between the ligand-bound LBD and the ^{23}FQNLFQSVREVIQ^{36} in α-helical conformation. The low population of NTD with the ^{23}FQNLFQSVREVIQ^{36} motif in helical conformation suggests that this interaction is more probable to occur as a *cis*-interaction, because ligand and binder are tethered together and have more probability to find each other.

The helical ^{23}FQNLFQSVREVIQ^{36} is then stabilized in this conformation and the self-affinity becomes higher, so that another ^{23}FQNLFQSVREVIQ^{36} in this conformation can be recruited. It is therefore very likely that two monomers ligand-bound and ^{23}FQNLF^{28}-bound start to establish the contacts for the dimerization from ^{23}FQNLFQSVREVIQ^{36}, which then nucleates all the other regions attributed to this interaction, described in (De Mol, in preparation).

This scenario clearly needs a considerable amount of data to be proven. However, providing this information is not a straightforward task with the techniques applied so far for studying the N/C interaction (374), the most successful of which has been Förster Resonance Energy Transfer (FRET). Generally, FRET probes are connected to the extremities of the protein, but, in this case, both the N-terminal and C-terminal of each monomer would be in close proximity, as the proteins are folding back upon themselves, producing positive results both for *trans* and for *cis* N/C interactions. Newly developed techniques or a new application of the same techniques need to be developed in order to address this issue.

Finally, it is difficult to find a clear-cut explanation for the function of polyQ. It might play the role of a spacer, preventing hyper-activation of AR, stabilize the dimerization via coil-coil interactions, or favor protein-protein interaction via the same mechanism(103).

10.4 CAN THE CASPASE-3-CLEAVED FRAGMENT BE CONSIDERED THE MINIMAL ETIOLOGIC ELEMENT FOR SBMA?

The findings reported in this project try to demonstrate that is possible to apply a reductionist approach to the biophysical characterization of the AR aggregation in the same way as huntingtin has been characterized in the current literature(307, 346, 349, 375), with the added value of working with a protein fragment that is observed in the cell(376).

The nuclear localization of caspase-3 has been reported as a downstream event of the apoptotic cascade(377, 378), but it is not clear at what stage AR undergoes proteolytic cleavage in an SBMA cell.

It is possible that the cleavage is in reality a relatively late event in the development of the diseases, but still crucial, because it releases a highly toxic species, due to the higher aggregation propensity.

In this picture, the formation of uncontrolled oligomers of AR NTD, in absence of chaperones, is prevented by the interaction with LBD, but an expanded polyQ escapes this control due to increased difficulty to perform the N/C interaction, forming early aggregates, the formation of which the cell cannot dispose. This would activate the apoptotic pathway, which in turn cleaves off the AR N-terminal fragment here characterized, decreasing its affinity to the LBD because it is no more part of the same protein and favoring the self-assembly.

A possible experiment to prove the point that AR N-terminal fragment cleaved by caspase-3 is representative of the AR aggregation, would be to induce the nuclear transport of this fragment with a range of polyQ lengths, by fusing a nuclear localization signal (NLS), and check if there is formation of aggregates and the onset of any toxic phenotype.

10.5 POSSIBLE METHODS FOR INHIBITION OF THE AGGREGATION OF AR IN SBMA

In this project we demonstrated that an already used strategy for blocking the aggregation of a polyQ protein(379) is not effective for AR, probably because of the distance between the nucleating region (^{23}FQNLFQSVREVIQ^{36}) and the polyQ.

Nevertheless, the knowledge generated on the system, both by this project and with the studies carried out by BaharehEftekharzadeh, allowed us to identify what species is toxic and how this species is generated. Fibrils of AR were found in muscle tissue and spinal cord of SBMA mice in advanced state of the diseases, while amorphous aggregates were found in healthy individuals (Eftekharzadeh et al., in preparation). From the current study, it is known that the amorphous aggregates form via an oligomerization process that involves interaction of amphipaticα-helices, while the fibrils are formed as a consequence of the α-to-β transition of the polyQ.

Therefore, any approach that prevents the formation or disposes of the fibrils has good potential to work as a remedy for this disease.

Disposing of the aggregates has been attempted by improving the functions of the proteasome complex(175, 239), whereas preventing the formation of the fibrils can be achieved by blocking one of the two transitions.

Increasing the chaperone activity would target the unfolded intermediate, isolating the monomeric protein, and trials in this direction have been already implemented (380–382). However, we propose a strategy that aims directly at the α-helical intermediate and prevents the α-to-ß transition, maintaining the aggregates in condition to be processed by the proteasome complex, without thwarting the whole physiology of the cell.

The most straightforward approach would be to design conformational-specific antibodies or nanobodies that selectively bind to the α-helical polyQ. To do this, we could take advantage of the plasticity of the polyQ, by further increasing the number of Leu residues up to a maximum value, which would force the whole polyQ in α-helical conformation.

Agadir simulation already predicted successfully the effect of the Leu-rich motif deletion from 25Q, so it is a good tool for screening what number of Leu residues is the best match for stabilizing the α-helix. From figure 10.2 emerges clearly that just by adding 1 residue to the Leu-rich motif, the helicity increases of a factor 3 and the highest values of helicity are reached by doubling the number of Leu residues, so that, according to the simulation, the polyQ should be completely folded as a α-helix with a total number of 7-8 Leu residues and could be used easily enough to screen for antibodies aimed against this conformation.

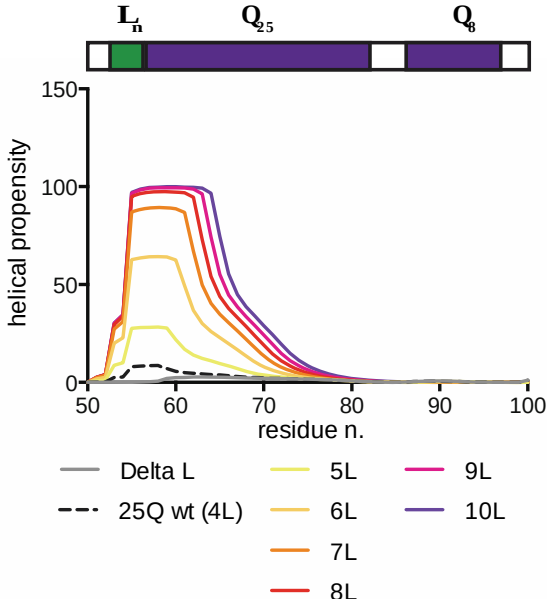


Figure 10.2 Agadir prediction of helicity of 25Q with variable number of Leu residues in ^{55}LLLL^{59}. ΔL is showed in gray, 25Q in dotted black while the increase of the number of Leu residues follows a gradient from yellow to violet. The region showed in the prediction corresponds to the part of the protein directly involved in the change of conformation and its close surroundings, represented in the cartoonish representation on top.

Similar approaches have been already attempted with the IAPP protein, where insulin have been found to interact with the helical intermediate species and stabilize this conformation, preventing the progression to fibrils(354, 383).

In any case, whether this is a viable strategy or not, the knowledge produced with this project opens new possibilities for treatment, not only of SBMA, but also for developing therapies for other polyQ diseases, thanks to the insights in the common mechanisms.

10.6 THINKING OUTSIDE BIOMEDICINE: CAN THE PROPERTIES OF POLYQ PEPTIDES AND PROTEINS HAVE OTHER APPLICATIONS?

The emerging field of synthetic biology has provided new perspectives for protein engineering, by extensively applying it in biotechnological research. The field of protein engineering has produced interesting studies on amyloid fibrils decorated with porphyrinic groups with the aim of developing biocompatible nano-wires(384), or decorated with mussel

foot protein (Msfp, used by *Mytilusgalloprovincialis* for tethering its shell to the sea floor) for developing self-assembling adhesive nanofibers that work in aqueous conditions(385).

Also, several studies have been published about the assembly of domains with known functions to perform new tasks. Interesting applications have already been discussed for the Src Homology 3 (SH3) domain(386) and coil-coil-prone domains (387) for designing artificially interacting proteins.

The coil-coil system proved particularly versatile for generating quaternary structures of increasing complexity, from oligomers with a controlled geometry(388) up to an organized lattice of coil-coil fibers(389).

On this topic, there is increasing interest in understanding how the polyQ tracts modulate coil-coil interactions, as these motifs are often found close to regions with high coil-coil propensity(105, 194). The inclusion of polyQ tracts in the protein engineering toolbox may provide a system for producing libraries of coil-coil interactors with finely regulated strengths.

Furthermore, we demonstrated the plasticity of the secondary structure of the polyQ tract and believe that this property may have applications for the production of linkers with secondary structure that can be tuned in response to changes in the environment. The Glu side chain, for instance, presents different propensity to form α-helix with pH dependence. Knowing that polyQ tracts acquire, to some extent, the secondary structure of their flanking regions, it would be possible to design polyQ proteins with properties controlledby changes in pH, like fibrils that are irreversibly deposited only upon a determined stimulus (pH change, in this case). In the same way, Ser residues placed in a strategic position of the flanking region can induce the formation of the helix upon phosphorylation, which then extends on the polyQ tract.

A similar study has recently been carried out on a *de novo* designed IDP, tailored to tether on a silica surface and protrude in the environment as entropic springs with variable excursion depending on pH and ionic strength, with the aim of coating surfaces for reducing non-specific interactions with cells or other proteins and exposing specific epitopes (390).

These examples of application are unlikely to be implemented in the near future, but further characterization of polyQ tracts might help to find a use for this motif in the protein engineering toolbox.

CHAPTER 11: CONCLUSIONS

Also in this case, the conclusions will be split in two blocks: methodological and scientific conclusions.

11.1 METHODOLOGICAL CONCLUSIONS

1. Three different recombinant proteins were designed corresponding to the fragment of AR cleaved by caspase-3. The three proteins were produced with a vestigial number of Gln (4Q), a physiologic number of Gln (25Q) and a pathogenic number of Gln (51Q). The three fragments, together with the 8 mutants of 25Q developed, were successfully expressed and purified with yields sufficiently high for sustaining the experimental requirements.

2. A method for reproducible disaggregation was implemented from previously available methodological approaches and adapted to the requirements of the AR recombinant fragments. The effectiveness of this method was tested with different biophysical techniques. This new methodology is described in an article currently in preparation.

3. A set of biophysical tools was adapted for studying the effect on aggregation of different lengths of the polyQ tract and mutation on the sequence of AR.

11.2 SCIENTIFIC CONCLUSIONS

1. The secondary structure of 25Q and 4Q was characterized in detail, producing high-resolution data on the secondary structure of the polyQ tract. The polyQ tract is in α-helical conformation, contrarily to what is reported for other polyQ proteins in the literature.

2. The residues that induce the helicity in the polyQ tract were identified as an ^{54}LLLL^{58} motif directly N-terminal to the polyQ tract. This motif substantially decreases the aggregation propensity of the AR fragment.

The observations of conclusions 1 and 2 of this block resulted in an article currently under review process.

3. The aggregation mechanism of the AR recombinant proteins was characterized. These proteins aggregate in a multistep process, where the nucleation is mediated by a region N-terminal and distant from the polyQ, ^{23}FQNLFQSVREVIQ^{36}. This region is prone to aggregation and forms amorphous aggregates, whereas the polyQ tract forms fibrils. Therefore, the AR recombinant proteins aggregate with a complex mechanism where ^{23}FQNLFQSVREVIQ^{36}and the polyQ tract act independently, producing amorphous aggregates or fibrils depending on the length of the tract. The rate of maturation of the aggregates to fibrils depends on the length of the polyQ tract.

4. The early oligomerization of 25Q was studied in detail, especially in relation to the ^{23}FQNLFQSVREVIQ^{36} motif. The key feature of the formation of the oligomer is the helical content of this motif. An increase in helical content causes a strong increase in the

aggregation rate of mutant 25Q, whereas perturbing the hydrophobicity of the motif does not decrease the aggregation rate of the protein. The key residues in the interaction were recognized as the Gln positions 24 and 28. These observations allowed us to propose that the nucleation of this protein is mediated by the Gln residues organized in a line along the axis of the α-helix, to form a 'polar spine', which interact with the 'polar spines' of other monomers by forming a polar zipper.

The observations of points 3 and 4 are currently material for an article in preparation.

BIBLIOGRAPHY

1. Soto C (2003) Unfolding the role of protein misfolding in neurodegenerative diseases. *Nat Rev Neurosci* 4(1):49–60.

2. Masters CL, Selkoe DJ (2012) Biochemistry of amyloid beta-protein and amyloid deposits in Alzheimer disease. *Cold Spring Harb Perspect Med* 2:a006262.

3. Haass C, Kaether C, Thinakaran G, Sisodia S (2012) Trafficking and proteolytic processing of APP. *Cold Spring Harb Perspect Med* 2:1–26.

4. Nussbaum JM, Seward ME, Bloom GS (2013) Alzheimer disease: a tale of two prions. *Prion* 7(1):14–9.

5. Arima K, et al. (1999) Cellular co-localization of phosphorylated tau- and NACP/alpha-synuclein-epitopes in lewy bodies in sporadic Parkinson's disease and in dementia with Lewy bodies. *Brain Res* 843:53–61.

6. Uversky VN, Eliezer D (2009) Biophysics of Parkinson's disease: structure and aggregation of alpha-synuclein. *Curr Protein Pept Sci* 10(5):483–499.

7. Safar JG (2012) Molecular pathogenesis of sporadic prion diseases in man. *Prion* 6:108–115.

8. Soto C, Estrada L, Castilla J (2006) Amyloids, prions and the inherent infectious nature of misfolded protein aggregates. *Trends Biochem Sci* 31(3):150–5.

9. Brundin P, Melki R, Kopito R (2010) Prion-like transmission of protein aggregates in neurodegenerative diseases. *Nat Rev Mol Cell Biol* 11:301–307.

10. Sipe JD, et al. (2010) Amyloid fibril protein nomenclature: 2010 recommendations from the nomenclature committee of the International Society of Amyloidosis. *Amyloid* 17:101–104.

11. Chiti F, Dobson CM (2006) Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem* 75:333–366.

12. Eisenberg D, Jucker M (2012) The amyloid state of proteins in human diseases. *Cell* 148(6):1188–203.

13. Sunde M, et al. (1997) Common core structure of amyloid fibrils by synchrotron X-ray diffraction. *J Mol Biol* 273(3):729–739.

14. Nelson R, et al. (2005) Structure of the cross-beta spine of amyloid-like fibrils. *Nature* 435:773–778.

15. Lu J-X, et al. (2013) Molecular structure of β-amyloid fibrils in Alzheimer's disease brain tissue. *Cell* 154(6):1257–68.

16. Tycko R (2011) Solid-state NMR studies of amyloid fibril structure. *Annu Rev Phys Chem* 62:279–99.

17. Jiménez JL, et al. (1999) Cryo-electron microscopy structure of an SH3 amyloid fibril and model of the molecular packing. *EMBO J* 18:815–821.

18. Schmidt M, et al. (2009) Comparison of Alzheimer Abeta(1-40) and Abeta(1-42) amyloid fibrils reveals similar protofilament structures. *Proc Natl Acad Sci U S A* 106:19813–19818.

19. Sawaya MR, et al. (2007) Atomic structures of amyloid cross-beta spines reveal varied steric zippers. *Nature* 447(7143):453–457.

20. Makin OS, Atkins E, Sikorski P, Johansson J, Serpell LC (2005) Molecular basis for amyloid fibril formation and stability. *Proc Natl Acad Sci U S A* 102(2):315–320.

21. Wetzel R (2006) Kinetics and thermodynamics of amyloid fibril assembly. *Acc Chem Res* 39(9):671–9.

22. O'Nuallain B, Shivaprasad S, Kheterpal I, Wetzel R (2005) Thermodynamics of A beta(1-40) amyloid fibril elongation. *Biochemistry* 44:12709–12718.

23. Bemporad F, De Simone A, Chiti F, Dobson CM (2012) Characterizing intermolecular interactions that initiate native-like protein aggregation. *Biophys J* 102(11):2595–604.

24. Calloni G, et al. (2008) Structure and dynamics of a partially folded protein are decoupled from its mechanism of aggregation. *J Am Chem Soc* 130(39):13040–50.

25. Baldwin AJ, et al. (2011) Metastability of native proteins and the phenomenon of amyloid formation. *J Am Chem Soc* 133(36):14160–14163.

26. Greenwald J, Riek R (2010) Biology of amyloid: structure, function, and regulation. *Structure* 18(10):1244–60.

27. Greenwald J, Riek R (2012) On the possible amyloid origin of protein folds. *J Mol Biol* 421(4-5):417–26.

28. Maurer-Stroh S, et al. (2010) Exploring the sequence determinants of amyloid structure using position-specific scoring matrices. *Nat Methods* 7(3):237–42.

29. Rousseau F, Schymkowitz J, Serrano L (2006) Protein aggregation and amyloidosis: confusion of the kinds? *Curr Opin Struct Biol* 16(1):118–26.

30. Riek R (2014) The Presence of an Air − Water Interface A ff ects Formation and Elongation of α - Synuclein Fibrils.

31. Hellstrand E, Boland B, Walsh DM, Linse S (2010) Amyloid β-protein aggregation produces highly reproducible kinetic data and occurs by a two-phase process. *ACS Chem Neurosci* 1(1):13–8.

32. Cheon M, et al. (2007) Structural reorganisation and potential toxicity of oligomeric species formed during the assembly of amyloid fibrils. *PLoS Comput Biol* 3(9):1727–38.

33. Knowles TPJ, et al. (2009) An analytical solution to the kinetics of breakable filament assembly. *Science* 326(5959):1533–7.

34. Cohen SI a, Vendruscolo M, Dobson CM, Knowles TPJ (2012) From macroscopic measurements to microscopic mechanisms of protein aggregation. *J Mol Biol* 421(2-3):160–71.

35. Cohen SI a, et al. (2013) Proliferation of amyloid-β42 aggregates occurs through a secondary nucleation mechanism. *Proc Natl Acad Sci U S A* 110(24):9758–63.

36. Berthelot K, Cullin C, Lecomte S (2013) What does make an amyloid toxic: morphology, structure or interaction with membrane? *Biochimie* 95(1):12–9.

37. Stroud JC, Liu C, Teng PK, Eisenberg D (2012) Toxic fibrillar oligomers of amyloid-β have cross-β structure. *Proc Natl Acad Sci U S A* 109(20):7717–22.

38. Glabe CG (2008) Structural classification of toxic amyloid oligomers. *J Biol Chem* 283(44):29639–43.

39. Von Bergen M, Barghorn S, Biernat J, Mandelkow E-M, Mandelkow E (2005) Tau aggregation is driven by a transition from random coil to beta sheet structure. *Biochim Biophys Acta* 1739(2-3):158–66.

40. Arnaudov LN, de Vries R (2005) Thermally Induced Fibrillar Aggregation of Hen Egg white Lysozyme. *Biophys J* 88(1):515–526.

41. Xu M, et al. (2007) The first step of hen egg white lysozyme fibrillation, irreversible partial unfolding, is a two-state transition. *Protein Sci* 16(5):815–832.

42. Colon W, Kelly JW (1992) Partial denaturation of transthyretin is sufficient for amyloid fibril formation in vitro. *Biochemistry* 31(36):8654–8660.

43. Mulligan VK, et al. (2012) Early steps in oxidation-induced SOD1 misfolding: implications for non-amyloid protein aggregation in familial ALS. *J Mol Biol* 421(4-5):631–52.

44. Vignaud H, et al. (2013) A structure-toxicity study of Aß42 reveals a new anti-parallel aggregation pathway. *PLoS One* 8(11). doi:10.1371/journal.pone.0080262.

45. Armen RS, Daggett V (2005) Characterization of two distinct beta2-microglobulin unfolding intermediates that may lead to amyloid fibrils of different morphology. *Biochemistry* 44(49):16098–107.

46. Abedini A, Raleigh DP (2009) A critical assessment of the role of helical intermediates in amyloid formation by natively unfolded proteins and polypeptides. *Protein Eng Des Sel* 22(8):453–459.

47. Duyao M, et al. (1993) Trinucleotide repeat length instability and age of onset in Huntington's disease. *Nat Genet* 4:387–392.

48. Sequeiros J, Seneca S, Martindale J (2010) Consensus and controversies in best practices for molecular genetic testing of spinocerebellar ataxias. *Eur J Hum Genet* 18:1188–1195.

49. Nakamura K, et al. (2001) SCA17, a novel autosomal dominant cerebellar ataxia caused by an expanded polyglutamine in TATA-binding protein. *Hum Mol Genet* 10(14):1441–8.

50. Tsuji S (2003) Dentatorubral-Pallidoluysian Atrophy (DRPLA). *Genetics of Movement Disorders*, pp 143–150.

51. La Spada AR, Wilson EM, Lubahn DB, Harding AE, Fischbeck KH (1991) Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature* 352:77–79.

52. Zoghbi HY (2006) Trinucleotide repeat disorders. *Principles of Molecular Medicine*, pp 1114–1122.

53. Robertson AL, Bate M a, Androulakis SG, Bottomley SP, Buckle AM (2011) PolyQ: a database describing the sequence and domain context of polyglutamine repeats in proteins. *Nucleic Acids Res* 39(Database issue):D272–6.

54. Robertson a L, Bottomley SP (2010) Towards the treatment of polyglutamine diseases: the modulatory role of protein context. *Curr Med Chem* 17(27):3058–68.

55. La Spada AR, Taylor JP (2010) Repeat expansion disease: progress and puzzles in disease pathogenesis. *Nat Rev Genet* 11:247–258.

56. Mirkin SM (2007) Expandable DNA repeats and human disease. *Nature* 447:932–940.

57. Pan XF (2006) Mechanism of trinucleotide repeats instabilities: The necessities of repeat non-B secondary structure formation and the roles of cellular trans-acting factors. *Acta Genet Sin* 33:1–11.

58. Freudenreich CH, Kantrow SM, Zakian VA (1998) Expansion and length-dependent fragility of CTG repeats in yeast. *Science* 279:853–856.

59. Kang S, Jaworski A, Ohshima K, Wells RD (1995) Expansion and deletion of CTG repeats from human disease genes are determined by the direction of replication in E. coli. *Nat Genet* 10:213–218.

60. McMurray CT (2010) Mechanisms of trinucleotide repeat instability during human development. *Nat Rev Genet* 11(11):786–99.

61. Ashley CT, Warren ST (1995) Trinucleotide repeat expansion and human disease. *Annu Rev Genet* 29:703–728.

62. Yoon S-R, Dubeau L, de Young M, Wexler NS, Arnheim N (2003) Huntington disease expansion mutations in humans can occur before meiosis is completed. *Proc Natl Acad Sci U S A* 100:8834–8838.

63. Gonitel R, et al. (2008) DNA instability in postmitotic neurons. *Proc Natl Acad Sci U S A* 105:3467–3472.

64. Swami M, et al. (2009) Somatic expansion of the Huntington's disease CAG repeat in the brain is associated with an earlier age of disease onset. *Hum Mol Genet* 18:3039–3047.

65. Kennedy L, et al. (2003) Dramatic tissue-specific mutation length increases are an early molecular event in Huntington disease pathogenesis. *Hum Mol Genet* 12:3359–3367.

66. De Mezer M, Wojciechowska M, Napierala M, Sobczak K, Krzyzosiak WJ (2011) Mutant CAG repeats of Huntingtin transcript fold into hairpins, form nuclear foci and are targets for RNA interference. *Nucleic Acids Res* 39(9):3852–63.

67. Bañez-Coronel M, et al. (2012) A pathogenic mechanism in Huntington's disease involves small CAG-repeated RNAs with neurotoxic activity. *PLoS Genet* 8(2):e1002481.

68. Nalavade R, Griesche N, Ryan DP, Hildebrand S, Krauss S (2013) Mechanisms of RNA-induced toxicity in CAG repeat disorders. *Cell Death Dis* 4(8):e752.

69. Wetzel R (2012) Physical chemistry of polyglutamine: Intriguing tales of a monotonous sequence. *J Mol Biol* 421:466–490.

70. Perutz MAXF, Johnson T, Suzuki M, Finch JT (1994) polar zippers: 91(June):5355–5358.

71. Crick SL, Jayaraman M, Frieden C, Wetzel R, Pappu R V (2006) Fluorescence correlation spectroscopy shows that monomeric polyglutamine molecules form collapsed structures in aqueous solutions. *Proc Natl Acad Sci U S A* 103:16764–16769.

72. Walters RH, Murphy RM (2009) Examining polyglutamine peptide length: a connection between collapsed conformations and increased aggregation. *J Mol Biol* 393(4):978–92.

73. Peters-Libeu C, et al. (2012) Disease-associated polyglutamine stretches in monomeric huntingtin adopt a compact structure. *J Mol Biol* 421(4-5):587–600.

74. Dougan L, Li J, Badilla CL, Berne BJ, Fernandez JM (2009) Single homopolypeptide chains collapse into mechanically rigid conformations. *Proc Natl Acad Sci U S A* 106(31):12605–10.

75. Zhou Z-L, et al. (2011) The possible structural models for polyglutamine aggregation: a molecular dynamics simulations study. *J Biomol Struct Dyn* 28(5):743–58.

76. Laghaei R, Mousseau N (2010) Spontaneous formation of polyglutamine nanotubes with molecular dynamics simulations. *J Chem Phys* 132(16):165102.

77. Rossetti G, Magistrato A, Pastore A, Persichetti F, Carloni P (2008) Structural properties of polyglutamine aggregates investigated via molecular dynamics simulations. *J Phys Chem B* 112(51):16843–50.

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

79. Klein FAC, et al. (2007) Pathogenic and Non-pathogenic Polyglutamine Tracts Have Similar Structural Properties: Towards a Length-dependent Toxicity Gradient. *J Mol Biol* 371:235–244.

80. Digambaranath JL, et al. (2011) An accurate model of polyglutamine. *Proteins* 79(5):1427–40.

81. Heck BS, Doll F, Hauser K (2014) Length-dependent conformational transitions of polyglutamine repeats as molecular origin of fibril initiation. *Biophys Chem* 185:47–57.

82. Masino L, Kelly G, Leonard K, Trottier Y, Pastore A (2002) Solution structure of polyglutamine tracts in GST-polyglutamine fusion proteins. *FEBS Lett* 513(2-3):267–72.

83. Masino L, et al. (2003) Domain architecture of the polyglutamine protein ataxin-3: a globular domain followed by a flexible tail. *FEBS Lett* 549(1-3):21–25.

84. Hoop CL, et al. (2014) Polyglutamine amyloid core boundaries and flanking domain dynamics in huntingtin fragment fibrils determined by solid-state nuclear magnetic resonance. *Biochemistry* 53(42):6653–66.

85. Chen S, Berthelier V, Hamilton JB, O'Nuallain B, Wetzel R (2002) Amyloid-like features of polyglutamine aggregates and their assembly kinetics. *Biochemistry* 41(23):7391–9.

86. Kar K, Jayaraman M, Sahoo B, Kodali R, Wetzel R (2011) Critical nucleus size for disease-related polyglutamine aggregation is repeat-length dependent. *Nat Struct Mol Biol* 18(3):328–36.

87. Landrum E, Wetzel R (2014) Biophysical underpinnings of the repeat length dependence of polyglutamine amyloid formation. *J Biol Chem* 289(15):10254–60.

88. Bhattacharyya AM, Thakur AK, Wetzel R (2005) Polyglutamine Aggregation Nucleation: Thermodynamics of a Highly Unfavorable Protein Folding Reaction. *Proc Natl Acad Sci U S A* 102(43):15400–5.

89. Kar K, et al. (2013) B-Hairpin-Mediated Nucleation of Polyglutamine Amyloid Formation. *J Mol Biol* 425(7):1183–97.

90. Poirier M a., et al. (2002) Huntingtin spheroids and protofibrils as precursors in polyglutamine fibrilization. *J Biol Chem* 277(43):41032–41037.

91. Darnell G, Orgel JPRO, Pahl R, Meredith SC (2007) Flanking polyproline sequences inhibit beta-sheet structure in polyglutamine segments by inducing PPII-like helix structure. *J Mol Biol* 374(3):688–704.

92. Duennwald ML, Jagadish S, Muchowski PJ, Lindquist S (2006) Flanking sequences profoundly alter polyglutamine toxicity in yeast. *Proc Natl Acad Sci U S A* 103(29):11045–50.

93. Apostol BL, et al. (2003) A cell-based assay for aggregation inhibitors as therapeutics of polyglutamine-repeat disease and validation in Drosophila. *Proc Natl Acad Sci U S A* 100(10):5950–5955.

94. Kim MW, Chelliah Y, Kim SW, Otwinowski Z, Bezprozvanny I (2009) Secondary Structure of Huntingtin Amino-Terminal Region. *Structure* 17(9):1205–1212.

95. Thakur AK, 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.

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

97. Jayaraman M, et al. (2012) Slow amyloid nucleation via α-helix-rich oligomeric intermediates in short polyglutamine-containing huntingtin fragments. *J Mol Biol* 415(5):881–99.

98. Masino L, 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):1021–1035.

99. Ellisdon AM, Thomas B, Bottomley SP (2006) The two-stage pathway of ataxin-3 fibrillogenesis involves a polyglutamine-independent step. *J Biol Chem* 281(25):16888–96.

100. Robertson AL, Bate M a, Buckle AM, Bottomley SP (2011) The rate of polyQ-mediated aggregation is dramatically affected by the number and location of surrounding domains. *J Mol Biol* 413(4):879–87.

101. Tobelmann MD, Murphy RM (2011) Location trumps length: polyglutamine-mediated changes in folding and aggregation of a host protein. *Biophys J* 100(11):2773–82.

102. Scarafone N, et al. (2012) Amyloid-like fibril formation by polyQ proteins: a critical balance between the polyQ length and the constraints imposed by the host protein. *PLoS One* 7(3):e31253.

103. Fiumara F, Fioriti L, Kandel ER, Hendrickson W a (2010) Essential role of coiled coils for aggregation and activity of Q/N-rich prions and PolyQ proteins. *Cell* 143(7):1121–35.

104. Ramazzotti M, Monsellier E, Kamoun C, Degl'Innocenti D, Melki R (2012) Polyglutamine repeats are associated to specific sequence biases that are conserved among eukaryotes. *PLoS One* 7(2):e30824.

105. Schaefer MH, Wanker EE, Andrade-Navarro MA (2012) Evolution and function of CAG/polyglutamine repeats in protein-protein interaction networks. *Nucleic Acids Res* 40(10):4273–87.

106. Butland SL, et al. (2007) CAG-encoded polyglutamine length polymorphism in the human genome. *BMC Genomics* 8:126.

107. Duennwald ML, Jagadish S, Giorgini F, Muchowski PJ, Lindquist S (2006) A network of protein interactions determines polyglutamine toxicity. *Proc Natl Acad Sci U S A* 103(29):11051–6.

108. Abou-Sleymane G, et al. (2006) Polyglutamine expansion causes neurodegeneration by altering the neuronal differentiation program. *Hum Mol Genet* 15(5):691–703.

109. Kumar R (2012) Role of androgen receptor polyQ chain elongation in Kennedy's disease and use of natural osmolytes as potential therapeutic targets. *IUBMB Life* 64(11):879–84.

110. Hipp MS, Park SH, Hartl FU (2014) Proteostasis impairment in protein-misfolding and -aggregation diseases. *Trends Cell Biol* 24(9):506–514.

111. Gidalevitz T, Ben-Zvi A, Ho KH, Brignull HR, Morimoto RI (2006) Progressive disruption of cellular protein folding in models of polyglutamine diseases. *Science* 311(5766):1471–4.

112. Morley JF, Brignull HR, Weyers JJ, Morimoto RI (2002) The threshold for polyglutamine-expansion protein aggregation and cellular toxicity is dynamic and influenced by aging in Caenorhabditis elegans. *Proc Natl Acad Sci U S A* 99(16):10417–10422.

113. David DC, et al. (2010) Widespread protein aggregation as an inherent part of aging in C. elegans. *PLoS Biol* 8(8):47–48.

114. Horwich AL, Fenton WA, Chapman E, Farr GW (2007) Two families of chaperonin: physiology and mechanism. *Annu Rev Cell Dev Biol* 23:115–45.

115. Haslbeck M, Franzmann T, Weinfurtner D, Buchner J (2005) Some like it hot: the structure and function of small heat-shock proteins. *Nat Struct Mol Biol* 12(10):842–846.

116. Kampinga HH, Craig E a (2010) The HSP70 chaperone machinery: J proteins as drivers of functional specificity. *Nat Rev Mol Cell Biol* 11(8):579–592.

117. Taipale M, Jarosz DF, Lindquist S (2010) HSP90 at the hub of protein homeostasis: emerging mechanistic insights. *Nat Rev Mol Cell Biol* 11(7):515–528.

118. Arndt V, et al. (2010) Chaperone-Assisted Selective Autophagy Is Essential for Muscle Maintenance. *Curr Biol* 20(2):143–148.

119. Saibil H (2013) Chaperone machines for protein folding, unfolding and disaggregation. *Nat Rev Mol Cell Biol* 14(10):630–42.

120. Yao T-P (2010) The role of ubiquitin in autophagy-dependent protein aggregate processing. *Genes Cancer* 1(7):779–786.

121. Komander D, Rape M (2012) The Ubiquitin Code. *Annu Rev Biochem* 81(1):203–229.

122. Kettern N, Dreiseidler M, Tawo R, Höhfeld J (2010) Chaperone-assisted degradation: Multiple paths to destruction. *Biol Chem* 391(5):481–489.

123. Kaushik S, Cuervo AM (2008) Chaperone-mediated autophagy. *Methods Mol Biol* 445:227–244.

124. López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G (2013) The hallmarks of aging. *Cell* 153(6):1194–217.

125. Tonoki A, et al. (2009) Genetic evidence linking age-dependent attenuation of the 26S proteasome with the aging process. *Mol Cell Biol* 29(4):1095–1106.

126. Fargnoli J, Kunisada T, Fornace AJ, Schneider EL, Holbrook NJ (1990) Decreased expression of heat shock protein 70 mRNA and protein after heat treatment in cells of aged rats. *Proc Natl Acad Sci U S A* 87(2):846–50.

127. Pahlavani MA, Harris MD, Moore SA, Weindruch R, Richardson A (1995) The expression of heat shock protein 70 decreases with age in lymphocytes from rats and rhesus monkeys. *Exp Cell Res* 218(1):310–8.

128. Hall DM, et al. (2000) Aging reduces adaptive capacity and stress protein expression in the liver after heat stress. *J Appl Physiol* 89(2):749–759.

129. Venkatraman P, Wetzel R, Tanaka M, Nukina N, Goldberg AL (2004) Eukaryotic proteasomes cannot digest polyglutamine sequences and release them during degradation of polyglutamine-containing proteins. *Mol Cell* 14(1):95–104.

130. Hipp MS, et al. (2012) Indirect inhibition of 26S proteasome activity in a cellular model of Huntington's disease. *J Cell Biol* 196(5):573–87.

131. Sittler A, et al. (2001) Geldanamycin activates a heat shock response and inhibits huntingtin aggregation in a cell culture model of Huntington's disease. *Hum Mol Genet* 10(12):1307–1315.

132. Sarkar S, Ravikumar B, Floto RA, Rubinsztein DC (2009) Rapamycin and mTOR-independent autophagy inducers ameliorate toxicity of polyglutamine-expanded huntingtin and related proteinopathies. *Cell Death Differ* 16(1):46–56.

133. Lee B-H, et al. (2010) Enhancement of proteasome activity by a small-molecule inhibitor of USP14. *Nature* 467(7312):179–184.

134. Papapetropoulos T, Panayiotopoulos CP (1981) *X-linked spinal and bulbar muscular atrophy of late onset (Kennedy-Stefanis disease?).*

135. Sułek A, et al. (2005) CAG repeat polymorphism in the androgen receptor (AR) gene of SBMA patients and a control group. *J Appl Genet* 46(2):237–239.

136. Rhodes LE, et al. (2009) Clinical features of spinal and bulbar muscular atrophy. *Brain* 132(Pt 12):3242–51.

137. Katsuno M, et al. (2006) Pathogenesis, animal models and therapeutics in Spinal and bulbar muscular atrophy (SBMA). *Exp Neurol* 200(1):8–18.

138. Tabrizi S, Beal MF, Lang AE, Ludolph AC (2005) Neurodegenerative Diseases: Neurobiology, Pathogenesis and Therapeutics. *J Neurol Neurosurg Psychiatry* 77(2):284.

139. Tanaka F, et al. (1996) Founder effect in spinal and bulbar muscular atrophy (SBMA). *Hum Mol Genet* 5(9):1253–1257.

140. Schmidt BJ, Greenberg CR, Allingham-Hawkins DJ, Spriggs EL (2002) *Expression of X-linked bulbospinal muscular atrophy (Kennedy disease) in two homozygous women.*

141. Pradat P-F (2014) [SBMA: A rare disease but a classic ALS mimic syndrome]. *Presse Med* 43(5):580–6.

142. Hama T, et al. (2012) Discrimination of spinal and bulbar muscular atrophy from amyotrophic lateral sclerosis using sensory nerve action potentials. *Muscle and Nerve* 45(2):169–174.

143. Chahin N, Klein C, Mandrekar J, Sorenson E (2008) Natural history of spinal-bulbar muscular atrophy. *Neurology* 70(21):1967–1971.

144. Bruson a, et al. (2012) CAG repeat length in androgen receptor gene is not associated with amyotrophic lateral sclerosis. *Eur J Neurol* 19(10):1373–5.

145. Katsuno M, Adachi H, Inukai A, Sobue G (2003) Transgenic mouse models of spinal and bulbar muscular atrophy (SBMA). *Cytogenet Genome Res* 100(1-4):243–251.

146. Sorarù G, et al. (2008) Spinal and bulbar muscular atrophy: Skeletal muscle pathology in male patients and heterozygous females. *J Neurol Sci* 264(1-2):100–105.

147. Li M, et al. (1998) Nuclear inclusions of the androgen receptor protein in spinal and bulbar muscular atrophy. *Ann Neurol* 44(2):249–54.

148. Li M, et al. (1998) Nonneural nuclear inclusions of androgen receptor protein in spinal and bulbar muscular atrophy. *Am J Pathol* 153(3):695–701.

149. Brooks BP, et al. (1997) Characterization of an expanded glutamine repeat androgen receptor in a neuronal cell culture system. *Neurobiol Dis* 3(4):313–23.

150. Adachi H, et al. (2005) Widespread nuclear and cytoplasmic accumulation of mutant androgen receptor in SBMA patients. *Brain* 128(Pt 3):659–70.

151. Chevalier-Larsen ES, et al. (2004) Castration restores function and neurofilament alterations of aged symptomatic males in a transgenic mouse model of spinal and bulbar muscular atrophy. *J Neurosci* 24(20):4778–86.

152. Li M, Chevalier-Larsen ES, Merry DE, Diamond MI (2007) Soluble androgen receptor oligomers underlie pathology in a mouse model of spinobulbar muscular atrophy. *J Biol Chem* 282(5):3157–64.

153. Katsuno M, et al. (2003) Leuprorelin rescues polyglutamine-dependent phenotypes in a transgenic mouse model of spinal and bulbar muscular atrophy. *Nat Med* 9(6):768–773.

154. Renier KJ, et al. (2014) Antiandrogen flutamide protects male mice from androgen-dependent toxicity in three models of spinal bulbar muscular atrophy. *Endocrinology* 155(7):2624–2634.

155. Monks DA, et al. (2007) Overexpression of wild-type androgen receptor in muscle recapitulates polyglutamine disease. *Proc Natl Acad Sci U S A* 104(46):18259–64.

156. Kai L, Levenson AS (2011) Combination of resveratrol and antiandrogen flutamide has synergistic effect on androgen receptor inhibition in prostate cancer cells. *Anticancer Res* 31(10):3323–3330.

157. Feldman BJ, Feldman D (2001) The development of androgen-independent prostate cancer. *Nat Rev Cancer* 1(1):34–45.

158. Chevalier-Larsen ES, Merry DE (2012) Testosterone treatment fails to accelerate disease in a transgenic mouse model of spinal and bulbar muscular atrophy. *Dis Model Mech* 5(1):141–5.

159. Jung J, Bonini N (2007) CREB-binding protein modulates repeat instability in a Drosophila model for polyQ disease. *Science* 315(5820):1857–1859.

160. Gunawardena S, et al. (2003) Disruption of axonal transport by loss of huntingtin or expression of pathogenic polyQ proteins in Drosophila. *Neuron* 40(1):25–40.

161. Higashiyama H, et al. (2002) Identification of ter94, Drosophila VCP, as a modulator of polyglutamine-induced neurodegeneration. *Cell Death Differ* 9(3):264–273.

162. Nedelsky NB, et al. (2010) Native functions of the androgen receptor are essential to pathogenesis in a Drosophila model of spinobulbar muscular atrophy. *Neuron* 67(6):936–52.

163. Palazzolo I, et al. (2008) The role of the polyglutamine tract in androgen receptor. *J Steroid Biochem Mol Biol* 108(3-5):245–53.

164. Ranganathan S, et al. (2009) Mitochondrial abnormalities in spinal and bulbar muscular atrophy. *Hum Mol Genet* 18(1):27–42.

165. Orr HT (2012) Polyglutamine neurodegeneration: expanded glutamines enhance native functions. *Curr Opin Genet Dev* 22(3):251–5.

166. Walcott JL, Merry DE (2002) Ligand promotes intranuclear inclusions in a novel cell model of spinal and bulbar muscular atrophy. *J Biol Chem* 277(52):50855–9.

167. Malena A, et al. (2013) Androgen-dependent impairment of myogenesis in spinal and bulbar muscular atrophy. *Acta Neuropathol* 126(1):109–121.

168. Katsuno M, et al. (2012) Pathogenesis and therapy of spinal and bulbar muscular atrophy (SBMA). *Prog Neurobiol* 99(3):246–256.

169. Grunseich C, et al. (2014) Stem cell-derived motor neurons from spinal and bulbar muscular atrophy patients. *Neurobiol Dis* 70:12–20.

170. Dossena M, et al. (2014) Human Adipose-Derived Mesenchymal Stem Cells as a New Model of Spinal and Bulbar Muscular Atrophy. *PLoS One* 9(11):e112746.

171. Hackam AS, et al. (1998) The influence of huntingtin protein size on nuclear localization and cellular toxicity. *J Cell Biol* 141(5):1097–1105.

172. Martín-Aparicio E, et al. (2001) Proteasomal-dependent aggregate reversal and absence of cell death in a conditional mouse model of Huntington's disease. *J Neurosci* 21(22):8772–8781.

173. Jochum T, et al. (2012) Toxic and non-toxic aggregates from the SBMA and normal forms of androgen receptor have distinct oligomeric structures. *Biochim Biophys Acta* 1822(6):1070–8.

174. Young JE, et al. (2009) NIH Public Access. 29(7):1987–1997.

175. Rusmini P, et al. (2007) Aggregation and proteasome: the case of elongated polyglutamine aggregation in spinal and bulbar muscular atrophy. *Neurobiol Aging* 28(7):1099–111.

176. Schmidt M, Finley D (2013) Regulation of proteasome activity in health and disease. *Biochim Biophys Acta*. doi:10.1016/j.bbamcr.2013.08.012.

177. Oddo S (2008) The ubiquitin-proteasome system in Alzheimer's disease. *J Cell Mol Med* 12(2):363–373.

178. Lam YA, et al. (2000) Inhibition of the ubiquitin-proteasome system in Alzheimer's disease. *Proc Natl Acad Sci U S A* 97(18):9902–9906.

179. Adachi H, et al. (2007) CHIP overexpression reduces mutant androgen receptor protein and ameliorates phenotypes of the spinal and bulbar muscular atrophy transgenic mouse model. *J Neurosci* 27(19):5115–5126.

180. Lin H-K, et al. (2002) Proteasome activity is required for androgen receptor transcriptional activity via regulation of androgen receptor nuclear translocation and

interaction with coregulators in prostate cancer cells. *J Biol Chem* 277(39):36570–36576.

181. Katsuno M, et al. (2005) Pharmacological induction of heat-shock proteins alleviates polyglutamine-mediated motor neuron disease. *Proc Natl Acad Sci U S A* 102(46):16801–16806.

182. Waza M, et al. (2005) 17-AAG, an Hsp90 inhibitor, ameliorates polyglutamine-mediated motor neuron degeneration. *Nat Med* 11(10):1088–95.

183. Koga H, Kaushik S, Cuervo AM (2012) Protein Homeostasis and Aging: the importance of exquisite quality control. *Ageing Res Rev* 10(2):205–215.

184. Beitel LK, Alvarado C, Mokhtar S, Paliouras M, Trifiro M (2013) Mechanisms mediating spinal and bulbar muscular atrophy: investigations into polyglutamine-expanded androgen receptor function and dysfunction. *Front Neurol* 4(May):53.

185. Petrakis S, et al. (2012) Identification of human proteins that modify misfolding and proteotoxicity of pathogenic ataxin-1. *PLoS Genet* 8(8):e1002897.

186. Michelitsch MD, Weissman JS (2000) A census of glutamine/asparagine-rich regions: implications for their conserved function and the prediction of novel prions. *Proc Natl Acad Sci U S A* 97(22):11910–11915.

187. McCampbell A, et al. (2000) CREB-binding protein sequestration by expanded polyglutamine. *Hum Mol Genet* 9(14):2197–2202.

188. Lieberman AP, Harmison G, Strand AD, Olson JM, Fischbeck KH (2002) Altered transcriptional regulation in cells expressing the expanded polyglutamine androgen receptor. *Hum Mol Genet* 11(17):1967–1976.

189. Vismara G, et al. (2009) Androgens inhibit androgen receptor promoter activation in motor neurons. *Neurobiol Dis* 33(3):395–404.

190. Gottlieb B, Beitel LK, Nadarajah A, Paliouras M, Trifiro M (2012) The androgen receptor gene mutations database: 2012 update. *Hum Mutat* 33(5):887–894.

191. Montie HL, Pestell RG, Merry DE (2011) SIRT1 modulates aggregation and toxicity through deacetylation of the androgen receptor in cell models of SBMA. *J Neurosci* 31(48):17425–36.

192. Beauchemin A, et al. (2001) Cytochrome c oxidase subunit Vb interacts with human androgen receptor: a potential mechanism for neurotoxicity in spinobulbar muscular atrophy. *Brain Res Bull* 56(3-4):285–297.

193. Davies P, et al. (2008) Consequences of poly-glutamine repeat length for the conformation and folding of the androgen receptor amino-terminal domain. *J Mol Endocrinol* 41(5):301–14.

194. Pelassa I, et al. (2014) Association of polyalanine and polyglutamine coiled coils mediates expansion disease-related protein aggregation and dysfunction. *Hum Mol Genet* 23(13):3402–20.

195. Piccioni F, et al. (2002) Androgen receptor with elongated polyglutamine tract forms aggregates that alter axonal trafficking and mitochondrial distribution in motor neuronal. 1418–1420.

196. Szebenyi G, et al. (2003) Neuropathogenic forms of huntingtin and androgen receptor inhibit fast axonal transport. *Neuron* 40(1):41–52.

197. Katsuno M, et al. (2006) Reversible disruption of dynactin 1-mediated retrograde axonal transport in polyglutamine-induced motor neuron degeneration. *J Neurosci* 26(47):12106–17.

198. Yu Z, et al. (2006) Androgen-dependent pathology demonstrates myopathic contribution to the Kennedy disease phenotype in a mouse knock-in model. *J Clin Invest* 116(10):2663–2672.

199. Pronsato L, Boland R, Milanesi L (2013) Non-classical localization of androgen receptor in the C2C12 skeletal muscle cell line. *Arch Biochem Biophys* 530(1):13–22.

200. Ellerby LM, et al. (1999) Kennedy ' s Disease: Caspase Cleavage of the Androgen Receptor Is a Crucial Event in Cytotoxicity.

201. LaFevre-Bernt M a, Ellerby LM (2003) Kennedy's disease. Phosphorylation of the polyglutamine-expanded form of androgen receptor regulates its cleavage by caspase-3 and enhances cell death. *J Biol Chem* 278(37):34918–24.

202. Heine EM, Berger TR, Pluciennik A, Orr CR, Merry DE (2015) Proteasome-Mediated Proteolysis of the Polyglutamine-Expanded Androgen Receptor is a Late Event in SBMA Pathogenesis. *J Biol Chem*:jbc.M114.617894.

203. Mangelsdorf DJ, et al. (1995) The nuclear receptor superfamily: the second decade. *Cell* 83(6):835–839.

204. Mcewan IJ (2009) The Nuclear Receptor Superfamily. *Methods Mol Biol* 505:3–18.

205. Robinson-Rechavi M, Escriva Garcia H, Laudet V (2003) The nuclear receptor superfamily. *J Cell Sci* 116(Pt 4):585–586.

206. Wahli W, Martinez E (1991) Superfamily of steroid nuclear receptors: positive and negative regulators of gene expression. *FASEB J* 5(9):2243–2249.

207. Quigley CA, et al. (1995) Androgen receptor defects: Historical, clinical, and molecular perspectives. *Endocr Rev* 16(3):271–321.

208. Gelmann EP (2002) Molecular Biology of the Androgen Receptor. *J Clin Oncol* 20(13):3001–3015.

209. Van Royen ME, van Cappellen W a, de Vos C, Houtsmuller AB, Trapman J (2012) Stepwise androgen receptor dimerization. *J Cell Sci* 125(Pt 8):1970–9.

210. Clarke BL, Khosla S (2009) Androgens and bone. *Steroids* 74(3):296–305.

211. Fargo KN, Foecking EM, Jones KJ, Sengelaub DR (2009) Neuroprotective actions of androgens on motoneurons. *Front Neuroendocrinol* 30(2):130–141.

212. McEwan IJ, Lavery D, Fischer K, Watt K (2007) Natural disordered sequences in the amino terminal domain of nuclear receptors: lessons from the androgen and glucocorticoid receptors. *Nucl Recept Signal* 5:e001.

213. Tompa P (2012) Intrinsically disordered proteins: a 10-year recap. *Trends Biochem Sci* 37(12):509–16.

214. Dyson HJ, Wright PE (2005) Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* 6(3):197–208.

215. Tompa P, Varadi M (2014) Predicting the predictive power of IDP ensembles. *Structure* 22(2):177–8.

216. Wright PE, Dyson HJ (2009) Linking folding and binding. *Curr Opin Struct Biol* 19(1):31–38.

217. Tompa P, Fuxreiter M (2008) Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends Biochem Sci* 33(1):2–8.

218. Weikl TR, Paul F (2014) Conformational selection in protein binding and function. *Protein Sci* 23(11):1508–18.

219. Csermely P, Palotai R, Nussinov R (2010) Induced fit, conformational selection and independent dynamic segments: An extended view of binding events. *Trends Biochem Sci* 35(10):539–546.

220. Zhou H-X (2012) Intrinsic disorder: signaling via highly specific but short-lived association. *Trends Biochem Sci* 37(2):43–8.

221. Uversky VN (2002) Natively unfolded proteins: a point where biology waits for physics. *Protein Sci* 11(4):739–756.

222. Lacy ER, et al. (2004) p27 binds cyclin-CDK complexes through a sequential mechanism involving binding-induced protein folding. *Nat Struct Mol Biol* 11(4):358–364.

223. Ou L, et al. (2011) Incomplete folding upon binding mediates Cdk4/cyclin D complex activation by tyrosine phosphorylation of inhibitor p27 protein. *J Biol Chem* 286(34):30142–30151.

224. Lavery DN, McEwan IJ (2005) Structure and function of steroid receptor AF1 transactivation domains: induction of active conformations. *Biochem J* 391(Pt 3):449–464.

225. Lavery DN, McEwan IJ (2008) Functional characterization of the native NH2-terminal transactivation domain of the human androgen receptor: Binding kinetics for interactions with TFIIF and SRC-1a. *Biochemistry* 47(11):3352–3359.

226. Callewaert L, Van Tilborgh N, Claessens F (2006) Interplay between two hormone-independent activation domains in the androgen receptor. *Cancer Res* 66(1):543–553.

227. Lavery DN, McEwan IJ (2006) The human androgen receptor AF1 transactivation domain: interactions with transcription factor IIF and molten-globule-like structural characteristics. *Biochem Soc Trans* 34(Pt 6):1054–1057.

228. Kumar R, Betney R, Li J, Thompson EB, McEwan IJ (2004) Induced alpha-helix structure in AF1 of the androgen receptor upon binding transcription factor TFIIF. *Biochemistry* 43(11):3008–3013.

229. Luse DS (2012) Rethinking the role of TFIIF in transcript initiation by RNA polymerase II. *Transcription* 3(4):156–159.

230. Heery DM, Kalkhoven E, Hoare S, Parker MG (1997) A signature motif in transcriptional co-activators mediates binding to nuclear receptors. *Nature* 387(6634):733–736.

231. Dubbink HJ, et al. (2004) Distinct recognition modes of FXXLF and LXXLL motifs by the androgen receptor. *Mol Endocrinol* 18(9):2132–50.

232. Hur E, et al. (2004) Recognition and accommodation at the androgen receptor coactivator binding interface. *PLoS Biol* 2(9):E274.

233. Langley E, Kemppainen JA, Wilson EM (1998) Intermolecular NH2-/carboxyl-terminal interactions in androgen receptor dimerization revealed by mutations that cause androgen insensitivity. *J Biol Chem* 273(1):92–101.

234. He B, et al. (2004) Structural basis for androgen receptor interdomain and coactivator interactions suggests a transition in nuclear receptor activation function dominance. *Mol Cell* 16(3):425–438.

235. He B, Wilson EM (2003) Electrostatic modulation in steroid receptor recruitment of LXXLL and FXXLF motifs. *Mol Cell Biol* 23(6):2135–2150.

236. Orr CR, et al. (2010) An interdomain interaction of the androgen receptor is required for its aggregation and toxicity in spinal and bulbar muscular atrophy. *J Biol Chem* 285(46):35567–77.

237. Burd CJ, Petre CE, Moghadam H, Wilson EM, Knudsen KE (2005) Cyclin D1 binding to the androgen receptor (AR) NH2-terminal domain inhibits activation function 2 association and reveals dual roles for AR corepression. *Mol Endocrinol* 19(3):607–620.

238. Bai S, He B, Wilson EM (2005) Melanoma antigen gene protein MAGE-11 regulates androgen receptor function by modulating the interdomain interaction. *Mol Cell Biol* 25(4):1238–1257.

239. Chandra S, et al. (2008) A common motif targets huntingtin and the androgen receptor to the proteasome. *J Biol Chem* 283(35):23950–5.

240. Dehm SM, Regan KM, Schmidt LJ, Tindall DJ (2007) Selective role of an NH2-terminal WxxLF motif for aberrant androgen receptor activation in androgen depletion-independent prostate cancer cells. *Cancer Res* 67(20):10067–10077.

241. He B, Kemppainen JA, Wilson EM (2000) FXXLF and WXXLF sequences mediate the NH2-terminal interaction with the ligand binding domain of the androgen receptor. *J Biol Chem* 275(30):22986–22994.

242. He B, Lee LW, Minges JT, Wilson EM (2002) Dependence of selective gene activation on the androgen receptor NH2- and COOH-terminal interaction. *J Biol Chem* 277(28):25631–25639.

243. Wang Q, et al. (2012) Polymorphism of CAG repeats in androgen receptor of carnivores. *Mol Biol Rep* 39(3):2297–2303.

244. Shen HC, Coetzee GA (2005) The Androgen Receptor: Unlocking the Secrets of Its Unique Transactivation Domain. *Vitam Horm* 71:301–319.

245. Patiño-García B, et al. (2007) Association between polymorphisms of the androgen and vitamin D receptor genes with prostate cancer risk in a Mexican population. *Rev Invest Clin* 59(1):25–31.

246. Buchanan G, et al. (2004) Structural and functional consequences of glutamine tract variation in the androgen receptor. *Hum Mol Genet* 13(16):1677–92.

247. Cortes CJ, et al. (2014) Polyglutamine-expanded androgen receptor interferes with TFEB to elicit autophagy defects in SBMA. *Nat Neurosci* 17(August):5–8.

248. MacLean HE, Brown RW, Beilin J, Warne GL, Zajac JD (2004) Increased frequency of long androgen receptor CAG repeats in male breast cancers. *Breast Cancer Res Treat* 88(3):239–46.

249. Callewaert L, et al. (2003) Implications of a polyglutamine tract in the function of the human androgen receptor. *Biochem Biophys Res Commun* 306(1):46–52.

250. Irvine RA, Yu MC, Ross RK, Coetzee GA (1995) The CAG and GGC microsatellites of the androgen receptor gene are in linkage disequilibrium in men with prostate cancer. *Cancer Res* 55(9):1937–1940.

251. Claessens F, et al. (2008) Diverse roles of androgen receptor (AR) domains in AR-mediated signaling. *Nucl Recept Signal* 6:e008.

252. Simental JA, Sar M, Wilson EM (1992) Domain functions of the androgen receptor. *J Steroid Biochem Mol Biol* 43(1-3):37–41.

253. Shaffer PL, Jivan A, Dollins DE, Claessens F, Gewirth DT (2004) Structural basis of androgen receptor binding to selective androgen response elements. *Proc Natl Acad Sci U S A* 101(14):4758–63.

254. Klug A, Schwabe JW (1995) Protein motifs 5. Zinc fingers. *FASEB J* 9(8):597–604.

255. Umesono K, Evans RM (1989) Determinants of target gene specificity for steroid/thyroid hormone receptors. *Cell* 57(7):1139–1146.

256. Chen Y, Young MA (2010) Structure of a thyroid hormone receptor DNA-binding domain homodimer bound to an inverted palindrome DNA response element. *Mol Endocrinol* 24(8):1650–1664.

257. Zhou ZX, Lane M V, Kemppainen JA, French FS, Wilson EM (1995) Specificity of ligand-dependent androgen receptor stabilization: receptor domain interactions influence ligand dissociation and receptor stability. *Mol Endocrinol* 9(2):208–218.

258. Clinckemalie L, Vanderschueren D, Boonen S, Claessens F (2012) The hinge region in androgen receptor control. *Mol Cell Endocrinol* 358(1):1–8.

259. Matias PM, et al. (2000) Structural evidence for ligand specificity in the binding domain of the human androgen receptor. Implications for pathogenic gene mutations. *J Biol Chem* 275(34):26164–26171.

260. Sack JS, et al. (2001) Crystallographic structures of the ligand-binding domains of the androgen receptor and its T877A mutant complexed with the natural agonist dihydrotestosterone. *Proc Natl Acad Sci U S A* 98(9):4904–9.

261. Bourguet W, Germain P, Gronemeyer H (2000) Nuclear receptor ligand-binding domains: three-dimensional structures, molecular interactions and pharmacological implications. *Trends Pharmacol Sci* 21(10):381–388.

262. Van de Wijngaart DJ, Dubbink HJ, van Royen ME, Trapman J, Jenster G (2012) Androgen receptor coregulators: Recruitment via the coactivator binding groove. *Mol Cell Endocrinol* 352(1-2):57–69.

263. Moras D, Gronemeyer H (1998) The nuclear receptor ligand-binding domain: structure and function. *Curr Opin Cell Biol* 10(3):384–391.

264. Danielian PS, White R, Lees JA, Parker MG (1992) Identification of a conserved region required for hormone dependent transcriptional activation by steroid hormone receptors. *EMBO J* 11(3):1025–1033.

265. Katzen F (2007) recombinational cloning: a biological operating system. *Expert Opin Drug Discov* 2(4):571–589.

266. Nallamsetty S, Austin BP, Penrose KJ, Waugh DS (2005) Gateway vectors for the production of combinatorially-tagged His6-MBP fusion proteins in the cytoplasm and periplasm of Escherichia coli. *Protein Sci* 14(12):2964–2971.

267. Tataurov A V, You Y, Owczarzy R (2008) Predicting ultraviolet spectrum of single stranded and double stranded deoxyribonucleic acids. *Biophys Chem* 133(1-3):66–70.

268. Miller C a, Tan X, Wilson M, Bhattacharyya S, Ludwig S (2010) Single plasmids expressing human steroid hormone receptors and a reporter gene for use in yeast signaling assays. *Plasmid* 63(2):73–8.

269. Dickinson DJ, Ward JD, Reiner DJ, Goldstein B (2013) Engineering the Caenorhabditis elegans genome using Cas9-triggered homologous recombination. *Nat Methods* 10(10):1028–34.

270. Santner A a., et al. (2012) Sweeping away protein aggregation with entropic bristles: Intrinsically disordered protein fusions enhance soluble expression. *Biochemistry* 51(37):7250–7262.

271. Chow MKM, Ellisdon AM, Cabrita LD, Bottomley SP (2006) Purification of polyglutamine proteins. *Methods Enzymol* 413(06):1–19.

272. Marley J, Lu M, Bracken C (2001) A method for efficient isotopic labeling of recombinant proteins. *J Biomol NMR* 20(1):71–5.

273. Xue B, Dunbrack RL, Williams RW, Dunker AK, Uversky VN (2010) PONDR-FIT: A meta-predictor of intrinsically disordered amino acids. *Biochim Biophys Acta - Proteins Proteomics* 1804(4):996–1010.

274. Munoz V, Serrano L (1994) Elucidating the Folding Problem of Helical Peptides Using Empirical Parameters. *Nat Struct Biol* 1(6):399–409.

275. Doig AJ (2008) Stability and design of alpha-helical peptides. *Prog Mol Biol Transl Sci* 83:1–52.

276. McDonnell a V, Jiang T, Keating a E, Berger B (2006) Paircoil2: improved prediction of coiled coils from sequence. *Bioinformatics* 22(3):356–8.

277. Berger B, et al. (1995) Predicting coiled coils by use of pairwise residue correlations. *Proc Natl Acad Sci U S A* 92(18):8259–63.

278. Jayaraman M, Thakur AK, Kar K, Kodali R, Wetzel R (2011) Assays for studying nucleated aggregation of polyglutamine proteins. *Methods* 53(3):246–54.

279. O'Nuallain B, et al. (2006) Kinetics and thermodynamics of amyloid assembly using a high-performance liquid chromatography-based sedimentation assay. *Methods Enzymol* 413(06):34–74.

280. Rahimi F, Bitan G (2010) Selection of aptamers for amyloid beta-protein, the causative agent of Alzheimer's disease. *J Vis Exp* (39):1–7.

281. Hoshino M, Katou H, Yamaguchi K, Goto Y (2007) Dimethylsulfoxide-quenched hydrogen/deuterium exchange method to study amyloid fibril structure. *Biochim Biophys Acta* 1768(8):1886–99.

282. Koo BONW, Miranker AD (2005) Contribution of the intrinsic disulfide to the assembly mechanism of islet amyloid. 231–239.

283. Monsellier E, Redeker V, Ruiz-Arlandis G, Bousset L, Melki R (2015) Molecular Interaction between the Chaperone Hsc70 and the N-terminal Flank of Huntingtin Exon 1 Modulates Aggregation. *J Biol Chem* 290(5):2560–2576.

284. Nichols MR, Moss MA, Reed DK, Hoh JH, Rosenberry TL (2005) Amyloid-β aggregates formed at polar-nonpolar interfaces differ from amyloid-β protofibrils produced in aqueous buffers. *Microsc Res Tech* 67(3-4):164–174.

285. Walgers R, Lee TC, Cammers-Goodwin A (1998) An indirect chaotropic mechanism for the stabilization of helix conformation of peptides in aqueous trifluoroethanol and hexafluoro-2- propanol. *J Am Chem Soc* 120(20):5073–5079.

286. Raja M (2010) Monitoring the effects of strong cosolvent hexafluoroisopropanol in investigation of the tetrameric structure and stability of K+-channel KcsA. *Arch Biochem Biophys* 498(1):1–6.

287. Wang Y-Q, et al. (2011) Relationship between prion propensity and the rates of individual molecular steps of fibril assembly. *J Biol Chem* 286(14):12101–7.

288. Hudson S a., Ecroyd H, Kee TW, Carver J a. (2009) The thioflavin T fluorescence assay for amyloid fibril detection can be biased by the presence of exogenous compounds. *FEBS J* 276(20):5960–5972.

289. Groenning M (2010) Binding mode of Thioflavin T and other molecular probes in the context of amyloid fibrils-current status. *J Chem Biol* 3(1):1–18.

290. Goodman JW (1976) Some fundamental properties of speckle. *J Opt Soc Am* 66(11):1145.

291. Pecora R (2000) Dynamic light scattering measurement of nanometer particles in liquids. *J Nanoparticle Res* 2:123–131.

292. Uversky VN (2012) *Intrinsically Disordered Protein Analysis* eds Uversky VN, Dunker AK (Humana Press, Totowa, NJ) doi:10.1007/978-1-61779-927-3.

293. Kelly SM, Jess TJ, Price NC (2005) How to study proteins by circular dichroism. *Biochim Biophys Acta* 1751(2):119–39.

294. Whitmore L, Wallace BA (2004) DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res* 32(WEB SERVER ISS.). doi:10.1093/nar/gkh371.

295. Whitmore L, Wallace BA (2008) Protein secondary structure analyses from circular dichroism spectroscopy: Methods and reference databases. *Biopolymers* 89(5):392–400.

296. Schuck P (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys J* 78:1606–1619.

297. Brown PH, Schuck P (2008) A new adaptive grid-size algorithm for the simulation of sedimentation velocity profiles in analytical ultracentrifugation. *Comput Phys Commun* 178:105–120.

298. Nölting B (2010) *Methods in modern biophysics: Third edition* doi:10.1007/978-3-642-03022-2.

299. Bruker (2011) Introduction to Brukers ScanAsyst and PeakForce Tapping Atomic Force Microscopy Technology AFM AN133. 1–12.

300. Fultz B, Howe JM (2008) *Transmission electron microscopy and diffractometry of materials* doi:10.1007/978-3-540-73886-2.

301. Kiselev NA, Sherman MB, Tsuprun VL (1990) Negative staining of proteins. *Electron Microsc Rev* 3(1):43–72.

302. Fontana A, et al. (2004) Probing protein structure by limited proteolysis. *Acta Biochimica Polonica*, pp 299–321.

303. Gao X, et al. (2005) High-throughput limited proteolysis/mass spectrometry for protein domain elucidation. *J Struct Funct Genomics* 6(2-3):129–34.

304. Williams AD, et al. (2005) Structural properties of Abeta protofibrils stabilized by a small molecule. *Proc Natl Acad Sci U S A* 102(20):7115–20.

305. Frare E, et al. (2006) Identification of the core structure of lysozyme amyloid fibrils by proteolysis. *J Mol Biol* 361(3):551–61.

306. Kheterpal I, Williams A, Murphy C, Bledsoe B, Wetzel R (2001) Structural features of the Abeta amyloid fibril elucidated by limited proteolysis. *Biochemistry* 40(39):11757–67.

307. Wang Z-M, Lashuel H a (2013) Discovery of a novel aggregation domain in the huntingtin protein: implications for the mechanisms of Htt aggregation and toxicity. *Angew Chem Int Ed Engl* 52(2):562–7.

308. Kobayashi Y, et al. (1998) Caspase-3 cleaves the expanded androgen receptor protein of spinal and bulbar muscular atrophy in a polyglutamine repeat length-dependent manner. *Biochem Biophys Res Commun* 252(1):145–50.

309. Dyson HJ, Wright PE, Scheraga HA (2006) The role of hydrophobic interactions in initiation and propagation of protein folding. *Proc Natl Acad Sci U S A* 103(35):13057–61.

310. Mamedov TG, et al. (2008) A fundamental study of the PCR amplification of GC-rich DNA templates. *Comput Biol Chem* 32(6):452–7.

311. Takahashi N, Sasagawa N, Suzuki K, Ishiura S (1999) Synthesis of long trinucleotide repeats in vitro. *Neurosci Lett* 262(1):45–48.

312. Tobelmann MD, Kerby RL, Murphy RM (2008) A strategy for generating polyglutamine “length libraries” in model host proteins. *Protein Eng Des Sel* 21(3):161–4.

313. Rocha EPC (2004) Codon usage bias from tRNA’s point of view: Redundancy, specialization, and efficient decoding for translation optimization. *Genome Res* 14(11):2279–2286.

314. Menzella HG (2011) Comparison of two codon optimization strategies to enhance recombinant protein production in Escherichia coli. *Microb Cell Fact* 10:15.

315. Husain-Ponnampalam R, Turnbull V, Tarlac V, Storey E (2010) Expression and purification of ataxin-1 protein. *J Neurosci Methods* 189(1):30–5.

316. Nallamsetty S, Waugh DS (2006) Solubility-enhancing proteins MBP and NusA play a passive role in the folding of their fusion partners. *Protein Expr Purif* 45(1):175–182.

317. Kapust RB, Waugh DS (1999) Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci* 8(8):1668–1674.

318. Sun P, Tropea JE, Waugh DS (2011) Enhancing the solubility of recombinant proteins in escherichia coli by using hexahistidine-tagged maltose-binding protein as a fusion partner. *Methods Mol Biol* 705:259–274.

319. Stark GR, Stein WH, Moore S (1960) Reactions of the Cyanate Present in Aqueous Urea with Amino Acids and Proteins. *J Biol Chem* 235(11):3177–3181.

320. Piai A, et al. (2014) “CON-CON” assignment strategy for highly flexible intrinsically disordered proteins. *J Biomol NMR* 60(4):209–18.

321. Tamiola K, Acar B, Mulder FAA (2010) Sequence-specific random coil chemical shifts of intrinsically disordered proteins. *J Am Chem Soc* 132(51):18000–18003.

322. Logan TM, Thériault Y, Fesik SW (1994) Structural characterization of the FK506 binding protein unfolded in urea and guanidine hydrochloride. *J Mol Biol* 236(2):637–648.

323. Lee CC, Walters RH, Murphy RM (2007) Reconsidering the mechanism of polyglutamine peptide aggregation. *Biochemistry* 46(44):12810–12820.

324. Kokona B, Rosenthal ZP, Fairman R (2014) Role of the Coiled-Coil Structural Motif in Polyglutamine Aggregation. *Biochemistry*.

325. Kokona B, Johnson KA, Fairman R (2014) Effect of Helical Flanking Sequences on the Morphology of Polyglutamine-Containing Fibrils. *Biochemistry*.

326. Petrakis S, Schaefer MH, Wanker EE, Andrade-Navarro M a. (2013) Aggregation of polyQ-extended proteins is promoted by interaction with their natural coiled-coil partners. *BioEssays* 35(6):503–507.

327. Streets AM, Sourigues Y, Kopito RR, Melki R, Quake SR (2013) Simultaneous measurement of amyloid fibril formation by dynamic light scattering and fluorescence reveals complex aggregation kinetics. *PLoS One* 8(1):e54541.

328. Chen S, Wetzel R (2001) Solubilization and disaggregation of polyglutamine peptides. 887–891.

329. Kar K, Arduini I, Drombosky KW, van der Wel PC a, Wetzel R (2014) D-polyglutamine amyloid recruits L-polyglutamine monomers and kills cells. *J Mol Biol* 426(4):816–29.

330. Evans KC, Berger EP, Cho CG, Weisgraber KH, Lansbury PT (1995) Apolipoprotein E is a kinetic but not a thermodynamic inhibitor of amyloid formation: implications for the pathogenesis and treatment of Alzheimer disease. *Proc Natl Acad Sci U S A* 92(3):763–767.

331. Wood SJ, Maleeff B, Hart T, Wetzel R (1996) Physical, morphological and functional differences between ph 5.8 and 7.4 aggregates of the Alzheimer's amyloid peptide Abeta. *J Mol Biol* 256(5):870–877.

332. Zhang S, et al. (2000) The Alzheimer's peptide a beta adopts a collapsed coil structure in water. *J Struct Biol* 130(2-3):130–141.

333. Mason PE, Neilson GW, Dempsey CE, Barnes AC, Cruickshank JM (2003) The hydration structure of guanidinium and thiocyanate ions: implications for protein stability in aqueous solution. *Proc Natl Acad Sci U S A* 100(8):4557–4561.

334. Natalello A, et al. (2011) A major role for side-chain polyglutamine hydrogen bonding in irreversible ataxin-3 aggregation. *PLoS One* 6(4). doi:10.1371/journal.pone.0018789.

335. Chow MKM, Ellisdon AM, Cabrita LD, Bottomley SP (2004) Polyglutamine expansion in ataxin-3 does not affect protein stability: Implications for misfolding and disease. *J Biol Chem* 279(46):47643–47651.

336. Yonezawa Y, et al. (2002) An insight into the pathway of the amyloid fibril formation of hen egg white lysozyme obtained from a small-angle X-ray and neutron scattering study. *J Mol Biol* 323(2):237–251.

337. Maroufi B, Ranjbar B, Khajeh K, Naderi-Manesh H, Yaghoubi H (2008) Structural studies of hen egg-white lysozyme dimer: Comparison with monomer. *Biochim Biophys Acta - Proteins Proteomics* 1784(7-8):1043–1049.

338. Zagorski MG, et al. (1999) *Amyloid, Prions, and Other Protein Aggregates* (Elsevier) doi:10.1016/S0076-6879(99)09015-1.

339. Baldwin RL (1996) How Hofmeister ion interactions affect protein stability. *Biophys J* 71(4):2056–2063.

340. Knight SD, Presto J, Linse S, Johansson J (2013) The BRICHOS domain, amyloid fibril formation, and their relationship. *Biochemistry* 52(43):7523–31.

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

342. Navarra G, Troia F, Militello V, Leone M (2013) Characterization of the nucleation process of lysozyme at physiological pH: primary but not sole process. *Biophys Chem* 177-178:24–33.

343. Li Y, Lubchenko V, Vekilov PG (2011) The use of dynamic light scattering and brownian microscopy to characterize protein aggregation. *Rev Sci Instrum* 82(5):053106.

344. Chen S, Berthelier V, Yang W, Wetzel R (2001) Polyglutamine aggregation behavior in vitro supports a recruitment mechanism of cytotoxicity. *J Mol Biol* 311(1):173–82.

345. Xue WF, Homans SW, Radford SE (2009) Amyloid fibril length distribution quantified by atomic force microscopy single-particle image analysis. *Protein Eng Des Sel* 22(8):489–496.

346. Legleiter J, et al. (2010) Mutant huntingtin fragments form oligomers in a polyglutamine length-dependent manner in vitro and in vivo. *J Biol Chem* 285(19):14777–90.

347. Barth A (2007) Infrared spectroscopy of proteins. *Biochim Biophys Acta - Bioenerg* 1767(9):1073–1101.

348. Vázquez-Fernández E, et al. (2012) Structural organization of mammalian prions as probed by limited proteolysis. *PLoS One* 7(11):e50111.

349. Jayaraman M, et al. (2012) Kinetically competing huntingtin aggregation pathways control amyloid polymorphism and properties. *Biochemistry* 51(13):2706–16.

350. Fawzi NL, Ying J, Torchia DA, Clore GM (2010) Kinetics of amyloid?? monomer-to-oligomer exchange by NMR relaxation. *J Am Chem Soc* 132(29):9948–9951.

351. Fersht AR (2000) Transition-state structure as a unifying basis in protein-folding mechanisms: contact order, chain topology, stability, and the extended nucleus mechanism. *Proc Natl Acad Sci U S A* 97(4):1525–1529.

352. Wood E (1987) Data for Biochemical Research (third edition). *Biochem Educ* 15(2):97.

353. Wei L, et al. (2009) Residual structure in islet amyloid polypeptide mediates its interactions with soluble insulin. *Biochemistry* 48(11):2368–2376.

354. Wiltzius JJW, Sievers S a., Sawaya MR, Eisenberg D (2009) Atomic structures of IAPP (amylin) fusions suggest a mechanism for fibrillation and the role of insulin in the process. *Protein Sci* 18(7):1521–1530.

355. Fändrich M (2012) Oligomeric intermediates in amyloid formation: Structure determination and mechanisms of toxicity. *J Mol Biol* 421(4-5):427–440.

356. Ping Cao, Peter Marek, Harris Noor, Vadim Patsalo, Ling-Hsien Tu HW, Andisheh Abedini and DPR (2013) Islet amyloid: From fundamental biophysics to mechanisms of cytotoxicity. *FEBS Lett* 29(6):997–1003.

357. Bhattacharyya A, et al. (2006) Oligoproline effects on polyglutamine conformation and aggregation. *J Mol Biol* 355(3):524–535.

358. Doig a J, Baldwin RL (1995) N- and C-capping preferences for all 20 amino acids in alpha-helical peptides. *Protein Sci* 4(7):1325–1336.

359. Lakhani V V, Ding F, Dokholyan N V (2010) Polyglutamine induced misfolding of huntingtin exon1 is modulated by the flanking sequences. *PLoS Comput Biol* 6(4):e1000772.

360. Tartari M, et al. (2008) Phylogenetic comparison of huntingtin homologues reveals the appearance of a primitive polyQ in sea urchin. *Mol Biol Evol* 25(2):330–8.

361. Huntley MA, Clark AG (2007) Evolutionary analysis of amino acid repeats across the genomes of 12 drosophila species. *Mol Biol Evol* 24(12):2598–2609.

362. Huntley M, Golding GB (2000) Evolution of simple sequence in proteins. *J Mol Evol* 51(2):131–140.

363. Lenz C, Haerty W, Golding GB (2014) Increased substitution rates surrounding low-complexity regions within primate proteins. *Genome Biol Evol* 6(3):655–665.

364. Project G, et al. (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature* 135(V):0–9.

365. Ding F, Borreguero JM, Buldyrey S V, Stanley HE, Dokholyan N V (2003) Mechanism for the alpha-helix to beta-hairpin transition. *Proteins* 53(2):220–228.

366. Suzuki Y, Brender JR, Hartman K, Ramamoorthy A, Marsh ENG (2012) Alternative pathways of human islet amyloid polypeptide aggregation distinguished by (19)f nuclear magnetic resonance-detected kinetics of monomer consumption. *Biochemistry* 51(41):8154–62.

367. Kunjithapatham R, et al. (2005) Role for the alpha-helix in aberrant protein aggregation. *Biochemistry* 44(1):149–156.

368. Liu G, et al. (2010) Mechanistic Studies of Peptide Self-Assembly: Transient alpha-Helices to Stable beta-Sheets. *J Am Chem Soc* 132(11):18223–18232.

369. Rolfsmeier ML, Lahue RS (2000) Stabilizing effects of interruptions on trinucleotide repeat expansions in Saccharomyces cerevisiae. *Mol Cell Biol* 20(1):173–180.

370. Sobczak K, Krzyzosiak WJ (2005) CAG repeats containing CAA interruptions form branched hairpin structures in spinocerebellar ataxia type 2 transcripts. *J Biol Chem* 280(5):3898–3910.

371. Kurosaki T, Ninokata A, Wang L, Ueda S (2006) Evolutionary scenario for acquisition of CAG repeats in human SCA1 gene. *Gene* 373(1-2):23–27.

372. Brahmachari SK (2003) Role of histidine interruption in mitigating the pathological effects of long polyglutamine stretches in SCA1□: A molecular approach. 953–962.

373. Kumar R, et al. (2011) Role of the androgen receptor CAG repeat polymorphism in prostate cancer, and spinal and bulbar muscular atrophy. *Life Sci* 88(13-14):565–71.

374. Schaufele F, et al. (2005) The structural basis of androgen receptor activation: intramolecular and intermolecular amino-carboxy interactions. *Proc Natl Acad Sci U S A* 102(28):9802–9807.

375. Burke K a, Godbey J, Legleiter J (2011) Assessing mutant huntingtin fragment and polyglutamine aggregation by atomic force microscopy. *Methods* 53(3):275–84.

376. Young JE, et al. (2009) Polyglutamine-expanded androgen receptor truncation fragments activate a Bax-dependent apoptotic cascade mediated by DP5/Hrk. *J Neurosci* 29(7):1987–1997.

377. Kamada S, Kikkawa U, Tsujimoto Y, Hunter T (2005) Nuclear translocation of caspase-3 is dependent on its proteolytic activation and recognition of a substrate-like protein(s). *J Biol Chem* 280(2):857–860.

378. Luo M, et al. (2010) Nuclear entry of active caspase-3 is facilitated by its p3-recognition-based specific cleavage activity. *Cell Res* 20(2):211–222.

379. Mishra R, et al. (2012) Inhibiting the nucleation of amyloid structure in a huntingtin fragment by targeting α-helix-rich oligomeric intermediates. *J Mol Biol* 415(5):900–17.

380. Pongtepaditep S, Limjindaporn T, Lertrit P, Srisawat C, Limwongse C (2012) Polyglutamined expanded androgen receptor interacts with chaperonin CCT. *Eur J Med Genet* 55(11):599–604.

381. Yang Z, et al. (2007) ASC-J9 ameliorates spinal and bulbar muscular atrophy phenotype via degradation of androgen receptor. *Nat Med* 13(3):348–53.

382. Månsson C, et al. (2014) DNAJB6 is a peptide-binding chaperone which can suppress amyloid fibrillation of polyglutamine peptides at substoichiometric molar ratios. *Cell Stress Chaperones* 19(2):227–39.

383. Cao P, Abedini A, Raleigh DP (2013) Aggregation of islet amyloid polypeptide: From physical chemistry to cell biology. *Curr Opin Struct Biol* 23(1):82–89.

384. Baldwin AJ, et al. (2006) Cytochrome Display on Amyloid Fibrils Cytochrome Display on Amyloid Fibrils. (Ii):2162–2163.

385. Zhong C, et al. (2014) Strong underwater adhesives made by self-assembling multi-protein nanofibres. *Nat Nanotechnol* 9(10):858–866.

386. Kay BK (2012) SH3 domains come of age. *FEBS Lett* 586(17):2606–8.

387. Thompson KE, Bashor CJ, Lim W a., Keating AE (2012) Synzip protein interaction toolbox: In vitro and in vivo specifications of heterospecific coiled-coil interaction domains. *ACS Synth Biol* 1(4):118–129.

388. Thomson a. R, et al. (2014) Computational design of water-soluble -helical barrels. *Science (80-)* 346(6208):485–488.

389. Sharp TH, et al. (2012) Cryo-transmission electron microscopy structure of a gigadalton peptide fiber of de novo design. *Proc Natl Acad Sci*. doi:10.1073/pnas.1118622109.

390. Srinivasan N, Bhagawati M, Ananthanarayanan B, Kumar S (2014) Stimuli-sensitive intrinsically disordered protein brushes. *Nat Commun* 5:5145.

391. Lonergan PE, Tindall DJ (2011) Androgen receptor signaling in prostate cancer development and progression. *J Carcinog* 10:20.

392. Greenfield NJ (2006) Using circular dichroism spectra to estimate protein secondary structure. *Nat Protoc* 1(6):2876–2890.