Genetic factors and analysis of protein misfolding in vivo
Holmberg, Mats

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date: 2015

Download date: 28-11-2021
Chapter 1: General Introduction

1.1 Protein aggregates, occurrences and characteristics

All living cells are in a constant thermodynamic struggle to maintain the high degree of order necessary for complex life. In the case of proteins this goes far beyond the correct linking of amino acids. Proteins are constantly recycled by degradation to respond to different environmental conditions, as well as to eliminate damaged proteins [1, 2]. The dependence of protein function on its three dimensional structure necessitates the maintenance of correct protein folding. Not only will misfolded proteins be non-functional, but they can also be highly toxic. In cells where the load of misfolded proteins exceeds the capacity of the cell to refold or degrade, non-native proteins often aggregate in dense inclusions [3]. These aggregates have been found to be a unifying property of a set of distinct neurodegenerative disorders which had previously been thought to lack a common cause [4]. We now know that their different symptoms arise from the selective loss of function of disease-specific neurons [5, 6, 7]. Naturally, the first assumption was that the aggregates were the cause of the illness [8] and this view has only recently been challenged by a more complex image that portrays aggregates as both toxic and benign [9, 10].

Generally, protein aggregates refer to any massed continuous body of peptides regardless of shape, composition or biochemical properties. Sizes vary from oligomers, which can readily be analysed using chromatographic methods, to inclusions up to several micrometers in diameter [3, 11]. These inclusions have been found to consist of many different misfolded proteins as well as components of the protein folding and degradation machinery, consistent with a model where the protein homeostasis machinery is concentrated for higher efficiency. Another possibility is that components of the cellular protein refolding machinery accidentally co-aggregate while attempting to rescue aggregated proteins. A possible benefit of protein aggregation for cellular fitness is the reduced surface exposure of toxic moieties [12].

1.2 Amyloid aggregation

One of the clearest distinctions separating one type of aggregates from another is the presence or absence of amyloid material. Amyloids are macromolecular structures consisting of ordered, repetitively stacked β-sheet structures, usually formed by one type of polypeptide [4]. These form fibres perpendicular to the β-sheet and can vary in length from just a few
molecules to micrometer long structures that in turn can associate into bundles or tangles structurally similar to the organisations of individual strands in a rope or cable (Figure 1). In general, amyloid is seen as a final and extremely stable protein state which is no longer in equilibrium with its surrounding [4]. Soluble proteins can attach and increase the length of fibres, which might sheer into smaller segments as has been shown for the *Saccharomyces cerevisiae* amyloidogenic protein Sup35 [13]. However, in most organisms only a minority of amyloid proteins are resolubilised [14]. The physical properties of amyloids, such as the conformation of constituent proteins, make their investigation difficult. Conventional methods to solve protein structures, such as nuclear magnetic resonance spectroscopy (NMR) and x-ray crystallography, are hindered by the size of the particles and their semi-ordered state, which has so far only allowed relatively low resolution structures to be generated [15].

**Figure 1: Fibril structure and assembly**

Upper panel: Adapted from Chiti and Dobson, 2006 [4]. An Aβ proto-fiber seen along its growth axis.

Lower panel: A schematic representation of an amyloid fiber (not to scale) showing the growth of the fiber through monomer addition and templating.
Although amyloids and research into amyloid formation are associated with a relatively small set of proteins, *in vitro* studies have shown that, given time and the right circumstances, any protein can form amyloids [16]. Mutations in critical amino acids, temperature and destabilising factors, such as loss of ligand-binding or metal coordination, can increase a protein’s propensity to form amyloids, as can low pH and high concentrations of the protein itself [16]. Amyloids occur in nature not only as failed, terminally sequestered proteins, but are in some cases formed by organisms to perform a specific function. In some prokaryotes, the stickiness of aggregates is used for adherence to smooth surfaces, the formation of biofilm or membrane permeabilisation [17]. In spiders and other arthropods, the strength of amyloid fibres is used in the formation of silk. The strength/density ratio of these amyloidogenic silk threads is several times that of steel [18]. Finally, pigmentation in humans and other animals is created by the formation of amyloid like light absorbing particles of the protein Melanin [19].

### 1.2.1 Protein aggregation diseases

Amyloid- or amyloid-like deposits are pathological hallmarks of several human diseases (Table 1, adapted from Chiti and Dobson, 2006 [4]). It is not known how and in which assembly-state the amyloid-forming proteins are involved in the diseases. Described below are four examples illustrating a role of protein aggregation in neurodegenerative diseases (NDD).
### Table 1: Human diseases associated with the formation of extracellular amyloid deposits or intracellular inclusions with amyloid-like characteristics

<table>
<thead>
<tr>
<th>Disease</th>
<th>Aggregating protein or peptide</th>
<th>Native structure</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Neurodegenerative diseases</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alzheimer’s disease</td>
<td>Amyloid β peptide; 40 or 42 residues</td>
<td>Natively unfolded</td>
<td>20, 21</td>
</tr>
<tr>
<td>Parkinson’s disease</td>
<td>α-Synuclein; 140 residues</td>
<td>Natively unfolded</td>
<td>22</td>
</tr>
<tr>
<td>Dementia with Lewy bodies</td>
<td>α-Synuclein; 140 residues</td>
<td>Natively unfolded</td>
<td>23</td>
</tr>
<tr>
<td>Frontotemporal dementia with Parkinsonism</td>
<td>Tau; 352-441 residues</td>
<td>Natively unfolded</td>
<td>24</td>
</tr>
<tr>
<td>Amyotrophic lateral sclerosis</td>
<td>Superoxide dismutase 1; 153 residues</td>
<td>All-β, Ig like b)</td>
<td>25</td>
</tr>
<tr>
<td>Huntington’s disease</td>
<td>Huntingtin with polyQ expansion; 3144 residues</td>
<td>Largely natively unfolded</td>
<td>26</td>
</tr>
<tr>
<td>Spinocerebellar ataxias</td>
<td>Ataxins with polyQ expansion; 816 residues</td>
<td>All-β⁶, AXH domain (residues 562-694)</td>
<td>27</td>
</tr>
<tr>
<td>Spinal and bulbar muscular atrophy</td>
<td>Androgen receptor with polyQ expansion; 919 residues</td>
<td>All-β⁶ nuclear receptor ligand-binding domain (residues 669-919)</td>
<td>28</td>
</tr>
<tr>
<td><strong>Non-neurodegenerative diseases</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amyloid light-chain (AL) amyloidosis</td>
<td>Immunoglobulin light chains of fragments; ~90 residues⁵</td>
<td>All-β⁶, Ig like b)</td>
<td>29</td>
</tr>
<tr>
<td>Amyloid A (AA) amyloidosis</td>
<td>Fragments of serum amyloid A protein; 76-104 residues²</td>
<td>All-α², unknown fold</td>
<td>30</td>
</tr>
<tr>
<td>Type II diabetes</td>
<td>Amylin, also called islet amyloid polypeptide; 37 residues</td>
<td>Natively unfolded</td>
<td>31</td>
</tr>
<tr>
<td>Cataract</td>
<td>γ-Crystallins</td>
<td>All-β⁶, γ-crystallin like</td>
<td>32</td>
</tr>
<tr>
<td>Inclusion-body myosis</td>
<td>Amyloid β peptide; variable, 40 or 42 residues³</td>
<td>Natively unfolded</td>
<td>33</td>
</tr>
</tbody>
</table>

¹¹ Predominantly sporadic, although in some cases hereditary forms associated with specific mutations are well documented.

²² Predominantly hereditary, although sporadic forms are documented in some cases.

³³ Fragments of various lengths are generated and have been reported to be present in ex vivo fibrils.

⁴⁴ Lengths shown refer to the normal sequences with nonpathogenic traits of polyQ.

⁵⁵ Length shown is for ataxin-1.

⁶⁶ Domains consisting of β-sheets according to the structural classification of proteins (SCOP).

⁷⁷ Domains consisting of α-helices according to the structural classification of proteins (SCOP).
1.2.1.1 Alzheimer’s disease

Alzheimer’s disease (AD) can be both familial and non-familial [34]. Non-familial AD most often has a late onset with diagnosis typically not before the age of 60, while familial forms of the disease generally appear much earlier. However, both forms share their symptoms, commonly including cognitive impairment such as thinking, memory and reasoning. AD is characterised by intracellular amyloid ‘tangles’ formed by the microtubule stabilising protein Tau as well as extracellular plaques of Amyloid-β (Aβ), which is the N-terminal fragment of the synaptic amyloid precursor protein (APP) [35]. APP cleavage most commonly leaves a 40 or 42 amino acid fragment, of which Aβ\textsubscript{42} is less soluble and more prone to toxicity and aggregation. Mutations in three genes in addition to mutations in APP strongly confer AD susceptibility [37] and all three have been connected to APP-processing and to the ratio of Aβ\textsubscript{40} to Aβ\textsubscript{42}. These are Presenilin 1 and 2 (PSEN1/2), which make up parts of the γ-secretase complex, a transmembrane protease responsible for the cleavage of Aβ and apolipoprotein E (APOE) which, despite not directly affecting the formation of Aβ isoforms, is known to bind Aβ [38-41]. Different alleles coding for isoforms of APOE confer an increased susceptibility to AD and have been associated with an increase in the load of amyloid Aβ [42]. Although the mechanism of APOE and Aβ interaction is not known in detail, it is hypothesised that especially the lipidated form of APOE reduces clearance and/or induces Aβ aggregation. In addition to the above mentioned modifiers, genome wide associations have identified several genes and loci carrying susceptibility to AD, including ACE (encoding the zink metalloprotease Angiotensin I converting enzyme) and CH25H encoding Cholesterol 25-hydrolase which act upstream of lipid metabolism gene regulation. [43, 44].

1.2.1.2 Parkinson’s disease

Parkinson’s disease (PD) is the second largest NDD in number of affected individuals and was identified by Dr. James Parkinson in 1817 with symptoms including tremor, slow movement and rigidity. Later, cognitive impairments including dementia, personality changes and emotional disturbances have been added as typical symptoms [45]. In PD, aggregation of the protein alpha-Synuclein (aSyn) into intracellular inclusions, called Lewy bodies, is the main histological symptom, while some cases can also display Tau aggregation similar to what is observed in AD [46, 47]. Another similarity with AD is the presence of both familial and sporadic cases where the latter makes up the majority of cases. So far, several gene variations have been linked to PD, which include mutations and multiplications of the aSyn gene [48,
These lead to increased aggregation by increasing the amount of available aSyn by higher expression levels, lower degradation rates, or by increasing aSyn aggregation propensity. Symptoms arise due to neuronal death with motor symptoms resulting from a 60-80% loss of dopamine producing cells in a brain region called the Substantia Nigra, identified by their dark coloring of Neuromelanin [50]. Administration of L-DOPA, a Dopamine precursor, can alleviate PD symptoms temporarily until resistance to the therapy occurs [51].

1.2.1.3 Huntington’s Disease
Huntington’s Disease (HD) is the most prevalent among a group of disorders where the aggregating protein, specific for each disease, contains an expanded polyglutamine (polyQ) region that is highly prone to aggregation [26]. The CAG-repeats encoding for the polyQ region can cause DNA polymerase slippage during replication resulting in an expanded polyQ stretch and increased aggregation propensity [52]. In HD, the polyQ containing protein is Huntingtin (HTT) with the expansion located in the N-terminal of the protein. PolyQ disorders are invariantly monogenic with the polyQ expansions being both necessary and sufficient to acquire the disease [53]. Both, aggregation propensity and age of onset are on average inversely proportional to the length of the polyQ stretch [54, 42]. Despite this, individuals with similar lengths of polyQ show variation in age of onset indicating the influence of other genes and environmental factors [53, 56].

1.3 Non-amyloid aggregation
Not all aggregation prone proteins form amyloid fibrils in a cell, but they can form other type of protein aggregates [57, 58]. The dynamic nature of most proteins results in a constant fight for balance between cellular folding and degradation pathways on one side and entropy and de novo synthesis on the other [59]. Non-amyloid aggregation does not depend on sequence homology between different aggregating polypeptides, does not form semi-organised structures, such as the cross-β, and is not irreversible.

1.3.1 Amyotrophic lateral sclerosis
Amyotrophic lateral sclerosis (ALS) selectively affects the upper and lower motor neurons, leaving neighboring sensory neurons intact [60]. Symptoms progress from muscle weakness
and twitching, to muscular atrophy and paralysis and generally start in the extremities and migrate inwards. The disease is invariably fatal with full body paralysis usually resulting in death by respiratory failure. More than 90% of cases are sporadic with no clear genetic influence [61]. Of the familial cases, 20% can be attributed to mutations in Superoxide dismutase 1 (SOD1), a reactive oxygen scavenging cytoplasmic protein [25]. Ubiquinated SOD1, together with several other ubiquinated proteins, can be found in inclusion bodies in affected tissues. Disease causing mutations have also been identified in other proteins including TAR DNA binding protein 43 (TDP43) and Fused in Sacroma (FUS) [62, 63]. The former is involved in splicing and the latter in transcription. Contrary to the aggregates found in AD, PD and HD, none of these proteins seems to form amyloid in vivo as visualised by the amyloid specific dyes ThioflavinT and Congo Red [57, 64].

1.4 Visualisation of aggregated proteins
Protein aggregates can be detected through a variety of techniques, each with some advantages and disadvantages. Visualisation in patient material is mainly achieved through amyloid-specific dyes such as Congo Red [65]. The stain was originally developed for the textile industry, but found a new use as gold standard for amyloidosis [66]. The binding of Congo Red to amyloid material, including some smaller oligomeric structures, induces a polarising effect in the dye and the appearance of a bright green light when viewed through a polarising filter, the so called birefringence. In living cells, the most common method to visualise aggregating proteins is through transcriptionally fused fluorescent tags such as GFP [67, 12]. While the attachment of a stable globular protein moiety to a misfolded protein will change protein turnover as well as aggregation dynamics, it is so far the most straightforward way to follow the aggregation process in vivo. A wide range of techniques have been developed, increasing the versatility of fluorescent microscopy. In Bimolecular Fluorescent Complementation (BiFC), the chromophore is divided into two parts with each part fused to a different peptide. If the two peptides interact, the chromophore is reconstituted and becomes fluorescent, allowing researchers to exclusively monitor oligomeric species. [68]. A property of some fluorescent probes is the dependence of the lifetime of the fluorescent signal on the availability of energy acceptors in the local environment. As aSYN aggregates can act as acceptors, this technique can be used to follow protein aggregation in vivo [69]. Despite these and other developments and today’s high resolution imaging techniques, models of structures
and pathways of aggregation rely mostly on *ex vivo* biochemical assays and much research is still needed to find out how these structures arise *in vivo* [70].

### 1.4.1 X-ray crystallography

X-ray crystallography is the most common method of protein structure determination applied today, used for the determination of approx. 90% of the structures submitted to the protein data bank (RCSB.org). As a first step, proteins are crystallised, creating an ordered pattern of identical unit cells. As electromagnetic radiation passes through the formed lattice, it diffracts creating a pattern of constructive and destructive interference. From this pattern, the form that caused the diffraction can be calculated. X-rays are used since meaningful diffraction only occurs if the wavelength does not exceed the size of the object being observed. X-ray crystallography works well with stably folded proteins, but x-ray diffraction is less suitable in the analysis of the semi-ordered structures of amyloids. However, x-ray defraction of fibrils resulting in what is called a cross-beta diffraction pattern remains a gold standard of amyloid fibrils [71]. More recently, peptide sequences from disease-associated amyloidogenic proteins have yealded high resolution structural models of amyloids [15, 72].

### 1.4.2 NMR

NMR is a technique sensitive to the specific chemical environment of a nucleus in a magnetic field. Each nucleus in a magnetic field absorbs radio frequency pulses at different magnetic field strengths due to shielding by electrons surrounding the nucleus. As the differences in field strength are very small, they are measured in parts per million (ppm) and in order to achieve a high signal-to-noise level highly concentrated samples have to be used. In molecular biology, NMR is used both for structural determination and the study of dynamic properties of the molecule, such as intramolecular mobility and binding. Like x-ray crystallography, it is an ensemble method and the obtained results reflect the average structure and behavior of the protein of interest. While NMR can be used to study proteins in solutions, there is a size limitation due to the need for molecular tumbling, which is inversely proportional to the size of the protein or protein complex. This makes the analysis of large structures such as amyloid fibrils impossible in liquid NMR. Although more prone to line broadening and thus to loss of resolution and signal intensity, solid state NMR can be used in the analysis of larger, ordered or semi-ordered structures including amyloid fibrils. This can
yield both general non-residue assigned information, such as homogeneity as well as residue specific structural characterisation [73].

1.4.3 Atomic force microscopy

Atomic force microscopy (AFM) in protein structure studies is a high resolution monomolecular visualisation technique. A small lever is used to probe a sample surface which can yield nanometer resolution, three orders of magnitudes higher than visual light microscopy. In the field of amyloidosis, AFM is commonly used to analyse fibres and oligomers and can be used to distinguish different populations of oligomers as well as compare samples from in vivo and in vitro sources [74]. As the observation is limited to the sample surface, only a limited amount of structural information is obtained. However, in contrast to other high magnification methods, such as Scanning Electron Microscopy, the technique yields three-dimensional information and does not rely on pre-staining of samples.

1.5 Aggregation and compartmentalisation

The cellular response to an aggregation prone protein seems to depend on several factors. Under physiologically healthy conditions, cells are thought to refold and degrade these proteins [75]. However, experiments using fluorescently labeled components of the protein folding machinery or aggregation prone proteins have shown that these labeled proteins can form microscopically visible foci during increased protein folding stress, such as heat-shock, and the presence of mutations in genes leading to destabilisation of proteins [76]. Although protein aggregation is a highly stochastic process and heavily reliant on nucleation events, different types of stress and aggregation prone proteins seem to lead to inclusions with specific compositions and properties [77, 4]. The characterisation of these different types of inclusions in different model systems has been the focus of many recent studies in the field. Aggregates are generally difficult to characterise and might differ qualitatively from each other. As a result, aggregates in literature either have the general description of the aggregate or foci, sometimes including their subcellular location. In some cases they have been given individual classification and some nomenclature seems to emerge, two of which will be presented below.
1.5.1 Insoluble Protein Deposit

Insoluble Protein Deposit (IPOD) is a protein aggregate in yeast associated with vacuoles in the cytoplasm. Proteins are amyloidogenic, non-ubiquinated and do not exchange material with their surroundings [78]. IPODs have been shown to require a functional cytoskeleton indicating that they are formed by an active cellular process. The compartment has been identified in yeast and is similar to mammalian inclusions in regards to insolubility and the association with homologous proteins.

1.5.2 The Juxta-Nuclear Quality control department

The Juxta-Nuclear Quality (JUNQ) control department was identified upon reduction in proteasome function. Formed by the cytoskeleton-dependent accumulation of ubiquinated, misfolded protein and the recruitment of 26S proteasomes, the JUNQ can exchange material with the rest of the cell. JUNQ-like inclusions have been found in samples from familial ALS patients [79].

1.6 Mechanisms of aggregation

Proteins vary widely in their structural stability from tightly folded to highly dynamic. The oxygen storage protein Myoglobin is exclusively α-helical and as a small and highly stable protein was the first protein to be structurally determined using x-ray crystallography [80]. On the opposite side of the spectrum aSyn exemplifies a protein that is natively unfolded, although it has been suggested that it acquires a well-defined structure when bound to membranes [81]. Sometimes these forms even occur within one protein: the heat-shock protein HSP70 is an example of a protein with an inflexible nucleotide binding domain and a C-terminal substrate binding tail, which is highly dynamic in the open and substrate ready state [82]. The high degree of freedom within the peptide chains of a protein allows for a wide range of folded, unfolded and aggregated protein structures (Figure 2), which also includes misfolded or aggregated states, where a part of the native structure and/or function of the protein is intact.
Figure 2: A schematic representation of the different possible stages in the life of proteins.
Proteins synthesised from mRNA by the ribosome fold and refold with the help of chaperones such as the Hsp70 machinery which also participates in diverting misfolded proteins to the Ubiquitin proteasome system for degradation. Misfolding is a common event which is exacerbated by stress which also enhances protein oligomerisation. Some oligomers, referred to as on-pathway oligomers can ultimately form amyloid fibrils and inclusions.
1.6.1 Templating

A hallmark feature of amyloids is their ability to catalyse the conversion of proteins with similar amino acid sequence from non-amyloid into amyloid structure. This seems to occur by the formation of β-sheet-like hydrogen bonds between the template and the substrate peptide, which then becomes a new template molecule with the daughter fibre inheriting the properties of the mother fibre [13, 83, 84]. This monomer addition leads to the growth of the amyloid fibril in a direction perpendicular to the aligned peptides [85]. Sheering of longer fibrils exposes more template surfaces and accelerates the conversion of native proteins [13, 86]. Seeding and templating has also been indirectly shown to occur in a cellular milieu. In a co-culture of cells expressing either Cherry fluorescent protein-Q71 or GFP-Q25, strong aggregation and seeding was observed of the normally non-aggregation prone GFP-Q25 [87]. Interestingly this was dependent on induced cell lysis of the Cherry-Q71 cells. It has also been shown in familial PD patients receiving cell grafts from healthy individuals that aSyn aggregates appear inside the donor cells after 11-16 years. The invasion is thought to occur from the outside inwards and despite the young age of the fetal cells strongly suggesting seeding from the disease afflicted patient cells [88]. In *in vitro* aggregation assays, where the supply of non-aggregated material is high, but finite, the existence of even minute/small quantities of amyloid precursor leads to a sigmoidal growth curve with an initial exponential growth followed by a rapid depletion of non-amyloid material. These assays have been used to quantify and detect amyloid material in biological samples for instance in the demonstration that Aβ42 expressing *C. elegans* deficient in HSF-1 signaling contain increased levels of seeding competent Aβ42 [89].

1.6.2 Prions and spreading

In 1982, approximately 150 years after the development of the germ theory, which proposed microorganisms as a cause of disease, the infectious agent of scrapie was identified as being composed of ‘protein only’ through its sensitivity to proteases and chemicals, while its resistance to UV indicated independence on nucleotide transmission [90]. Experiments suggest that prion proteins can be infectious even as monomers [91]. It is now known that prion diseases, such as Scrapie and Creutzfeldt-Jakob disease, spread from individual to individual though the consumption of food containing prion proteins with homology to human prions with cross-β structure [92]. This allows for the possibility of cross-species infection, something which caused public fear after the spread of bovine spongiform encephalopathy
among cattle destined for meat production [93]. Although proteins might not be identical between animal species, sufficient homology is present to pose a risk of protein conversion. This templating results in specific strains of infectious prion species distinguishable by the fragments formed by proteolytic cleavage (83, 84). There are strong similarities in structure and growth of misfolded protein species between prions and other disease associated aggregation prone proteins. This has suggested that amyloid disorders such as AD, HD and PD can be transferred using a prion like mechanism as well [94].

1.6.3 On-and-off-pathway oligomers
Many proteins prone to amyloid formation have also been shown to form oligomeric structures, which can be a required step in the de novo formation of amyloids, i.e. nucleation rather than monomeric addition [4, 95, 96]. For several disease-associated proteins, more than one type of oligomers has been described, e.g. for aSyn [97, 98]. These have been shown to differ in shape and toxicity, as well as whether or not they can form amyloids. In the case of aSyn, incompatibility to aggregation correlates to a higher level of toxicity. In general, toxicity testing in protein misfolding disorders is performed on cell cultures as delivering specific protein folding intermediates and following their processing in animal models is technically very challenging. However, antibodies raised against specific folding isoforms have been used to identify these oligomers in vivo [99, 100].

1.7 Aggregation and toxicity
As amyloids were discovered in neuronal tissue from patients suffering from NDDs, they were thought to be causing the neuronal damage and the same has been assumed for other types of aggregates. However, more recent research has revealed the need for a more nuanced hypothesis as microscopically visible protein aggregates have been shown to be either uncorrelated with neuronal death or found to be protective in several models of NDDs [9, 10, 101]. Additionally, several mouse models of NDDs, including overexpression of the AD associated proteins Aβ and Tau, recapture many of the behavioral phenotypes of the disease they model, without the presence of aggregates [102]. This indicates that although microscopically visible aggregates are a late consequence of NDDs, they cannot be assumed to be the primary source of toxicity.
A difficulty when analysing the cause of toxicity of misfolded proteins is the clustering of proteins involved in protein folding and degradation around aggregates. This can either be interpreted as a response to the protein aggregate, or the aggregate itself can be seen as a benign end product of an active protective mechanism. In HD, it has been demonstrated that the presence of inclusions correlate with survival of cultured striatal neurons [9, 10]. Instead, cytotoxicity correlates in these experiments with the amount of diffuse mutant Huntingtin (mHTT). This suggests that the formation of inclusions leads to reduced toxicity by reducing the amount of soluble mHTT. Still, the formation of amyloids is unlikely to be harmless. Several studies have demonstrated toxic properties of fibrils in vitro [103, 104] and numerous cellular processes have been suggested as a target of cytotoxic protein aggregation. Some of these are described more in detail below.

1.7.1 Damage to intracellular transport
Damage to intracellular transport such as the one facilitated by the microtubule network, has been proposed as a common denominator in NDDs [105, 106]. The unique form of neurons, with their connections to other neurons far removed from the cell body and with almost all of bulk cell volume in the axonal compartment, suggests that these cells would be especially vulnerable to disruptions in axonal transport. Experimental evidence seem to corroborate this as genes involved in axonal transport have been found to be linked to several degenerative diseases [107-109]. Tau, a microtubule associated protein found in neurofibrilar tangles in AD, has also been implicated in PD and other NDDs [110]. Transport along microtubuli is facilitated by the motor proteins Kinesin and Dynein which bind cargo through the scaffold protein Dynactin. Mutations in the Dynactin subunit p150 were found to be associated with ALS [111]. Another possibility could be a catch 22 mechanism, in which reduced transport impedes clearance of misfolded proteins and might also lead to locally elevated concentration of aggregation prone proteins [112].

1.7.2 Mitochondrial dysfunction
Mitochondrial dysfunction has been linked to neurodegeneration through imperfect metabolic regulation leading to oxidative stress and ultimately to damage of proteins, DNA and lipids. Oxidation of proteins can lead to reversible and irreversible changes in protein structure, such as cross-linking and carbonylation [113]. It has been shown that exposure to reactive oxygen
species (ROS) increases protein hydrophobicity, a strong indicator of protein unfolding which is, besides targeting them for proteasomal degradation, a trigger of protein aggregation [114, 115]. Oxidative stress has been linked to several of the major NDDs. The most direct connection exists in ALS, where one of the aggregating proteins is the globular, dimeric apoprotein SOD1, which clears the cell of superoxide radicals [25]. However, oxidative stress resulting from protein aggregation seems to be unrelated to loss of function of SOD1 as the effect persists in the presence of wild type levels of reactive oxygen scavenging. In AD, increased oxidative damage of proteins and membranes has been found in those brain regions most affected by the disease and genetic evidence links ApoE, a putative regulator of oxidative damage, to familial forms of AD. Furthermore, the dopamine metabolism in the Substantia Nigra produces superoxide as a byproduct, linking PD with oxidative stress.

1.7.3 ROS
ROS have been shown to act as a trigger of apoptosis, meaning that even if a cell manages to cope with any damage caused, its programming might still cause it to die [116, 117]. An additional connection between protein damage and oxidative stress is that critical cellular defenses in both cases are regulated by the insulin signaling pathway [89, 118, 119]. This includes the ROS scavenging protein family SOD described earlier. Animal models with loss of the mitochondrial ROS scavenger SOD2 show increased susceptibility to neuronal damage [120].

1.7.4 Membrane permeation
Membrane permeation resulting in a leakage of ions and small molecules can be caused by oligomers of aSyn, Aβ, mHtt as well as other aggregation prone proteins [121]. It is hypothesised that the confined space close to membranes in combination with the availability of hydrophobic interaction partners can facilitate nucleation of protein aggregation [122, 123]. The oligomers formed have been shown in vitro as well as in patient material to form annular, i.e. doughnut shaped structures with a pore opening of one to two nanometers in diameter [124-128]. Ion leakage and the resulting loss of membrane potential is, just like in intracellular transport, likely to affect neurons differently based on the importance of ion gradients in neuronal signaling.
1.7.5 Protein level imbalance or insufficiency
Protein level imbalance or insufficiency can also be caused by aggregation. Firstly, sequestration of vital cellular components, such as transcription factors, has been proposed to lead to transcriptional imbalance [129, 130]. One example is the cAMP response element binding protein CREB, which usually localises to the nucleus, but which associates with cytoplasmic polyQ aggregates in disease models and patient material [131, 132]. Secondly, many proteins require chaperones to fold correctly as well as components of the degradation machinery for their removal, either upon damage or for the removal of excess protein. There is evidence that amyloidogenic peptides, such as polyQ, can bind the proteasome in the entrance of the proteolytic cavity without being processed further, thus blocking proteasome dependent protein degradation [133].

The broad cytotoxic effects of protein aggregation suggest that it is likely to be more than one toxic species and more than one type of cellular injury for any disease-related protein. This would mean that apoptosis can be triggered by different insults in different tissue types or individuals as imbalances build up during the disease progression. This would also render experimental outcomes highly variable based on experimental setup as they all would differ slightly in the stress imposed on different cellular functions.

1.7.6 Toxic oligomers
The association of misfolded proteins into non-functional or gain-of-function oligomers has lately gained more attention as the possible source of toxicity [11]. While oligomers are a heterogeneous population which in most situations are likely to make up only a minority of the total amount of aggregation prone proteins, they can be expected to display a wide spectrum of \textit{in vivo} properties and their small size and relative stability might result in many opportunities to interact with their cellular environment. Despite the heterogeneity of aggregation oligomers, some characteristics have been proposed as general indicators of toxicity. A strong link is the correlation between exposed hydrophobicity and oligomer toxicity [12]. HYPF-N is a model protein composed of the 91 amino acid N terminal region of an \textit{Escherichia coli} protein with the capacity to form amyloid as well as stable oligomers. HYPF-N has been shown to form distinct types of oligomers that, except for the difference of hydrophobic exposure, are morphologically similar. Still, viability of human SH-
SY5Y cells seems compromised in the presence of the oligomers with the higher hydrophobic exposure. On the other hand, the reliability of the protective amyloid hypothesis is reinforced by the finding that the exposure of cells to off-pathway oligomers, \textit{i.e.} those that \textit{in vitro} have been found to have negligible amyloid formation, induces more cellular deaths than the exposure to their on-pathway isoforms [134]. As oligomers exist in so many varieties even within one disease, and any interactions can be specific to that cell type, forming a clear model that describes oligomer toxicity and interactions will be difficult. While some correlations, such as the one between hydrophobicity and toxicity are possible, other generalisations might have little predictive value. The discovery that most amyloidogenic proteins have prion-like activity, has put more emphasis on to what extent oligomers can carry infection from cell to cell.

1.8 Cellular defences to proteotoxicity

Cellular proteins range from stable to highly dynamic and can be affected by reactive compounds in their environment [75, 135]. This leads to a constant challenge on protein function, structure and stability. This pressure is matched in healthy cells by refolding and degradation through a complex system of cellular defenses. Cells have several responses to protein aggregation which are generally well conserved indicating the ubiquitous need to deal with aberrant protein folding [136]. These responses aim to refold, remove or restrict exposure to misfolded or otherwise damaged proteins. Refolding is mainly facilitated by molecular chaperones, a class of proteins involved in recognising and assisting in the folding of misfolded proteins [137]. They reversibly bind to exposed hydrophobic patches which are characteristic for misfolded proteins. For example, the eukaryotic CCT/TRiC complex and its prokaryotic homologue GroEL are central to the correct folding of newly synthesised proteins as well as misfolded proteins. Another example is Hsp70, a highly conserved ATP-dependent chaperone, which has been found to bind to exposed hydrophobic residues of other proteins. It is a central component of protein folding- and degradation and works in collaboration with many other protein groups such as DNAJ proteins. These are generally involved in \textit{de novo} recognition of unfolded proteins and nucleotide exchange factors (NEF) that accelerate the ADP to ATP exchange for Hsp70 and thus its activity [138]. NEFs have also been reported to be involved in the regulation of proteins that will be refolded and those that will be degraded [139]. The latter pathway involves the ubiquitin-proteasome cascade where the small protein ubiquitin is attached one by one in a chain to a misfolded protein by ubiquitin ligases and
serve as a recognition signal for downstream degradation. The marked protein is delivered to the proteasome, a 2 MDa barrel-like structure that unfolds the protein and leads it through its centre pore, where it is cleaved into small peptides [2]. Proteins that cannot be processed though this system, as well as larger cellular structures, can be processed though autophagy [140].

1.9 Modifiers of protein aggregation
Despite the remaining uncertainty regarding the sources of toxicity in protein aggregation, it is important to find genes influencing the rate and pathway of protein aggregation to acquire tools for further investigation and to identify potential drug targets. Genetic screens for aggregation and toxicity in cell culture as well as in animal models have so far found modifiers of aggregation in many different protein classes, cellular functions and through-out the cellular compartments [56, 141-144]. Meta-analysis of such screens show an overrepresentation in modifiers of aggregation and toxicity among genes in protein folding and degradation pathways [145]. Other cellular processes have also been identified in multiple screens, but appear to be more disease specific, such as vesicle transport for models PD and RNA processing for HD.

1.10 Aims and outline of this thesis
New biochemical tools developed in recent years have enabled big steps in the understanding of protein aggregation in vitro, while advances in genetics have led to the identification of modifiers of aggregation as well as cellular pathways and networks involved. However, there are still large gaps in our understanding on how cellular components affect the aggregation process in vivo, partly because there is a lack of mechanistic understanding about modifiers identified in screens and partly because these modifiers are often highly dependent on other factors leading to complex answers. One way forward would be to try to reduce complexity in screens for modifiers. This can be accomplished by using model systems without the specific vulnerabilities of different types of cells responsible for the variations in symptoms of different NDDs. Furthermore, removing non-amyloid components of the aggregating peptide reduces the number of interactions, hopefully increasing the chances of finding amyloid-specific modifiers.
Chapter 2
In order to understand the cellular regulation of protein aggregation, a *C. elegans* model for polyQ aggregation was used in a screen using forward mutagenesis to identify modifiers of aggregation which we named Modifiers of aggregation or MOAG. The mutation in *moag-4* mapped to a gene on chromosome 1, which encoded a protein of unknown function, hereafter referred to as MOAG-4. The *in vivo* and *in vitro* characterisation of this modifier of aggregation yielded a model where the modifier acted on a small, possibly monomeric, pre-aggregate polyQ conformer. MOAG-4 also demonstrated a general enhancing effect on amyloid aggregation as the effect extended to aSyn and Aβ, but not to SOD1 aggregation and this effect is conserved in SERF1A and SERF2, the human orthologs of MOAG-4. For clarity, the author of this thesis participated in the research design and conducted and analysed the experiments resulting in figures 1c, 2, 6a, S1, S2, S4 (shared) and supervised research and analysis for figure 4.

Chapter 3
In order to efficiently search for changes in the distribution of polyQ protein folding and multimerisation states, we developed an electrophoretic screening method called NAGE used in screening modifiers of aggregation for changes in aggregation intermediates. It employs the common use of fluorescent tags to follow aggregation prone proteins *in vivo* for electrophoretic separation and analysis.

Chapter 4
A method was developed to facilitate rapid cloning as well as efficient expression and purification or proteins through yeast homologous recombination. The method uses a His-SUMO-tag which allows efficient recovery through the His tag and highly specific cleavage through yeast SUMO protease ylp-1. Additionally, globular tags such as SUMO can increase solubility for misfolded or otherwise aggregation prone proteins. The vector and cloning described provides a high selection for correct inserts with small time investments and has a high recombination rate that is insensitive to multiple fragment cloning.

The scientific findings of this thesis are summarised and brought into the context of the field of NDD research in Chapter 5 and in Chapter 6 an outlook is given on further investigations into the effect of Moag-4/SERF.
Reference list

24


28


