Universitat Autònoma de Barcelona

A multidisciplinary insight into the determinants of protein aggregation

Anita Čarija

ibb
Institut de Biotecnologia
i de Biomedicina

UAB
Universitat Autònoma
de Barcelona

Universitat Autònoma de Barcelona
Departament de Bioquímica i Biologia Molecular
and
Institut de Biotechnologia i Biomedicina

A multidisciplinary insight into the determinants of protein aggregation

Anita Čarija

Bellaterra, September 2017

Universitat Autònoma de Barcelona
Departament de Bioquímica i Biologia Molecular
and
Institut de Biotechnologia i Biomedicina

A multidisciplinary insight into the determinants of protein aggregation

Doctoral thesis submitted by Anita Čarija in candidacy for the degree of Ph.D. in Biochemistry, Molecular Biology and Biomedicine from the Universitat Autònoma de Barcelona.

The described work has been performed at the Department of Biochemistry and Molecular Biology and at the Institute of Biotechnology and Biomedicine, under the supervision of Prof. Salvador Ventura Zamora.

Anita Čarija Prof. Salvador Ventura Zamora

Bellaterra, September 2017

Summary

Chronic neurodegenerative disorders, the medical conditions that strike primarily mid- to late-life population, represent a major issue of modern society. Therefore, finding new diagnostic and therapeutic approaches to treat these disorders is a goal of increasing urgency. The neurodegenerative disorders are characterised by intra/extracellular protein misfolding, resulting in the formation of ordered aggregates that are responsible for the onset of these diseases. On the other hand, aggregation represents a major limitation in the industrial production of proteinaceous therapeutic agents. To elucidate the causes behind the formation of these insoluble deposits, the mechanisms by which they mediate cellular toxicity, and how evolution confronts this risk, we employed a multidisciplinary approach to study the determinants of aggregation reactions, using different protein models such as, amyloid β-peptide and α-synuclein.

The research presented in this thesis seeks to understand the determinants of protein aggregation and its associated toxicity, in both the cellular environment and *in vitro* conditions, as well as to investigate how the selective pressure that acts to to avoid the aggregation has shaped the cellular proteomes along the evolution. Implementing a novel structure-based bioinformatic tool, we identify the structural determinants of protein aggregation using bacteria as a model organism. Little is known on the structural determinants that drive the aggregation of a protein to a particular pathway, resulting in diverse aggregated macromolecular structures displaying different toxicity. Here, we address this issue using the Parkinson's disease-associated intrinsically disordered protein, α-synuclein. Finally, using another intrinsically disordered protein, the amyloid β-peptide and mutants thereof, we identify the conformational species responsible for the cellular oxidative damage caused by the aggregation of this Alzheimer's linked peptide, employing yeast as a model system.

Overall, the work in this thesis attempts to understand fundamental aspects of protein aggregation processes, in both prokaryotic and eukaryotic organisms, highlighting how the interplay between different disciplines might improve our understanding on the impact of protein aggregation in health and disease.

Resum

Els trastorns neurodegeneratius crònics, condicions mèdiques que afecten a la població principalment a la seva darrera etapa de vida, representen un problema molt important a la societat moderna. Per això, trobar nous mètodes de diagnòstic i teràpies per tractar aquestes patologies representa un objectiu que cada vegada es presenta com més urgent. Les malalties neurodegeneratives estan caracteritzats pel malplegament proteic intra i extracel•lular, que deriva en la formació d'agregats ordenats responsables de l'inici d'aquestes patologies. Per altra banda, l'agregació representa la major limitació en la producció d'agents terapèutics de caràcter proteic. Per desvetllar les causes que es troben darrera de la formació d'aquests dipòsits insolubles, els mecanismes pels quals provoquen toxicitat cel•lular i com l'evolució enfronta aquest risc, hem utilitzat un enfoc multidisciplinar per tal d'estudiar els determinants de les reaccions d'agregació, utilitzant diferents models proteics com el pèptid β-amiloide i l'α-sinucleïna.

La recerca que es presenta en aquesta tesi persegueix entendre els determinants de l'agregació proteica i la toxicitat que s'hi associa, tant en l'entorn cel•lular com en condicions *in vitro*, així com investigar com la pressió selectiva ha modelat els proteomes al llarg de l'evolució per tal d'evitar l'agregació. Implementant una nova eina bioinformàtica basada en l'estructura tridimensional de les proteïnes, hem identificat els principals determinants estructurals de l'agregació proteica utilitzant bacteris com a organisme model. Es coneix molt poc sobre els determinants estructurals que condueixen l'agregació d'una proteïna cap a una via determinada i com aquests resulten en diverses estructures macromoleculars agregades que exhibeixen diferent toxicitat. En aquest treball, abordem aquest problema usant l'α-sinucleïna, proteïna intrínsecament desordenada associada a la malaltia de Parkinson. Finalment, utilitzant una altra proteïna intrínsecament desordenada, el pèptid β-amiloide i mutants d'aquest, hem identificat les espècies conformacionals responsables del dany oxidatiu cel•lular causat per l'agregació d'aquest pèptid relacionat amb l'Alzheimer, utilitzant llevat com a sistema model.

De manera global, el treball d'aquesta tesi pretén entendre aspectes fonamentals del procés d'agregació proteic, tant en organismes procariòtics com en eucariòtics, il•lustrant cóm la la integració de diferent disciplines pot millorar el nostre coneixement sobre l'impacte de l'agregació de proteïnes en la salut i la malaltia.

Abbreviations

A3D	Aggrescan3D
AD	Alzheimer's disease
AICD	Amyloid precursor protein intracellular domain
APP	Amyloid precursor protein
APR	Aggregation-prone region
ATM	Atomic force microscopy
Aβ	Amyloid-β-peptide
Bis-ANS	Bis(1-anilinonaphthalene 8-sulfonate)
BPTI	Bovine pancreatic trypsin inhibitor
CAT	Catalase
CD	Circular dichroism
CHC	Central hydrophobic cluster
CR	Congo Red
CV	Coefficient of variance
DAVID	Database for Annotation, Visualization and Integrated Discovery
DMPS	Dimyristoylphosphatidylserine
DNPH	Dinitrophenylhydrazine
DOPC	Dioleoylphosphatidylcholine
EP	Essential protein
ER	Endoplasmatic reticulum
ESI-MS	Electrospray Ionization Mass Spectrometry
FC	Flow cytometry
FTIR	Fourier-transformed infrared spectroscopy
GO	Gene ontology
H	Enthalpy
HD	Huntington's disease

HDAC6	Histone deacetylase 6
IAM	Iodoacetamide
IDP	Intrinsically disordered protein
IM	Inner membrane
IP	Propidium iodide
IPOD	Insoluble protein deposit
JUNQ	Juxtanuclear quality-control compartment
LUV	Unilamellar vesicles
M	Protein active as monomer
MG	Molten Globule
MLV	Multilamellar vesicle
MToC	Microtubule-organizing centre
MW	Molecular weight
NAC	Nascent polypeptide-associated complex
NAC	Non-amyloid-β component
NMR	Nuclear Magnetic Resonance
NP	Non-essential protein
NP	Nucleation-dependent polymerization
O	Protein active as oligomer
OM	Outer membrane
PD	Parkinson's disease
PDB	Protein Data Bank
PI	Protein inclusion
PK	Proteinase K
PQC	Protein Quality Control
PrLD	Prion-like domain
PSSM	Position-specific scoring matrix
PTM	Post-translational modification

RAC	Ribosome-associated complex
ROS	Reactive oxygen specie
RP-HPLC	Reverse-Phase High Performace Liquid Chromatography
S	Entropy
SAP	Structural aggregation propensity
SC-URA	Synthetic complete ura- dropout medium
SOD1	Superoxide dismutase 1
SSI	Structural supersaturation index
T2DB	Type 2 diabetes mellitus
TEM	Transmission electron microscopy
TF	Trigger factor
TFE	Trifluoroethanol
Th-T	Thioflavin-T
Tm	Melting temperature
TS	Transmembrane segment
TTR	Transthyretin
UPS	Ubiquitin-proteasome system
α-syn	α-synuclein

Content

1. Introduction

1.1. Protein folding

Proteins, or polypeptides, are biological polymers which are made from a combination of 20 amino acids that possess distinct chemical and physical properties due to their caracteristic side chains (Table 1.1). Proteins have a central role as executors of diverse fundamental biological functions, in many occasions through the establishment of transient interaction networks with other proteins or biological molecules. In most cases, in order for a protein to be able to perform its intended biological function, it has to adopt a specific three-dimensional structure, known as the native state (Anfinsen, 1973; Dobson, 2003). The process in which a polypeptide chain acquires its native state three-dimensional structure is referred as protein folding.

Protein structure is organized in hierarchical levels, ranging from primary to quaternary structure (Branden, 1998; Herczenik and Gebbink, 2008). After the polypeptide chain is synthesized on the ribosome in a linear form, the establishment of local intramolecular contacts, driven by a specific primary sequence, determines the secondary and tertiary structure. Such local interactions, dominated by the establishment of hydrogen bonds within the polypeptide backbone, result in the formation of regular secondary structure elements, such as α-helices and β-sheets. Further folding of these elements into more compact conformations through the establishment of novel hydrogen bonds, electrostatic interactions, Van der Waals packing and disulfide bonds, results in formation of tertiary structure. Finally, the spatial combination of folded polypeptide chains into homomeric or multimeric structures, usually driven by the establishment of intermolecular non-covalent interactions, adds an additional level of complexity referred as a quaternary structure (Herczenik and Gebbink, 2008).

The timescale in which proteins fold into their functional, three-dimensional structure can range from milliseconds to seconds (Anfinsen, 1973). During the process of protein folding, there is a tremendously high number of different conformations that a protein can adopt, regarded as folding intermediates, some of these conformations being more preferential than others (Dill and Chan, 1997). This structural plasticity is indeed behind the functional versatility of proteins (Creighton, 1993). However, through millions years of evolution protein sequences have been optimized to fold efficiently and robustly into their native states (Itzhaki and Wolynes, 2008), in such a way that even subtle changes in the functional conformational arrangement of proteins would require multiple mutations in the original protein sequence.

Table 1.1. Table with 20 natural amino acids. Depending on the physic-chemical properties of the side chains, each amino acid is coloured in corresponding colour: hydrophobic are in grey, positively charged are in orange, negatively charged are in blue, uncharged are in green, aromatic are in purple and special cases are in beige.

Letter Code	Amino acid	Abbreviation	Side chain
A	Alanine	Ala	Hydrophobic
C	Cysteine	Cys	Thiol group
D	Aspartic acid	Asp	Negative
E	Glutamic acid	Glu	Negative
F	Phenylalanine	Phe	Aromatic
G	Glycine	Gly	Tiny
H	Histidine	His	Positive
I	Isoleucine	Ile	Hydrophobic
K	Lysine	Lys	Positive
L	Leucine	Leu	Hydrophobic
M	Methionine	Met	Hydrophobic
N	Asparagine	Asn	Polar uncharged
P	Proline	Pro	Structurally rigid
Q	Glutamine	Gln	Polar uncharged
R	Arginine	Arg	Positive
S	Serine	Ser	Polar uncharged
T	Threonine	Thr	Polar uncharged
V	Valine	Val	Hydrophobic
W	Tryptophan	Trp	Aromatic
Y	Tyrosine	Tyr	Aromatic

The understanding of molecular basis of protein folding and unfolding has become a key topic in biology because the mistakes in the acquisition or the maintenance of the native state are being linked to an increasing number of fatal diseases (Chiti and Dobson, 2006). Moreover, the understanding of protein folding principles represents a major clue to enable the successful manipulation of protein structures and functions which is the main goal of a relatively young discipline called protein engineering.

1.1.1. The concept of energy landscapes within different protein folding models

From a biophysical point of view, protein folding is a physical process driven by classical kinetic and thermodynamic principles in which the polypeptide gains compactness as it approaches its native state. In order for a polypeptide chain to spontaneously fold, both kinetic and thermodynamic requirements must be fulfilled. In this sense, the native state should be thermodynamically stable and have a sufficient low activation energy barrier to permit the polypeptide folding within an acceptable timeframe. Furthermore, proteins under natural selection are expected to have evolved an optimal stability to perform their function in cellular environment (Doig and Williams, 1992).

Thermodynamic protein stability is commonly described by the difference in Gibbs free energy (ΔG) between the native and the unfolded state. In this sense, a negative ΔG indicates that the native state is more stable than the denaturated state, due to the stabilising forces in the native fold (Branden, 1998). The two factors that determine ΔG are changes in enthalpy (ΔH) and entropy (ΔS), described by the equation:

$$\Delta G = \Delta H - T\Delta S$$

where T denotes the absolute temperature in Kelvin.

The enthalpic difference in protein folding process derives from the energy of non-covalent interactions, such as hydrogen bonds, electrostatic forces and Van der Waals interactions, and it represents molecular stability. On the other hand, the entropic difference comes from the structural flexibility. The basic principle of protein folding is to establish a balance between enthalpy and entropy, therefore, for protein to acquire sufficient structural stability to maintain the native conformation, but also to keep certain flexibility to be able to perform its biological functions. In the native state, the non-covalent interactions are more frequent compared to the denaturated state and consequently, their energy contribution is larger but this comes at expenses of a lower system entropy.

The process of protein folding is usually composed of series of states that correspond to local minimums of energy fluctuations in the polypeptide conformation in an aqueous solvent, between which the interconversion can easily happen. The transient nature of the species populated during the folding process, make it difficult to explore the structural details of these intermediates and accordingly the molecular mechanism leading to protein folding. Even though the importance of intermediates in many protein folding pathways, their existence is not an essential requirement for an efficient folding as it has been demonstrated by various studies describing folding of small monomeric proteins through a simple two-state kinetic mechanism ($U \leftrightarrow N$) (Jackson, 1998) (Fig. 1.1A). When folding reactions are more complex and include intermediates, they can be either on- ($U \leftrightarrow I_{on} \leftrightarrow N$) or off-pathway ($I_{off} \leftrightarrow U \leftrightarrow N$) (Fig. 1.2B).

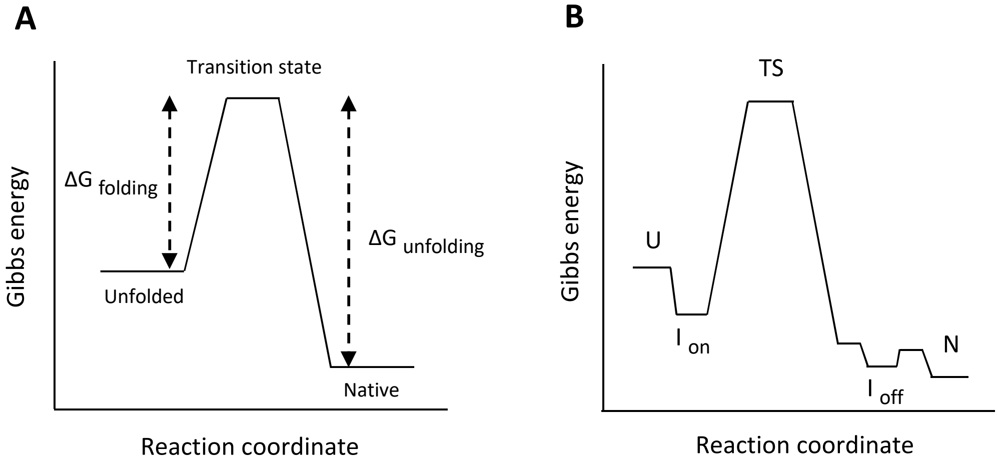


Figure 1.1. Schematic representations of the Gibbs energy diagrams for different protein folding models. (A) Two-state model; (B) Multi-state model.

In the case of the first model, only two states can be populated, the unfolded or denaturated (U), in which the protein lacks secondary structure and other higher levels of structural organization, and the native state (N). In this context, the native state of a protein is defined by the most stable, ordered and functional structure, while the unfolded state is characterised by lower stability, lower degree of order, and high entropy. For some proteins, such as cytochrome c (Ohgushi and Wada, 1983), it has been demonstrated that there exists a third thermodynamic metastable state different both from the native and the denatured state, called the molten globule (MG), in which a protein has a native-like secondary structure content and a relatively dynamic tertiary structure.

Among numerous models that have been proposed in order to describe the common folding mechanism necessary for proteins to reach their native state, the hydrophobic collapse model assumes that hydrophobic forces promote a rapid compaction of the structure, reducing the conformational search towards the native structure (Baldwin, 1989). A general hydrophobic collapse might promote the formation and stabilization of partially formed secondary and tertiary structures (Fersht, 1995; Itzhaki, 1995), so in later variants of this model, the molten globule intermediate was incorporated, combining the formation of a secondary structure with the hydrophobic collapse mechanism (Ptitsyn, 1995).

The versatility of possible conformations that a polypeptide chain can acquire towards its functional conformation is represented by theoretical diagrams called energy landscapes (Fig. 1.2). The funnel model represents the process of protein folding as a progression from a highly heterogeneous ensemble of unfolded states towards the native structure along an energy landscape (Dill and Chan, 1997) assuming the existence of multiple folding pathways involving different intermediates (Radford, 2000; Jahn and Radford, 2008). The shape of an energy landscape depends on the polypeptide sequence and the environmental conditions. In this sense, the size of proteins determines the appearance of their energy landscapes, large proteins having visually more complex, crooked landscapes with numerous local minima corresponding to different intermediates, since their folding pathway progresses through numerous partially folded conformations (Dill *et al*., 2007; Dill and MacCallum, 2012; Jahn and Radford, 2008).

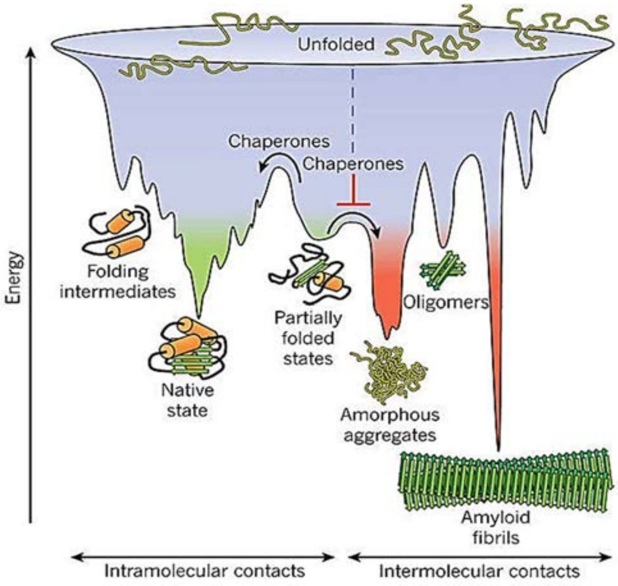


Figure 1.2. Energy landscape of protein folding and misfolding. Driving forces towards the acquisition of the native state, associated with an increase in conformational stability, are energetically favourable intramolecular interactions (represented with green colour). Alternatively, intermolecular contacts can lead to the formation of toxic species (represented with red colour). From the unfolded state towards the natively folded state, protein folding and protein aggregation are competing reactions. Adapted from: (Hartl *et al*., 2011).

1.1.2. Folding of disulfide-rich proteins

The folding of proteins enriched in disulfide bonds represents a specific case within protein folding due to the important influence that disulfide formation has on the kinetics and thermodynamics of folding process. The role of disulfide bridges in protein folding was initially addressed using Rnase A and bovine pancreatic trypsin inhibitor (BPTI) as a disulfide protein models. In order to describe variations in the mechanism of folding, more than 30 small disulfide-rich proteins have been characterised in the last 20 years.

The folding of polypeptides enriched in disulfide bonds is called oxidative folding. During the oxidative folding, there are two events that need to occur: the formation of native cysteine pairings, and the acquisition of the native conformation. As it has been demonstrated, all disulfide-rich proteins populate intermediates due to the marginal stabilization provided by formation of disulfide bridges. The establishment of covalent cysteine bonding allows both native and non-native cysteine pairings to be formed along the folding pathways. The intrachain covalent crosslinking between cysteines allows a stabilization of the native state, consequently reducing the entropy of the unfolded state and increasing the stability of protein (Trivedi *et al.*, 2009).

This property is especially important for proteins who need to be active to perform their biological role in harsh environmental conditions, such as the extracellular space. Proteins enriched in disulfide bonds are rarely found in the cytoplasm due to the reductive nature of this cellular compartment. Unlike the cytoplasm, the periplasm of bacteria, the endoplasmatic reticulum (ER) together with the intermembrane mitochondrial space of eukaryotic cells, represent oxidizing environments where the formation of disulfide bonds naturally occurs. From the evolutionary point of view, the possibility of crosslinking regions that are far away in the sequence through disulfide bridges represents an important feature that seems to be conserved, since when it is disrupted by mutation it tends to have a deleterious effect. The cysteine residues involved in disulfide bonds tend to be conserved through evolution because of the effect they have both on folding kinetics and the stability of the native state (Hamada *et al.*, 2009).

Furthermore, bioinformatics analysis of human proteome demonstrated for the extracellular proteins that the high intrinsic sequential aggregation propensity that these proteins display is correlated with a high occurrence of disulfide bridges (Mossuto *et al.*, 2011), suggesting the importance of disulfide bond formation in the context of extracellular space where the Protein Quality Control (PQC) machinery cannot longer assist proper folding, and therefore, the risk of protein aggregation is higher.

1.2. Protein aggregation

There is a large number of devastating human diseases that have been related to the formation of insoluble protein deposits, such as Alzheimer's disease (AD), Parkinson's disease (PD), Huntington's disease (HD), prion-associated transmissible spongiform encephalopathies, and type II diabetes (Fernandez-Busquets *et al.* 2008, Invernizzi *et al.*, 2012; Dobson, 2003). The living cell is an environment that is inherently crowded by various macromolecules, and as such needs a tight regulation of their interactions. Under normal conditions, nascent polypeptides efficiently assemble into their unique native state. However, the failure of polypeptides to acquire their native state can result in their misfolding, and consequently, in the exposure of hydrophobic stretches that can cause identical polypeptides to cluster into high order assemblies. Due to the medical relevance of the diseases caused by protein self-association, the understanding of protein aggregation mechanism represents a key stepping stone that needs to be mastered in order to enable further progress towards finding efficient therapeutic strategies for these pathologies.

1.2.1. Competition between protein folding and aggregation

Protein folding and aggregation are both driven by the same forces, the hydrophobic effect and hydrogen bonding. The kinetic competition between these two pathways occurs when the hydrophobic groups of proteins become exposed. During protein translation on the ribosomes, the elongation of the nascent polypeptide chain is often faster than the folding process, which consequently can increase the presence of totally or partially unfolded species inside the cell. If protein folding is successful, the hydrophobic groups would be buried in the protein core by intramolecular interactions. Alternatively, the burial can also occur in an intermolecular fashion (Baynes and Trout, 2004; Vendruscolo and Dobson, 2013). Furthermore, when the intrinsic folding rate of the protein is slow, it is possible that crowding after synthesis also enhances aggregation by promoting the accumulation of totally or partially unfolded species (Golding and Cox, 2006; Ellis and Pinheiro, 2002). In this context, it has been reported that the charge of primary sequence influences the elongation rate, the positively charged sequences (rich in Arg/Lys residues) slowing down the translation, and the neutral and negative ones promoting the elongation at the higher rate (Requiao *et al.,* 2017).

Under normal conditions, the native state (N) should dominate over the unfolded (U) and partially folded intermediate (I) conformations. However, even small changes in environmental conditions, such as thermal fluctuations, can result in crossing of the low-energy barrier of unfolding, and generate locally unfolded states (N*). Biologically, these states are relevant and required in order to permit functions, such as ligand binding and enzyme catalysis (Kern and Zuiderweg, 2003). If the generated N* states possess sufficient exposure of hydrophobic groups, the protein residing in this state can become prone to aggregation through hydrophobic interactions. Moreover, the establishment of intermolecular hydrogen bonding between amide NH and the amide carbonyl groups of the backbone belonging to hydrophobic stretches further enhances the stability of aggregates once protein aggregation process started (Fitzpatrick *et al*., 2011). Indeed, aggregation can occur from any conformational state where a protein has hydrophobic groups exposed, the fully unfolded state, any partially folded states or even the folded state (Chiti and Dobson, 2006) (Fig. 1.3).

During the evolution, a number of complementary structure and sequence-based strategies have been developed to reduce the aggregation risk, as well as the PQC system composed of molecular chaperones and proteases, that guarantee correct folding of polypeptides and degradation of misfolded forms (Monsellier and Chiti, 2007). The ribosome-binding chaperones bind and shield hydrophobic patches in the newly synthesized chain from the environment. These regions can also have a crucial role in the formation of the quaternary structure of proteins or in the assembly of protein complexes (Castillo and Ventura, 2009) but when exposed for a long time they represent an inherent aggregation risk.

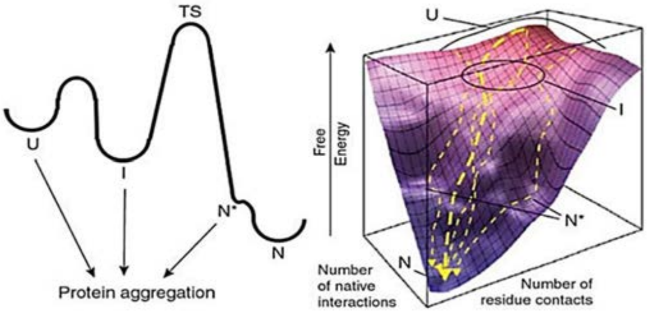


Figure 1.3. Kinetic profile and thermodynamic energy landscape indicating different stages at which protein aggregation can occur. U = unfolded, I = intermediate, N = native, N*= native-like locally unfolded state with low energy barrier, TS = transition state. Adapted from: (Chiti and Dobson, 2009).

1.2.2. Determinants of protein aggregation

The proteins involved in the pathologic amyloidogenic disorders do not have any sequential or structural similarity (Chiti and Dobson, 2006). However, independently of their sequence and native conformation, when these proteins aggregate into their toxic species they undergo the conversion to highly ordered macromolecular assemblies that share a common fibrillar architecture, known as amyloid fibrils (Fernàndez-Busquets *et al.*, 2008).

Amyloid fibrils are characterised by a particular polypeptide backbone organization in a cross-β supersecondary structures, where parallel β-sheets extend with their strands facing to each other leading to a succession of contiguous β-strands stacked perpendicularly to the fibril axis (Serpell *et al*., 2000; Stromer and Serpell, 2005; Chiti and Dobson, 2006) (Fig. 1.5). This particular structure of amyloidogenic insoluble deposits allows their characterisation using high resolution structural techniques like solution NMR and X-ray diffraction, as well as morphological analysis using techniques such as transmission electron microscopy (TEM) and atomic force microscopy (ATM), or by staining with amyloid-tropic dyes such as Thioflavins and Congo Red, whose spectral properties change upon binding to amyloid fibrils (Carrio *et al*., 2005; Nilsson, 2004). In the case of X-ray diffraction, it is possible to detect the diffraction pattern of cross-β supersecondary structure, characteristic for amyloids, amyloid-like fibrils, inclusion bodies or amorphous aggregates (Fig. 1.4). The diffraction pattern displayed at 4.7 Å is consistent with the spacing between extended protein chains running approximately perpendicular to the fibril axis (inter-strand distance) while the diffraction at 8–10 Å indicates the spacing between adjacent β-sheets in the protofibril.

The conformational changes that occur during the conversion of amyloidogenic polypeptides to amyloid fibrils can be followed and identified by secondary structure analysis techniques such as Fourier-transformed infrared spectroscopy (FTIR), where we can monitor the appearance of the characteristic intermolecular β-sheet band at 1625 (Calero and Gasset, 2004; Sabate and Ventura, 2013; Natalello and Doglia, 2015) or by circular dichroism (CD). Also, the resistance to proteolytic digestion with proteinase K is applied to detect the presence of buried and densely packed β-strands. Seeding experiments are also employed to investigate the seeding capacity characteristic of amyloid assemblies. The addition of pre-formed amyloid nuclei of the same protein should reduce the lag phase of the amyloidogenesis reaction.

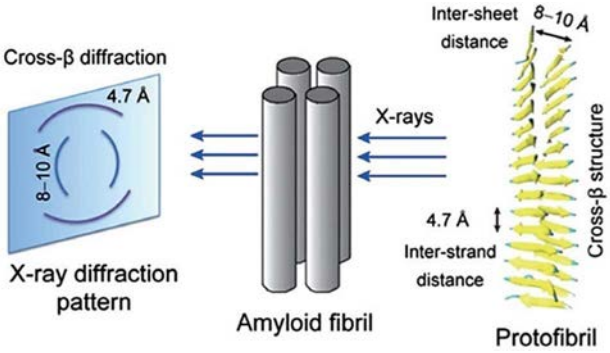


Figure 1.4. A diagram representing the x-ray diffraction pattern and characteristic cross-β structure feature of amyloid fibrils consisted of multiple protofibrils. The diffuse reflection at 4.7 Å spacing along the meridian (vertical) shows extended protein chains running approximately perpendicular to the fibril axis. The increasingly diffuse reflection at 8–10 Å spacing along the equator (horizontal) shows that the extended chains are organized into sheets spaced 8–10 Å apart. Adapted from: (Wen-Hui and Guang-Hong, 2016).

The process of protein aggregation into abnormal insoluble deposits is influenced by properties that are intrinsic to the polypeptide chains, as well as by the external elements (DuBay *et al.*, 2004). Among these inherent physico-chemical properties of amyloidogenic peptides that have been described as favourable determinants of the protein aggregation, hydrophobicity, the structural suitability to adopt β-conformations, the presence of β-strands in the native state, and total polypeptide charge have been intensively studied. Hydrophobicity is one of the major determinants of protein aggregation (Chiti *et al.*, 2003). Mutagenesis studies have shown that the replacement of polar residues by the non-polar ones tends to increase the overall aggregation propensity and the deposition rate of polypeptides.

The propensity to adopt specific secondary structure motifs also increases the likeness of protein aggregation. In this sense, Pro and Gly, so-called β-breakers, have been identified as amino acids that significantly disrupt aggregation-prone stretches when introduced into a primary sequence (Rauscher *et al.*, 2006; Jiang *et al.*, 2009). The charge of amino acids that compose the peptide chains can also disrupt the protein aggregation into insoluble deposits, acting as the basis for intermolecular repulsion and in such a way diminishing the possibility of establishment of the set of contacts necessary to initiate the protein aggregation process.

The extrinsic determinants can also influence the likeness of a polypeptide to converge into insoluble deposits. The most relevant extrinsic determinants are pH, the temperature of the system and the ionic strength of the solution where the proteins reside (DuBay *et al*., 2004). The pH influences the net charge of the polypeptides through the protonation/deprotonation of amino acids side chain, and in such a way it modifies the repulsion/attraction between individual molecules. The presence of salts can modulate the aggregation kinetics and the morphology of aggregates (Morel *et al*., 2010), usually resulting in shorter lag phases.

Furthermore, the interactions with other proteins, such as chaperones, proteases and the components of the ubiquitin-proteasome system, can also modify the tendencies of proteins to self-assemble. For example, in the case of α-synuclein, a protein involved in Parkinson's disease which belongs to the intrinsically disordered proteins (IDPs), it has been reported that some evolutionarily conserved proteins, such as MOAG-4, can promote its aggregation, due to the establishment of transient interactions with the terminal regions of synuclein (Yoshimura *et al*., 2017) (Fig. 1.5).

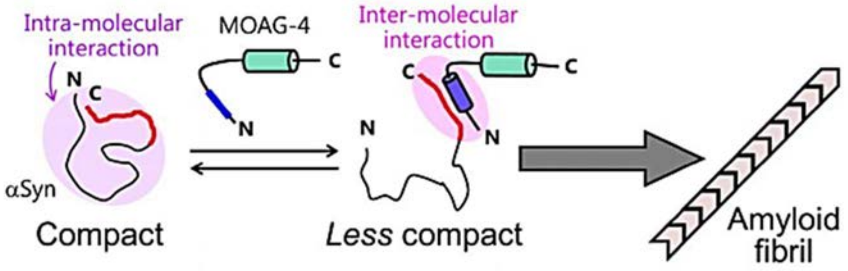


Figure 1.5. A schematic representation showing the effect of the established intermolecular interaction between negatively charged C-terminus of α-synuclein and a positively charged segment of MOAG-4. Due to the loss of the existing intramolecular interactions between C- and N-terminus of α-synuclein, the aggregation kinetics have been enhanced and the self-assembly of the protein has been promoted. Adapted from: (Yoshimura *et al*., 2017).

1.2.3. Mechanism of protein aggregation

As described, folding and protein aggregation are processes that kinetically compete in the cell. In order for the functional native globular proteins to undergo the multi-step self-assembling process, destabilization by mutations or external factors should usually occur. This destabilization commonly exposes sequence regions of a protein that are hidden in the native structures, and are essential for the protein to adopt its native fold, that at this point become available to establish non-functional intermolecular contacts (Vendruscolo *et al*., 2003; Castillo and Ventura, 2009). These regions of sequence called aggregation-prone regions

(APRs) are small amino acid stretches which promote and guide protein deposition and are commonly enriched in hydrophobic, aliphatic (Val, Le, Ile) and aromatic (Phe, Trp, Tyr) residues (Rousseau *et al*., 2006). As mentioned above, several structurally diverse conformations are able to act as precursors of protein aggregation, such as unfolded polypeptides, oligomeric species or even IDPs (Fig. 1.6).

Although different models have been proposed to explain the mechanism of amyloid formation, perhaps the model that has been more widely accepted is the Nucleation-dependent Polymerization (NP) mechanism that describes the formation of amyloid-like structures as a triphasic process (Jarrett and Lansbury, 1993) (Fig. 1.7). The first step is a slow thermodynamically unfavourable process in which the protein concentration plays an important role and it includes the formation of oligomers by an initially monomeric form. Only when the monomer concentration exceeds a certain level known as critical concentration, the lag phase can be observed (Merlini and Bellotti, 2003). This stage can be reduced by the addition of preformed prefibrillar aggregates, known as seeds (Sabate *et al.*, 2010).

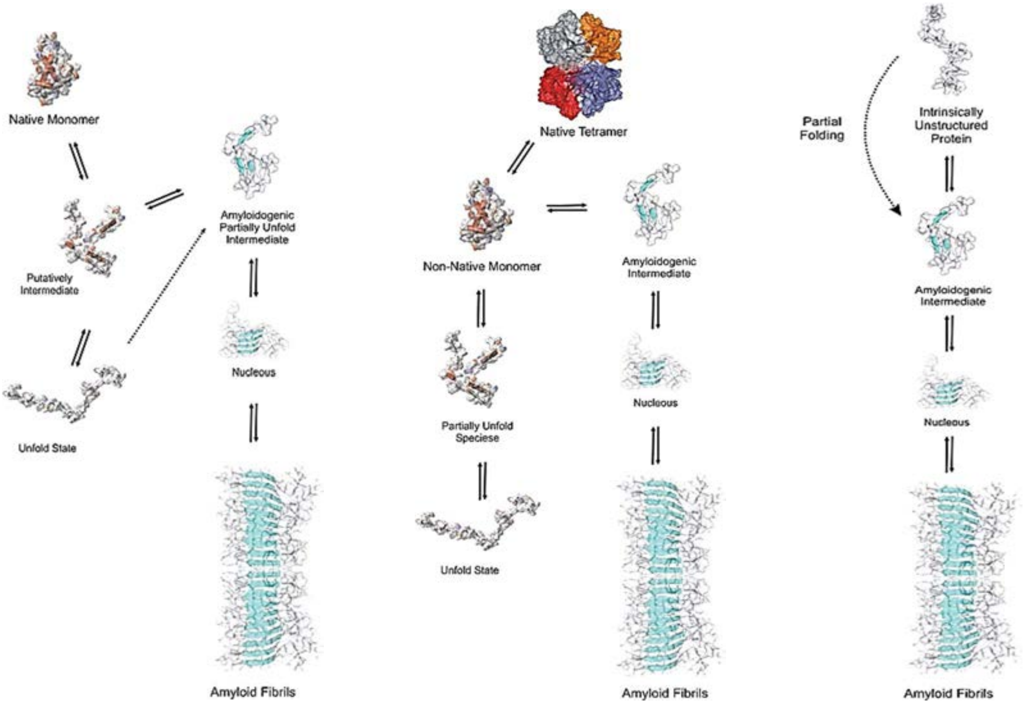


Figure 1.6. A schematic diagram showing different pathways of amyloid fibril formation. *Left panel:* Native monomeric precursor, due to the lack of conformational stability, populates partially unfolded states, which can easily convert into amyloidogenic intermediates and further lead to the amyloid fibril formation. *Middle panel:* Native homo-tetrameric protein precursor dissociates into a non-native monomer, which can undergo conformational changes and expose hydrophobic patches. This event can eventually lead to the formation of amyloid fibrils. *Right panel:* IDP precursor partially folds into β-sheet intermediate with high aggregation propensity. Adapted from: (Quintas, 2013).

Self-seeding (seeding with a protein exposing the same APR) is much more efficient than cross-seeding (seeding with a protein exposing another APR). This observation is confirming the specificity of the assembly process. Importantly, the reversed or scrambled APRs of a given protein cannot cross-seed with the original APRs, indicating that together with the sequence composition, the specific position of the amino acids is an important determinant as well (Sabate *et al.*, 2010).

After the formation of a sufficiently large, stable oligomer (nucleus), the monomers can be incorporated rapidly during the process of fibril elongation which is a thermodynamically favourable phase. Then, fibrillation reaches a stage called the saturation phase, when the fibrils stop growth due to the exhaustion of monomers. However, the small amount of monomers that remain present are in a continuous assembly/disassembly equilibrium (Mankar *et al.*, 2011).

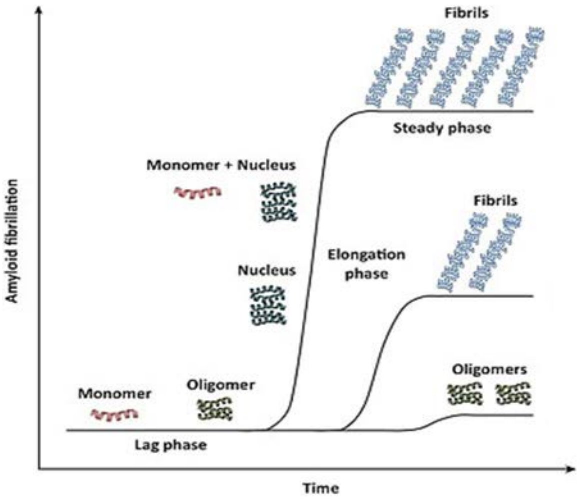


Figure 1.7. A schematic diagram showing the nucleation-dependent polymerization model of amyloid fibril formation. Aggregation processes resulting in mature amyloid fibrils produce a sigmoidal kinetic curve that is depicted in the figure. In the initial stages, monomeric forms of amyloid-forming peptides self-assemble into soluble oligomers. During the elongation phase, the formed nucleus is being rapidly extended into fibrils. Adapted from: (Hajipour *et al.*, 2017).

1.3. Protein folding and aggregation in the cell

The cell is a highly crowded environment where 20-30% of the available volume is occupied by various macromolecules. Consequently, due to the large reduction in the volume accessible to molecules, the effective protein concentration increases 10-100 fold (van den Berg *et al.*, 1999). That means that when compared to *in vitro* conditions, the protein misfolding and aggregation risk is higher *in vivo*.

Protein synthesis and folding are coupled processes in the cell. The truncated translation products, misfolded intermediates and incomplete assembled oligomeric proteins can potentially have their hydrophobic stretches exposed at the surface, which can lead to the establishment of non-functional intermolecular interactions (Dobson, 2003; Hartl and Hayer-Hartl, 2009). Cells have evolved a molecular machinery assigned to preserve protein quality by aiding proteins in attaining their native state, preventing them from establishing inappropriate intermolecular interactions, or degrading them in case the damage is too extensive or conditions for protein repair become unfavourable.

1.3.1. Molecular chaperones and degradation machinery

During and after its synthesis at the ribosome, the protein folds through various intermediates to its native, three-dimensional structure. Cellular stresses, mutations in the synthesized protein or translational errors can cause protein misfolding. Chaperones are folding assistants that minimize non-native interactions by shielding hydrophobic groups until they are buried inside the native structure (Walter and Buchner, 2002; Gregersen, 2006). They are usually classified based on their molecular weight (in kDa), and proteins belonging to the same family have a high sequence homology and are structurally and functionally related (Walter and Buchner, 2002). For instance, in prokaryotes, the trigger factor (TF) binds to the nascent polypeptide chains. After ATP dependent release of TF from the chain, either the folding to its native state can be initiated or the polypeptide chain can be transferred to the chaperones acting downstream of the ribosome. The ribosome-binding chaperones in other organisms, such as ribosome-associated complex (RAC) in yeast, or nascent polypeptide-associated complex (NAC) in archaea and eukaryotes, bind to the newly synthesized chains and shield the exposed hydrophobic stretches.

Proteins that cannot spontaneously fold into the native conformation after being released from the ribosome-binding chaperones are transferred to Hsp70 chaperone system (Kim *et al.*, 2013). Other than assisting the folding of nascent polypeptide chains, they participate in the translocation of the proteins, refolding of misfolded and/or aggregated proteins, and introduction of conformational changes (Rodriguez *et al.*, 2008). Studies have shown that Hsp70 chaperones preferentially bind to short hydrophobic stretches of extended polypeptides which are flanked by charged residues (Rudiger *et al.*, 2000). These features are consistent with the characteristics of an APR flanked by gatekeeper residues. Hsp70 chaperones cooperate with members of the Hsp40 family, which mediate substrate binding to Hsp70 but are inefficient themselves in preventing aggregation (Kampinga and Craig, 2010).

Moreover, Hsp70 members also cooperate with the Hsp100 family to achieve protein disaggregation. The exact mechanism is still elusive, however, a recent model postulates that Hsp70 chaperones disentangle polypeptides from already formed aggregates and deliver them to the Hsp100 unfoldase. Refolding of the polypeptide chains emerging from Hsp100 involves Hsp70 and Hsp60 chaperones. The Hsp60 class, known as chaperonins, form large complexes with double ring structure, often described as “Anfinsen cage”, where they encapsulate substrates in order to provide a protective folding environment where the isolated polypeptide chain can fold far from the crowded cellular environment, and far from the risk of establishing aberrant intermolecular interactions (Ellis, 1993; Horwich *et al.*, 2007). The Hsp90 family also cooperates with Hsp70 in the folding and maturation of substrates, such as eukaryote signalling proteins (Taipale *et al.*, 2010).

Another process assisted by molecular chaperones is the protein degradation. Despite the preference for protein repair, damaged proteins are targeted for protein degradation when there is no other alternative (Buchberger *et al.*, 2010). Even though the degraded proteins can be subsequently used for *de novo* synthesis, from the thermodynamic point of view, the biogenesis of proteins is unfavourable. Proteolytic systems involved in the intracellular protein turnover are ubiquitin-proteasome system (UPS) and lysosomal system. Lysosomes deal with the aggregates through the process called macroautophagy thanks to the hydrolases that they contain. Indeed, it has been shown that up to 30% of newly synthesized proteins are being degraded by the proteasome due to the ineffectiveness of the folding reaction (Schubert *et al.*, 2000). It has been proposed that E3 ubiquitin ligases might be able to recognize misfolded proteins and tag them by attaching several ubiquitin moieties to lysine residues which serves as a signal for degradation by the UPS (Goldberg, 2003; Ciechanover, 2005).

1.3.2. Sequestration of protein aggregates

An increasing amount of evidences suggest that soluble oligomers themselves, rather than the insoluble amyloids, as the final stage of protein aggregation, are cytotoxic species (Chiti and Dobson, 2006). Therefore, the formation of highly ordered amyloid aggregates may play a cytoprotective function, at least for certain proteins (Kitada *et al.*, 1998; Arrsate *et al.*, 2004; Tanaka *et al.*, 2004; Douglas *et al.*, 2008). Furthermore, after protein aggregation has already occurred, directing aggregated proteins to specific compartments can protect the cellular environment from these potentially deleterious species. The spatial sequestration of aggregated proteins can be viewed as a second cellular response that is initiated when the cellular quality-control machinery that assists refolding or degradation of misfolded proteins has been overrun. Depending on the particular aggregation-prone protein, organism, and the stress conditions that caused protein misfolding, there are different pathways for the cellular sequestration of protein aggregates (Fig. 1.8).

In bacteria, the aggregates of proteins can form under various conditions, such as exposure to increased temperature, oxidative stress conditions (Laskowska *et al.*, 2004), or due to the overexpression of heterologous proteins. These aggregates are typically localised to the old cell pole but the mechanisms by which the aggregating protein reaches the deposition sites are still unclear. There have been contradictory reports, one suggesting that the active, energy-dependent process is responsible for polar localization (Rokney *et al.*, 2009), and the second demonstrating that nucleoid occlusion itself is necessary and sufficient for controlling the polar localization of bacterial aggregated proteins.

The pattern of aggregation depends on the substrate and there may be various ways of sequestering different proteins, as it has been described in yeast. Misfolded species can be sequestered into one of two identified specialized quality-control compartments for the deposition of aggregated proteins (Kaganovich *et al.*, 2008). One compartment is a structure adjacent to the nuclear membrane, termed the juxtanuclear quality-control compartment (JUNQ). JUNQ is regarded as a temporary deposition site where misfolded ubiquitylated proteins are being accumulated and further targeted for refolding or degradation.

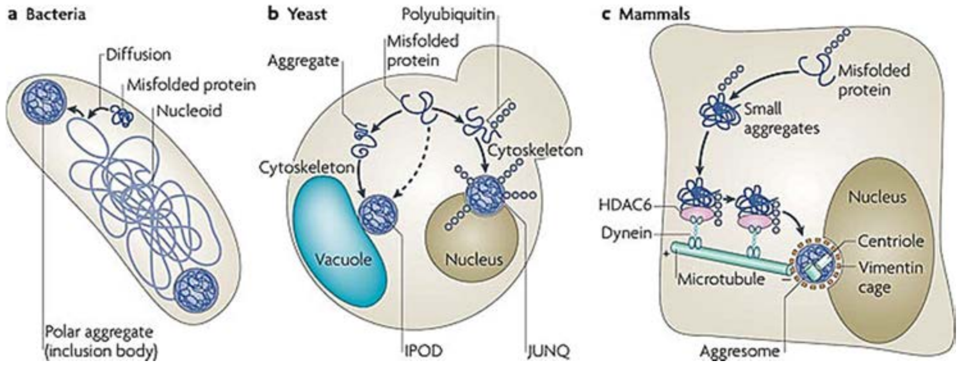


Figure 1.8. A schematic representation of different pathways for the sequestration of protein aggregates. (A) In bacteria, misfolded proteins can accumulate in inclusion bodies located at the periphery of the cell. Nucleoid exclusion is sufficient to control the localization of aggregated proteins at the specific cellular site. (B) In yeast cells, soluble, misfolded, ubiquitylated proteins can be disposed at the juxtanuclear quality-control compartment (JUNQ), while insoluble, terminally aggregated proteins can accumulate at the insoluble protein deposit (IPOD). (C) In mammalian cells, protein aggregates are being sequestered into aggresome, a vimentin-enwrapped structure located at an indentation of the nucleus. Aggresome formation requires the adaptor histone deacetylase 6 (HDAC6), which binds to ubiquitylated substrates and the microtubule minus-end motor protein dynein. Adapted from: (Tyedmers *et al.*, 2010).

The second compartment adjacent to the vacuole, termed the insoluble protein deposit (IPOD), harbours terminally aggregated, insoluble proteins that cannot be further refolded of degraded by the UPS. However, substrates that are targeted for either compartment can be directed to the other compartment by experimentally manipulating the ubiquitin proteasome system or impairing ubiquitylation of misfolded substrates. Aggresomes are specialized form of inclusion bodies that form in the mammalian cells, and are usually located at the proximity of the nuclear envelope, at the microtubule organizing centre (MToC). The aggresome is encapsulated by a cage-like shell formed by the intermediate filament vimentin. The initial steps of aggresome formation include the formation of smaller aggregates in the periphery, which are transferred along the microtubule cytoskeleton to the final perinuclear site at the MToC. The transport of aggregated proteins to aggresomes is mediated by histone deacetylase 6 (HDAC6), which binds polyubiquitin chains of terminally misfolded proteins and the microtubule motor protein dynein, and mediates the transport of substrate along microtubules towards the MToC during aggresome formation (Kawaguchi *et al.*, 2003). The formation of the aggresome likely functions as a cytoprotective mechanism rather than a pathogenic one (Kopito, 2000), isolating the terminally misfolded proteins and preventing their interactions with the other proteins or the components of the PQC.

1.4. Bioinformatic approaches to predict aggregation propensity

The vast knowledge about the intrinsic determinants that influence the deposition of proteins has permitted the development of *in silico* computational methods that aim to predict the propensity of proteins to form insoluble deposits. These methods are especially useful when it comes to analysing large datasets that can provide an insight into cellular evolutionary strategies that have been developed against protein aggregation or designing the therapeutics that could modulate the aggregation-propensity of disease-linked proteins.

There are two main approaches that can be identified among the more than 20 computational tools that have been published to date, the empirical and the structure-based strategies. Empirical methods rely on estimating aggregation propensity by taking into account, in an explicit or implicit manner, the physico-chemical properties of individual amino acids that compose the polypeptide chain. On the other hand, structure-based methods evaluate the compatibility between protein structural features and known cross-β structures or the propensity to form β-structures.

1.4.1. Sequence-based algorithms

One of the first methods that allowed the evaluation of the tendency of a protein to aggregate using a protein sequence as an input was TANGO (Fernandez-Escamilla *et al.*, 2004). This method analyses the tendency to adopt β-sheet structure taking into account the extrinsic determinants of aggregation such as pH, temperature and ionic strength, or factors like trifluoroethanol (TFE) concentration. TANGO predicts a segment of the sequence that has a tendency to aggregate with a minimum of 5 residues. TANGO was thus the first method to allow the detection of APRs within a protein sequence.

The empirical method AGGRESCAN, for instance, is based on the *in vivo* gathered information obtained by analysing the deposition properties of a generated set of Aβ point mutants fused to GFP in prokaryotic *E.coli*. AGGRESCAN exploits this experimentally obtained scale of aggregation propensity of 20 natural amino acids to predict aggregation propensity using a primary sequence of a polypeptide as an input (de Groot *et al*., 2006). AGGRESCAN predicts aggregation propensity from the sequence by computing for each amino acid the average aggregation propensity of a variable window, depending on the size of a protein, centred at this position. The predictor allows determining the overall aggregation propensity, the presence of hot spots of aggregation or the effect of point mutations on the deposition properties (Conchillo-Sole *et al*., 2007).

The Simple Algorithm for Sliding Averages (SALSA) identifies APRs as regions of polypeptide sequence with strong β-strand propensity. To each residue, it assigns the mean β-sheet propensity of different averaging windows centred on it, according to the secondary structure propensity defined by Chou and Fasman (Zibaee *et al*., 2007).

ZYGGREGATOR predicts aggregation propensity of polypeptides taking into account physico-chemical properties of amino acids and implements a new parameter in order to account for the impact of gatekeeper residues against aggregation. In contrast to previous developments, ZYGGREGATOR takes into account the influence of local structural flexibility on the basis of the prediction of flexibility and solvent accessibility of the polypeptide chain. This is a useful approach since the protein in its native state has some APRs buried inside the hydrophobic core, unable to participate in the intermolecular interactions driving protein aggregation (Tartaglia *et al*., 2008).

1.4.2. Structure-based algorithms

As mentioned above, the structure-based methods rely on characteristic structural features that have been associated with the formation of ordered amyloid-like aggregates. The first method derived from structural information is Net-CSSP algorithm (Yoon and Welsh, 2004) which is based on the observation that the regular secondary structure adopted by a polypeptide chain is dependent on its tertiary contacts (Minor and Kim, 1996). This indicates that the theoretical predictions differ from the experimental determinations due to the observation that certain sequences without β-conformation, in the context of their natural structural environment, may indeed exhibit a propensity to form β-sheets. The mentioned algorithm was developed in order to identify these stretches in the protein sequences that may have hidden β-sheet tendency and therefore, higher risk to aggregate.

Another method, FoldAmyloid, predicts the amyloid fibril-forming regions based on the mean packing density of individual residues. A mean packing density scale was determined for the

20 natural amino acids, on the basis of the average number of contacts observed for each residue, according to the information in the SCOP structural database (Garbuzynskiy *et al.*, 2010). According to the algorithm, the protein segments identified to contain more than 5 consecutive residues exhibiting a strong packing density are considered amyloidogenic.

Prediction of Amyloid Structure Aggregation (PASTA) is another method based on the assumption that β-strands constituting the amyloid fibril have a preference for an in-register parallel or anti-parallel arrangement with a minimum energy. Creating a dataset with these strictly defined secondary structure alignment allowed the calculation of a pairing energy for each possible pair of residues, which is then used to score all possible stretches in a sequence. Those pairings of stretches with a bonding energy below a certain threshold are considered to present an increased likelihood to form the cross-β core of amyloid-like structures (Trovato *et al.*, 2007).

BETASCAN estimates the most probable β-strand pairing between pairs of segments of the polypeptide chain with the same length relying on the table of probability for a couple of residues to be hydrogen-bonded in a set of amphiphilic β-strands selected from the Protein Data Bank (PDB) (Bryan *et al.*, 2009).

Waltz offers a way to distinguish between amorphous aggregates and amyloid fibrils combining empirical information and structural studies of a large set of amyloidogenic hexapeptides. Waltz is a prediction algorithm intended to identify amyloidogenic stretches in the polypeptide sequences by exploiting a sequence based position-specific scoring matrix (PSSM), based on the dataset including both positive and negative peptides for fibril formation, and combined with a set of physico-chemical properties and a structure based PSSM (Maurer-Stroh *et al.*, 2010). An update of the Waltz scoring function (pWaltz) was used to identify amyloid cores in putative prion domains through the PrionW server, that allows the recognition of prion-like domains (PrLDs) scanning for the presence of amyloid cores in sequences rich in glutamine/asparagine (Q/N) (Sabate *et al.*, 2015; Zambrano *et al.*, 2015a).

AGGRESCAN3D (A3D) is an evolved version of AGGRESCAN, where a structure-based approach is implemented to resolve the limitations of sequence-based algorithms. It offers the detection of APRs in globular proteins, taking into consideration the dynamic fluctuations and spatial clustering of residues distant in the primary sequence of a protein. The predictor allows determination of the effect of mutations on the overall aggregation propensity and an accurate discrimination of the effect of pathogenic mutations (Zambrano *et al.*, 2015b).

Finally, the consensus algorithm AMYLPRED, in the beginning fusing 5 of the described methods (Frousios *et al.*, 2009), but lather on updated to combine 11 different methods (Tsolis *et al.*, 2013), generates integrated predictions at the residue level that allows simultaneous comparison between the various outputs. The combination of the outputs provided by different algorithms may increase the accuracy of the prediction and minimize methods-specific bias towards the overprediction of certain types of aggregation.

1.4.3. Evolutionary strategies against protein aggregation

The presence of at least one APR in most proteins belonging to different proteomes (Reumers *et al*., 2009) implies that the ability to form cross-β enriched structures represents a generic property of polypeptides. However, aggregation as a competing process of protein folding is driven by the establishment of intermolecular contacts. Therefore, burying APRs inside the hydrophobic core of stable globular protein acts as a very efficient mechanism to avoid the establishment of the aberrant intermolecular contacts that drive the process of protein aggregation. The analyses of different sets of globular proteins with experimentally solved native states revealed that indeed aggregation-prone stretches tend to be buried (Linding *et al.,* 2004; Buck *et al*., 2013; Watters *et al*., 2007), mostly as a part of hydrophobic core, which means that under physiological conditions, these stretches cannot longer participate in non-native intermolecular contacts. Thus, the ability of proteins to efficiently attain a compact globular form through burial of hydrophobic side chains is not only the main driving force for protein folding and a strong determinant of protein stability but also a supplementary evolutionary strategy aimed to minimize the probability of protein aggregation (Monsellier and Chiti, 2007). On the other hand, the analyses of sequences belonging to structurally flexible proteins and IDPs revealed that these groups of proteins are depleted in APRs, while an enrichment is observed in charged amino acids and β-breaker residues such as Pro (Tompa, 2002), thus counteracting aggregation.

Large-scale analyses of different proteomes have allowed the identification of those determinants upon which evolution has imposed restrictions to limit protein aggregation. In spite of the ubiquitous presence of APRs in a large fraction of proteins in various proteomes, the predicted aggregation propensity is not homogenous across species and decreases with the increasing complexity and longevity of the corresponding organism (Tartaglia *et al.*, 2005; Rousseau *et al*., 2006). The frequency of detected APRs was found to be dependent on some other factors as well, such as the polypeptide length, concentration, half-life, and the complexity of the structural organisation.

Longer proteins undergo slower and more complicated folding processes which suggests that these proteins are expected to populate partially unfolded states for a larger time and therefore, that hydrophobic APRs may become exposed and potentially lead to protein aggregation (Ivankov *et al*., 2003, Monsellier *et al*., 2008), if they are not shielded by molecular chaperones.

Protein aggregation is a high-order reaction, and, at the same time, the efficiency of the aggregation reaction is influenced by the degree of sequential identity between interacting molecules (Krebs *et al*., 2004; Wright *et al*., 2005). For this reason, protein abundance represents an important and regulated determinant due to the increasing aggregation risk imposed by high local concentrations of identical polypeptide chains. This is especially true for proteins that perform their biological role as homo-oligomers and proteins that are found to be essential for cell survival. Subsequently, proteins exhibiting mentioned properties generally exhibit a lower predicted aggregation propensity and lower frequency of detected APRs relative to other polypeptides, suggesting that these groups of proteins experience selective pressure that aims to minimize their tendency to aggregate.

In a similar manner to the folding reaction, the formation of stable quaternary structures and protein complexes, aside from its functional implications, could have potentially evolved as a protective strategy to avoid the establishment of non-functional intermolecular contacts. As an example, proteins such as transthyretin (TTR), superoxide dismutase 1 (SOD1) and even α-synuclein have been reported to possess stable native oligomeric quaternary structures. The undesired dissociation of quaternary structures of these proteins into their monomeric subunits leads to pathogenesis due to exposure of aggregation-prone regions that have been physically shielded and hidden on the interfaces of quaternary structures, which then become involved in self-assembly processes.

Other strategies have been identified to have apparently been incorporated in cellular proteomes in order to confront the risk of aggregation. The relevance of protective shielding strategy is further supported by the presence of disulfide bonds and attractive electrostatic interactions in the proximity of interfaces, enhancing the stability and specificity of interactions (Pechmann *et al.*, 2009). It has been observed that protein structures with disulfide bonds present a high predicted aggregation propensity compared to proteins devoid of these covalent links which is especially true for extracellular proteins (Mossuto *et al.*, 2011; Graña-Montes *et al.*, 2012). This suggests that the presence of disulfide bonds increases the tolerance for a greater aggregative potential, particularly in proteins that function in harsh environments where the maintenance of their native states might be compromised. Furthermore, the disulfide cross-linking has a stabilising effect on native conformations (Grantcharova and Baker, 2001), but it also constraints the kinetics of aggregation and leads to a reduced cytotoxicity of the resulting aggregates (Mossuto *et al.*, 2011; Graña-Montes *et al.*, 2012).

1.5. Protein models: intrinsically disordered proteins

A commonly accepted structure-to-function paradigm which indicates that proteins need to attain a native three-dimensional structure to be able to perform their intended biological role has been silenced after a large number of proteins that lack a defined equilibrium conformation in its native state have been identified (Wright and Dyson, 1999; Tompa, 2012). These intrinsically disordered peptides/proteins (IDPs) cover a spectrum of states from fully unstructured to partially structured and include random coils, (pre-)molten globules, and large multi-domain proteins connected by flexible linkers. A first extensive study of IDPs revealed that these proteins possess a high net charge and low hydrophobicity (Uversky *et al.*, 2000). They are usually enriched in polar and charged residues (Lys, Arg, Ser, Gln, Glu) and Pro, but depleted in aromatic (Trp, Phe, Tyr) and aliphatic amino acids (Ile, Leu, Val), as well as Asn and Cys, compared to structured globular proteins (Radivojac *et al.*, 2007).

The analysis of proteomes from all three domains of life, using various computational methods to predict structural disorder from a sequence (Ferron *et al.*, 2006; Jin and Liu, 2013), have permitted the estimation of the IDP content in these proteomes. It has been revealed that in prokaryotes 10-35% proteins possess large disordered regions and could be classified as IDPs, while in eukaryotes 15-45% of proteins have been estimated to possess these properties (Dunker *et al.*, 2008; Tompa, 2012). IDPs are involved in important cellular activities, such as regulation and signalling (Tompa, 2012). IDPs have been implicated in key cellular processes such as transcription, translation, differentiation, and different stages of cell cycle. It is the conformational plasticity of IDPs that allows these proteins to perform various molecular functions. IDPs can display multiple interaction sites and in such a way that they interact with a high variety of targets simultaneously or sequentially (Uversky 2002; Wright and Dyson, 1999).

Even though the origin of protein aggregation and amyloid formation is not well understood for IDPs, since they lack a fixed three-dimensional structure in physiological conditions, there have been great efforts made in order to enhance this knowledge. Intensively studied amyloidogenic IDPs include the amyloid-β (Aβ) peptide, and tau protein, associated with AD, the human islet amyloid polypeptide, associated with a type 2 diabetes mellitus (T2DB), and α-synuclein associated with PD.

1.5.1. Amyloid-β peptide

Alzheimer's disease (AD) is a conformational disease characterised by two main hallmarks: amyloid plaques consisting of aggregated amyloid β-peptides found in the extracellular space, and neurofibrillary tangles formed by filaments of hyperphosphorylated tau protein found intracellularly (Ballard *et al.*, 2011). Symptomatically, AD patients first exhibit inability to retain recent memories which later evolves to a full cognitive and functional impairment (Jakob-Roetne and Jacobsen, 2009). Amyloid-β (Aβ) peptides are generated when amyloid precursor protein (APP) is proteolytically digested. APP is a growth factor that is involved in proliferation and differentiation of neuronal stem cells. APP can be cleaved by two different pathways: the non-amyloidogenic pathway that involves α- and γ-secretases, in which amyloid precursor protein intracellular cytoplasmic/C-terminal domain (AICD) peptide is generated, and amyloidogenic pathway in which Aβ peptides, ranging from 38 to 43 amino acids, are generated when APP is cleaved by γ-secretases (Fig. 1.9).

Alzheimer's amyloid plaques consist mostly of peptides 40 and 42 residues long (Aβ40 and Aβ42) peptides. Aβ42 is considered more toxic than Aβ40, although the latter one is more abundant in neurons (Iwatsubo *et al.*, 1994). Mutagenesis experiments of Aβ42 have allowed the identification of the peptide region called central hydrophobic cluster composed of five residues that acts as a hot spot of aggregation. During the amyloidogenic reaction, Aβ peptides self-assemble into intermediate oligomeric species that evolve into protofibrils that subsequently self-assemble into amyloid fibrils.

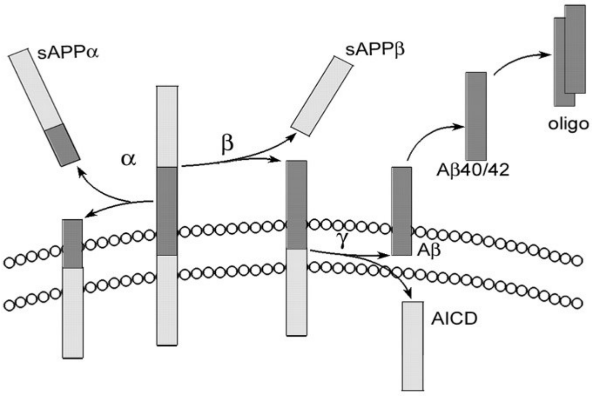


Figure 1.9. A schematic representation of different pathways in which APP gets cleaved: the non-amyloidogenic pathway that involves α- and γ-secretases, in which AICD peptide is generated, and amyloidogenic pathway in which Aβ peptides, ranging from 38 to 43 amino acids, are generated when APP is cleaved by γ-secretases. Adapted from: (Nixon, 2007).

1.5.2. α-synuclein

α-synuclein (α-syn) is a 140 amino acid protein that is the major component of Lewy's bodies which are considered to be the major pathological hallmark of Parkinson's disease (PD). Clinical signs of PD, such as rigidity, tremor and bradykinesia, are caused by the neuronal loss in the substantia nigra (Braak *et al.*, 2003; Tolosa *et al.*, 2006). The function of α-syn has not yet been well established but it's thought to be involved in the regulation of dopamine neurotransmission and stabilization of lipid membranes.

The primary structure of α-syn features an N-terminal domain (1-60), the hydrophobic fibrillation-triggering NAC region (61-95), and a C-terminal acidic region (96-140) (Fig. 1.10). Rare point mutations of α-syn, such as A53T, A30P, H50Q, A53E, E46K and G51D that result in autosomal dominant familial PD, all occur in the N-terminal region, suggesting an important function for this region of the protein. N-terminal domain encodes for a series of imperfect 11 amino acid repeats with consensus motif of KTKEGVT which shares similarities with the lipid-binding domain of apolipoproteins (Wales *et al*., 2013; Breydo *et al.*, 2012). The central hydrophobic region (non-amyloid-β component or NAC region) is 12 amino acid-stretch that has been described as the most important region for the aggregation properties of synuclein (Jethva *et al.*, 2011). The C-terminal tail contains mostly negatively charged residue and is largely unfolded. The C-terminal was found to interact with both the NAC domain and the N-terminal of α-syn stabilising its monomeric forms acting as an intramolecular chaperone (Bertoncini *et al.*, 2005). In agreement with this observation, α-syn truncated forms missing C-terminal region aggregate at the higher rate than the full-length protein *in vitro* (Crowther *et al.*, 1998).

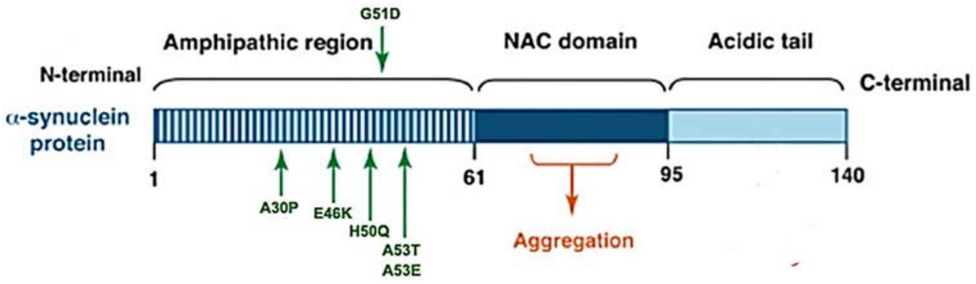


Figure 1.10. A scheme representing α-synuclein (α-syn) protein domain structure. The sequence of α-syn can be divided into three regions with distinct structural characteristics: amphipathic, highly conserved N-terminal domain, the central hydrophobic region (non-amyloid-β component or NAC domain) and largely unfolded, acidic C-terminal. Adapted from: (Xu and Chan, 2015).

Due to its conformational plasticity, α-syn structure and fibrillation rate are largely modulated by the protein environment (Uversky and Eliezer, 2009; Silva *et al.*, 2013). An increase in the ionic strength induces α-syn partial folding because it minimizes the electrostatic repulsion between the residues in the C-terminal domain of α-syn (Hoyer *et al.*, 2002), which can consequently promote aggregation. Furthermore, solution conditions that are close to the α-syn isoelectric point (Uversky and Eliezer, 2009), together with any condition which strengthens the hydrophobic effect, such as increased temperature (Baldwin, 1986), also increase the rate of α-syn aggregation. The equilibrium shift toward the partially folded intermediates can be modulated by other factors, such as presence of metal ions, herbicides, pesticides, the interaction with some proteins (Yoshimura *et al.*, 2017), the introduction of mutations or post-translational modifications (PTMs). Depending on the factor, the resulting generated α-syn species can exhibit various structural and biological properties (Fig. 1.11).

As expected, during α-syn aggregation kinetics, while progressing from oligomers to protofibrils, and further to fibrils, an increase in β-sheet content has been observed (Apetri *et al.*, 2006). Despite protofibrils share structural properties with mature fibrils, such as linear morphology and the ability to interact with specific antibodies, there have been some intraresidue contacts identified in protofibrils that are not present in mature fibrils, suggesting that the increase in the number of residues involved in β-strands, involves the remodelling of the pre-existent β-sheet structure (Petty *et a*l., 2005). Furthermore, it has been suggested that α-syn protofibrils might indeed grow by a mechanism of oligomer addition and coalescence (Modler *et al.*, 2003), where a conformational conversion of oligomers may occur either before or simultaneously with the oligomer association to the extremes of the protofibrils.

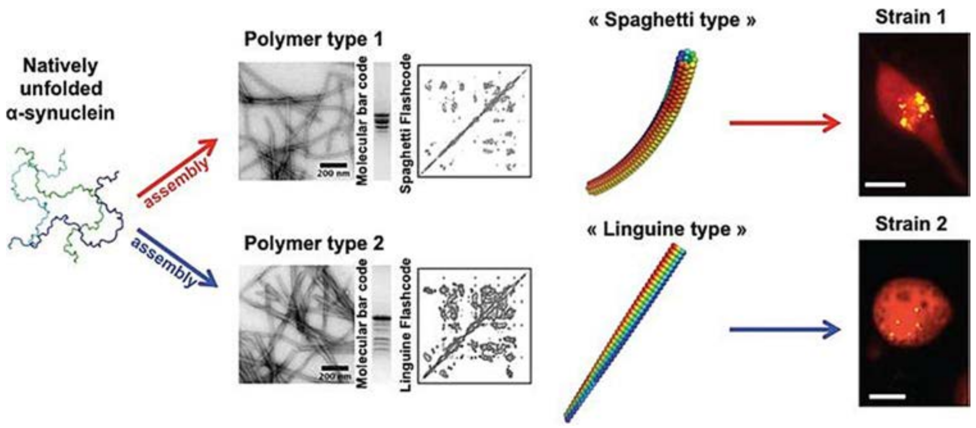


Figure 1.11. α-syn strains. Starting from the natively unfolded α-syn, the self-assembly occurs resulting in two distinct types of polymers exhibiting different morphologies, as observed by electron microscope. Furthermore, the two assemblies exhibit distinct limited proteolysis digestion patterns, referred as molecular barcodes, and solid-state NMR spectra, referred as a structural fingerprint. The two α-syn strains have different functional properties when injected into neuronal cultures, such as toxicity, binding propensity to cells and seeding capacities. Adapted from: (Melki, 2015).

α-syn exists at equilibrium as a conformational ensemble of heterogeneous conformational isomers that exhibit long-range intramolecular interactions. These diverse conformational isomers exhibit differential properties for aggregation/fibrillization. This structural promiscuity, and the fact that the diseases linked to α-syn aggregation cannot be considered homogeneous, since there are different clinical phenotypes identified within the same kind of synucleinopathy (McCann *et al*., 2014), a number of studies have emerged suggesting the existence of strains in α-syn pathology. The existence of α-syn strains, that have distinct physical and biological properties, would support the concept that synucleinopathies share important features with archetypic prion diseases.

These different conformational states that α-syn can adopt in its aggregated state exhibit differences in fibril morphology, X-ray diffraction patterns, the resistance to proteolytic digestion and seeding capacity. Initially, these different strains were generated *in vitro* through the modulation of solution conditions, temperature, or repetitive seeded fibrillization. It has been shown that PTMs of α-syn (i. e. α-syn phosphorylated at serine 129) can induce different strains formation (Ma *et al*., 2016). The additional evidence for the existence of putative *α*-syn strains came from the lesion profiles observed after newly formed *α*-syn fibrils assembled directly from recombinant protein, or after repeated seeding, were injected to primary neurons and transgenic mice (Guo *et al*., 2013). Two distinct strains of synthetic α-syn fibrils, exhibiting differential efficiency of cross-seeding tau aggregation, both in neuron cultures and *in vivo*, have been discovered. Furthermore, the proteinase K degradation profiles of the insoluble protein fractions isolated from patients diagnosed with PD, or PD and AD, revealed conformational differences between α-syn deposits. Interestingly, the existence of Aβ strains has been demonstrated as well, after amyloid plaques that were extracted from the brain of patients developing AD, with distinct clinical history, were used to seed the aggregation of Aβ *in vitro* (Guo *et al*., 2013).

2. Objectives

The main objective of this thesis is to study the determinants of protein aggregation, implementing a multidisciplinary approach to decipher the fundamental intrinsic and extrinsic factors that lead to the aggregation of proteins in the so-called conformational disorders. In the last years, different methods have been applied to the study of protein aggregation, ranging from *in vivo* and *in vitro* studies to *in silico* approaches. The main aims of this thesis can be summarized in the following points:

- Providing insights into the connection between the structural aggregation propensity of folded proteins and the evolutionary strategies adopted by nature to confront the aggregation in biological environments, using *E. coli* as a model system.
- Identifying the conformational species of the Alzheimer's linked peptide, amyloid β-peptide (Aβ), responsible for the cellular oxidative damage, using *S. cerevisiae* as a model organism.
- Understanding the structural determinants that modulate the formation of conformationally different amyloid structures and their associated toxicity in the case of the Parkinson's disease-associated protein α-synuclein (α-syn).

3. A structure-based protein aggregation profile of the Escherichia coli proteome

3.1. Introduction

Proteins are central components of almost all cellular processes and are involved in a variety of complex protein-protein interaction networks in the living cells. These interactions can be functional as well as non-functional, or promiscuous. Levy *et al.* have defined the propensity of proteins to interact as protein “stickiness” (Levy and Teichmann, 2012). The overall “stickiness” of a protein is influenced by a frequency of appearance of “sticky” amino acids in the sequence or at the solvent-accessible surface of the protein. Also, it has been reported that the protein abundance plays an important role in the evolution of proteomes, and therefore, on chemical and physical properties that proteins have, in this case the protein “stickiness” (Levy and Teichmann, 2012). The native fold of a protein is responsible for correct assembly into functional protein complexes. The different physiological conditions present in the cellular context can reshape the native fold of a protein and result in the exposure of the aggregation-prone regions (APRs) on the protein surface. These APRs are commonly part of the hydrophobic core and in such a way protein is protected from aggregation. However, the protein misfolding, which leads to the exposure of APRs on the protein surface, can infringe the thermodynamic stability of a protein, increase the overall protein “stickiness” and enhance the probability of non-functional or promiscuous interactions with the rest of the proteome. Therefore, protein misfolding is usually associated with the deterioration of essential cellular processes. In many cases, the formation of misfolded polypeptides results in their assembly into aggregates or β-sheet enriched structures known as amyloid fibrils, which can have a deleterious effect on cell fitness. However, it is important to mention that not all proteins are folded or adopt a globular conformation under natural physiological conditions to perform their biological function. These intrinsically disordered proteins (IDPs), which have essential regulatory roles such as controlling cell cycle, involvement in signalling activities, DNA and RNA binding, are structurally characterised by a high structural flexibility.

Previously, different computational algorithms have been implemented in order to perform large-scale aggregation analysis of protein sets of different organisms (Castillo *et al*., 2012; Belli *et al*., 2011). Most of these algorithms are linear aggregation predictors that require only the protein sequence as the input. Also, some of these web servers allow the identification of amino acid mutations that could reduce protein aggregation by reducing the intrinsic aggregation propensity of the sequence and can be applied in protein engineering (Van Durme *et al*., 2016).

The current study emphasizes the importance of conformational properties when predicting the aggregation propensity of globular proteins. Linear aggregation predictors assume that the identified APRs are exposed to solvent, while neglecting the possibility that these regions might be hidden inside the hydrophobic core or involved in the series of non-covalent interactions that sustain the tertiary and quaternary structure of proteins. For this reason, we performed a surface aggregation propensity analysis of *E.coli* proteins with available 3D-structures using the AGGRESCAN3D (A3D) algorithm (Zambrano *et al.,* 2015b). Overall, we confirm that the avoidance of protein aggregation seems to act as a strong and generic evolutionary constraint acting on protein folded states in bacteria, although the selective pressure against aggregation is stronger in certain groups of protein structures, depending on protein essentiality, biological function, protein active form and protein size.

3.2. Results and discussion

3.2.1. Relationship between protein abundance and structural aggregation propensity

Non-functional interactions usually compromise fitness, either because they sequester interaction partners (Vavouri *et al*., 2009) or because they result in the formation of toxic aggregates. According to the law of mass action, the probability to establish a non-functional interaction should be proportional to the protein abundance (Levy and Teichmann, 2012). In addition, most protein aggregation processes follow a nucleation-polymerization scheme, in which the formation of the initial aggregation nuclei represents the rate-limiting step of the overall process. Nucleation processes correspond to the second-or higher-order reactions and therefore the rate of protein aggregation is also strongly dependent on the initial protein concentration. Therefore, an abundant protein with a sticky aggregation-prone surface is expected to be more deleterious than a low-abundance protein with the same surface stickiness.

Thus, if cellular crowding facilitates promiscuous interactions and aggregation/co-aggregation, we would expect a negative correlation between protein structural aggregation propensity (SAP) and protein abundance. Thus, we decided to explore the relationship between the abundance and the aggregation properties of bacterial proteome as determined by A3D. We obtained the abundance data for the 612 proteins in the complete dataset and we proceeded by $\log_{10}$ transforming the reported abundance for statistical analysis, since the number of mRNAs in the bacterial cytosol encoding a given protein can vary greatly from 1 to 100,000 (Selinger *et al*., 2000). The proteins from the dataset were divided into 10 groups according to their protein abundance. The average A3D scores were calculated for each group and the two parameters were compared (Fig. 3.1A). A highly significant correlation ($r = 0.984$; $p \leq 0.0001$) was observed, indicating the link between the concentrations of bacterial proteins in the cell and their overall aggregation tendencies when folded. This data confirms previously reported evolutionary selection of the bacterial proteins that acts in order to reduce the probability of promiscuous interactions between highly abundant proteins (Levy and

Teichmann, 2012), which surface has adapted to be more soluble than lowly abundant proteins. The standard error of A3D average scores is clearly higher in the low abundance group than in those corresponding to highly expressed polypeptides. This indicates that low abundance proteins exhibit a wider distribution of structural aggregation propensity (SAP), which is consistent with the observation that the bimodal solubility distribution is characteristic of low abundant bacterial proteins (Niwaa *et al.*, 2009).

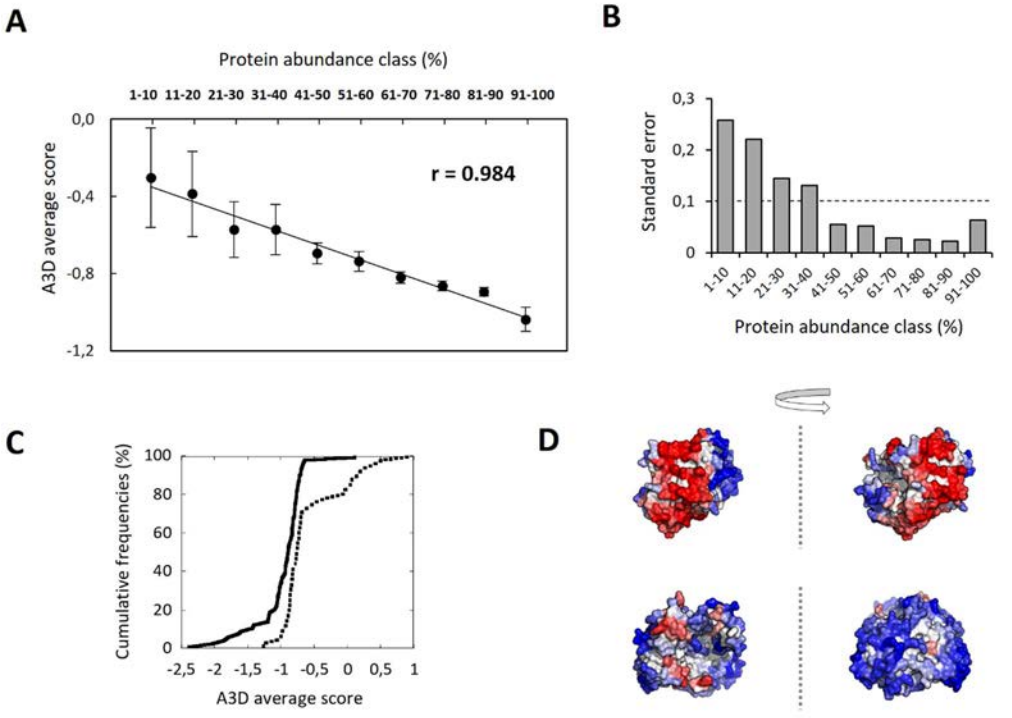


Figure 3.1. The relationship between surface aggregation propensity (SAP) and abundance for bacterial proteome. (A) The 612 bacterial proteins belonging to the complete set are divided into 10 classes regarding their protein abundance and the average A3D scores for each group, indicating SAP, were calculated (r = 0.984; p ≤ 0.0001). (B) Bar graph representing standard errors of A3D scores in different protein abundance classes. The standard error of the complete set is 0.1 (dashed line). (C) A cumulative distribution of SAP for the 10% most- (full line) and the 10% least abundant proteins (dashed line) (p ≤ 0.0001). (D) A3D analysis of representative low-abundance (PDB: 2WSX:A; *upper panel*) and high-abundance proteins (PDB: 1DFO:A; *lower panel*). The A3D average scores for 2WSX:A and 1DFO:A are 0.052 and –0.799, while $\log_{10}$ values of their protein abundance are -1.520 and 3.420, respectively. The protein surface is coloured according to A3D score in a gradient from red (high-predicted aggregation propensity) to white (negligible impact on protein aggregation) to blue (high-predicted solubility).

In contrast, highly abundant proteins are more tightly regulated in terms of SAP (Fig. 3.1B). The comparison of the aggregation propensity distribution of the 10% most and least abundant proteins in this dataset is shown in Fig. 3.1C, which illustrates again how highly abundant proteins and proteins present at low concentrations exhibit differential aggregation propensities ($p \leq 0.0001$), with low abundant proteins exhibiting surfaces that can support much higher aggregation propensity than those of high abundant proteins (Fig. 3.1D).

The higher solubility of high abundant proteins would work to prevent the aggregation of these proteins, even if they become concentrated at specific subcytosolic locations. Moreover, because of their high concentrations, their low deposition propensity would contribute significantly to decrease the overall cellular aggregation tendency and prevent the initiation of spontaneous, non-specific aggregation processes that can deplete the cell of less represented and/or functionally important proteins. This would imply that lowly abundant proteins might be more tolerant than abundant proteins to amino acid changes that significantly affect their SAP, which might be one of the underlying reasons accounting for well-established correlation between protein abundance and evolutionary conservation.

Different studies have reported previously a relationship between the abundance of proteins and the aggregation-propensity of linear protein sequences (Monsellier *et al*., 2008; de Groot and Ventura, 2010; Castillo *et al*., 2011). In order to test whether this relationship also exists in our complete data set, we used our linear aggregation-predictor AGGRESCAN to calculate the aggregation propensities of all sequences in the dataset and then grouped them as described above. The overall correlation between the aggregation propensities of the sequences and their length is significant ($r = 0.873$; $p \leq 0.0001$) (Fig. 3.2A), but lower than the one obtained using SAP values. Again, the 10% most abundant proteins display more soluble sequences than the 10% least abundant (Fig. 3.2B).

However, the significance of the solubility differences between these two subclasses is lower when considering sequences than when we consider structures, which strongly suggests that it is the aggregation propensity of protein structures which is really constrained by evolution according to the protein abundance required to exert any particular function, although, this propensity is somehow imprinted in the primary sequence. In other words, it seems that misassembly, rather than misfolding, accounts for the negative correlation between SAP and abundance. This implies that SAP is adjusted to the gene expression levels required for an optimal cell function. The consequence is that there is little margin of response in front of changes that decrease the protein surface solubility or increase expression levels (Tartaglia *et al*., 2007), both effects resulting in an increase of the probability to establish anomalous protein-protein interactions.

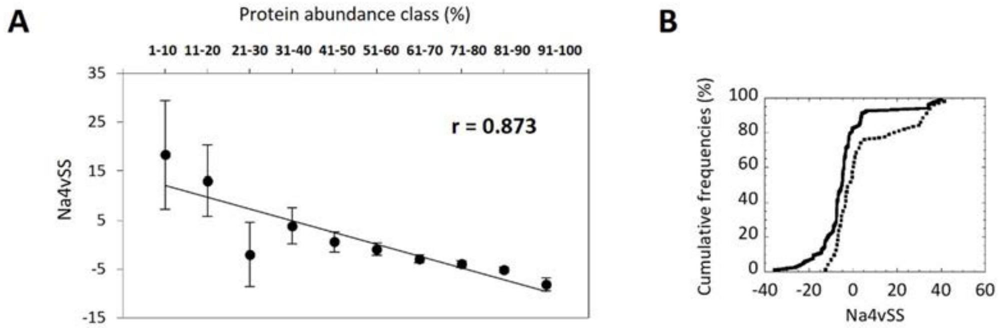


Figure 3.2. The relationship between sequential aggregation propensity and abundance for bacterial proteome. (A) The 612 bacterial proteins are divided into 10 classes regarding their protein abundance and average NA4vSS values for each group were calculated (r = 0.873; p ≤ 0,0001). (B) A cumulative distribution of Na4vSS values for the 10% most- (full line) and the 10% least abundant proteins (dashed line) (p ≤ 0.01).

3.2.2. Relationship between protein length and structural aggregation propensity

Previous studies have suggested that the length of a given protein might be an important determinant of its aggregation propensity. In this way, it was shown that longer proteins were more likely to co-aggregate and be sequestered *in vivo* by artificially designed aggregation-prone polypeptides. Also, it has been reported that the least soluble proteins of three distinct eukaryotic organisms share several common traits, one of them being generally longer in comparison to the highly soluble proteins (Albu *et al*. 2015). Our, A3D analysis report similar observations about the relationship between the aggregation-propensity of bacterial protein structures and their length. Proteins displaying transmembrane domains were excluded from this analysis, resulting in a total of 573 structures. This set was divided into 8 groups according to their length and the A3D average scores of the proteins in these subsets calculated and averaged. A significant correlation between length and aggregation-propensity was observed (r = 0.755; p ≤ 0.0001) (Fig. 3.3A). However, the correlation between size and SAP is not uniform along all the set. For proteins below 500 Da in size, the SAP increases logarithmically with the size of the protein, the smaller protein structures being clearly more soluble (r = 0.986; p ≤ 0.001). Above the length of 500 Da, the SAP doesn't increase anymore indicating that it exists an aggregation limit that cannot be overpassed even for very long proteins (Fig. 3.3A). The comparison of the 10% of the shortest and the 10% of the longest polypeptides of the same protein set again demonstrates that both groups exhibit different tendencies to aggregate (p ≤ 0.0001) (Fig. 3.3B). The shortest proteins exhibit a wider distribution of A3D average scores. This suggests that the dynamic range of aggregation propensities in this group is larger when compared to the group composed of the 10% of the longest polypeptides.

Accordingly, the shortest protein group is enriched in highly soluble proteins but also contains certain aggregation-prone structures exhibiting thus a certain bimodal distribution of surface aggregation propensities as proposed by Niwaa and co-workers (Niwaa *et al.*, 2009). Proteins exhibiting high solubility are very rare in the larger protein subset. To confirm the relationship between size and solubility, we compared the size the 10% more soluble proteins, according to their A3D average score, and the 10% less soluble. These two sets correspond to proteins with different size properties ($p \leq 0.0001$) (Fig. 3.3B).

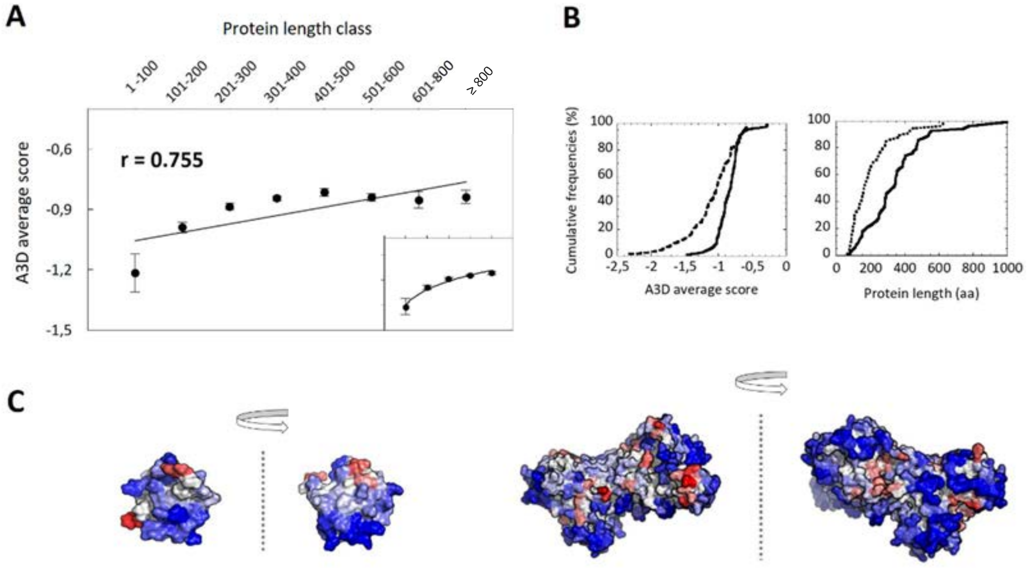


Figure 3.3. The relationship between SAP and length of bacterial proteins. (A) The 573 bacterial proteins that do not have transmembrane segments (TS) are divided into 8 classes regarding their length and the average A3D scores for each group were calculated ($r = 0.755$; $p \leq 0.0001$). The *inset* corresponding to groups of proteins below 500 Da in size shows the logarithmic increase of SAP with the size of the protein ($r = 0.986$; $p \leq 0.001$). (B) *Left panel:* a comparison between cumulative distribution of SAP for the 10% longest (full line) and the 10% shortest proteins (dashed line) ($p \leq 0.0001$). *Right panel:* cumulative distribution of protein length for the 10% proteins with the highest SAP (full line) and the 10% proteins with the lowest SAP (dashed line) ($p \leq 0.0001$). (C) A3D analysis of representative short (PDB: 1DCJ:A; *left panel*) and long (PDB: 2JGD:A; *right panel)* proteins. The A3D average scores for 1DCJ:A and 2JGD:A are -0.991 and -0.747, while $\log_{10}$ values of their protein abundance are 1.690 and 2.690, respectively. Colour code is as in Figure 3.1.

As expected, a group of low aggregation-prone proteins is enriched in small polypeptides and does not include very long proteins, whereas the proteins in the high aggregation-prone set are clearly larger (Fig. 3.3B). A visual inspection of the aggregation surfaces of short and long proteins suggests that the higher number of exposed aggregation-prone structural patches in the longest proteins is an important determinant of their higher SAP (Fig. 3.3C). The presence of these patches likely corresponds to the fact that these proteins can interact simultaneously with a larger number of partners or through larger protein interfaces, but this functionality comes at the expenses of a higher risk of a large number of non-native interactions of their folded structures.

A relationship between the length and the aggregation-propensity of protein sequences has been already described (Monsellier *et al.*, 2008). In order to discard that the observed relationships between A3D scores and protein length were indeed just a reflection of this correlation, we used our linear aggregation-predictor AGGRESCAN to calculate the aggregation propensities of the 573 sequences and then grouped them as described above. The overall correlation between the aggregation propensities of the sequences and their length in this dataset is poor (r = 0.014; p ≤ 0.0001) (Fig. 3.4A). In good agreement with our previous analysis of the complete *E. coli* proteome (de Groot and Ventura, 2010), there is a linear increase of the aggregation propensity until the size of 300 Da and from this size, the aggregation propensities of sequences decrease with the size of the protein. This makes that the overall aggregation propensities of the 10% shortest and the longest protein sequences do not differ significantly (p ≥ 0.05) (Fig. 3.4B).

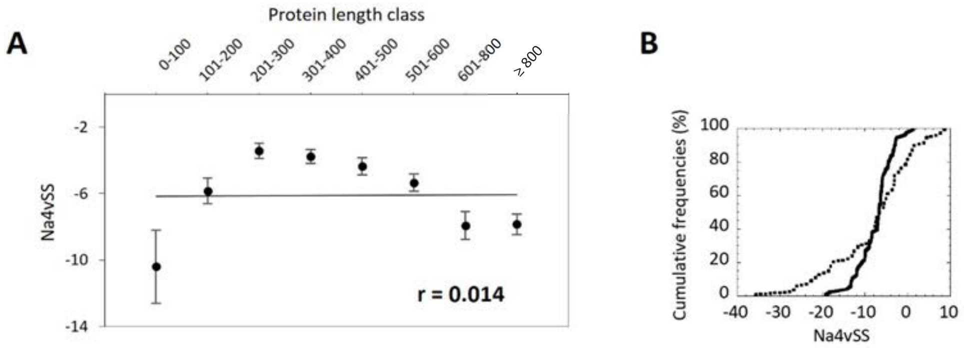


Figure 3.4. The relationship between sequential aggregation propensity and length of bacterial proteins. (A) The 573 bacterial proteins are divided into 8 classes regarding their length and the NA4vSS values for each group were calculated (r = 0.014; p ≤ 0.0001). (B) A cumulative distribution of Na4vSS values for the 10% longest bacterial proteins (full line) and the 10% shortest ones (dashed line) (p ≥ 0.05).

Thus, sequences and structures are both constrained by aggregation propensity but in different ways, which makes sense when we think that many of the sequential aggregation-prone regions are hidden inside the structure when the protein is folded and that many structural aggregation-prone regions are composed of residues that are not consecutive in the sequence.

3.2.3. Comparison between sequential and structural aggregation propensity predictions

As discussed above, since the sequential and structural determinants of protein aggregation might somehow overlap, we decided to compare systematically how they are connected in our dataset (Fig. 3.5), using Venn diagrams and hypergeometric tests. The aggregation propensity of the proteins in the dataset was predicted using both A3D (A3D score) and AGGRESCAN (Na4vSS). 34% and 40% of the predicted 20% most soluble and most insoluble proteins overlap, respectively, in the predictions of both algorithms.

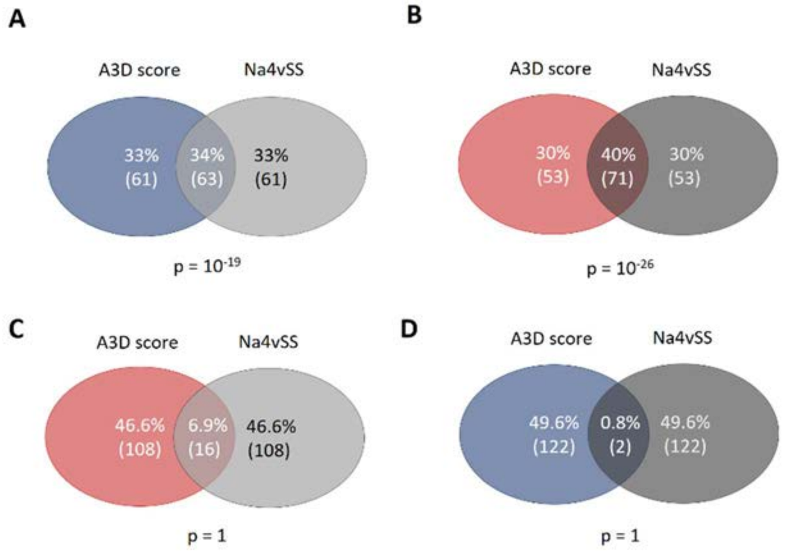


Figure 3.5. The comparison between sequential and structural aggregation propensity predictions. (A) The Venn diagram showing the overlapping between 20% most soluble proteins according to A3D (blue circle) with 20% most soluble proteins according to AGGRESCAN (light grey circle) ($p \leq 0.0001$). (B) The diagram showing the overlapping between 20% most aggregation-prone proteins according to A3D (red circle) with the 20% most aggregation-prone proteins according to AGGRESCAN (dark grey circle) ($p \leq 0.0001$). (C) The diagram showing the overlapping between 20% most aggregation-prone proteins according to A3D with the 20% most soluble proteins according to AGGRESCAN ($p \geq 0.05$). (D) The diagram showing the overlapping between 20% most soluble proteins according to A3D with the 20% most aggregation-prone proteins according to AGGRESCAN ($p \geq 0.05$).

This indicates that despite the fact that these two parameters are related, highly aggregation-prone sequences not necessarily give rise to very aggregation-prone structures, and vice versa (Fig. 3.5A and Fig. 3.5B). Indeed, there is a small proportion of proteins encoded by highly soluble sequences rendering highly aggregation-prone structures (6,9%), likely because scattered aggregation-prone residues in the sequence come together in the structure to form interaction interfaces with an inherent aggregation propensity (Fig. 3.5C). In contrast, highly soluble structures encoded by highly aggregation prone-sequences are rare (0.8%) (Fig. 3.5D).

The correlation between A3D scores and Na4vSS in the complete dataset is high (r = 0.839; p ≤ 0.0001) (Fig. 3.6A), but a significant contribution to this correlation comes from proteins with transmembrane domains, in which both local sequences and local structures are highly hydrophobic since they should be embedded in lipidic membranes. When these proteins are removed from the dataset, the correlation is still significant, but drops to r = 0.635 (p ≤ 0.0001) (Fig. 3.6B), indicating that despite these parameters are connected, they do not exactly reflect the same evolutionary constraints.

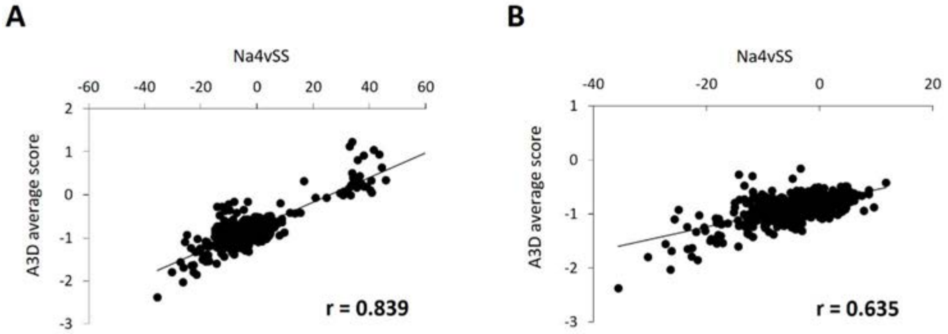


Figure 3.6. Correlation between aggregation propensity predicted by sequential and structural predictors for two different datasets. (A) Graph showing the correlation between Na4vSS values and A3D average scores for the dataset containing 619 proteins among which the proteins with transmembrane domains were included (r = 0.839; p ≤ 0.0001). (B) Graph showing the correlation between Na4vSS values and A3D average scores for the dataset containing 573 proteins without transmembrane domains (r = 0.635; p ≤ 0.0001).

3.2.4. The relationship between structural aggregation propensity and protein function in bacteria

We addressed whether there is any link between protein function and SAP in bacteria. The 20% of the proteins exhibiting the highest A3D average scores and the lowest A3D average scores belonging to the complete set, were selected for Gene Ontology (GO) enrichment analysis and clustered manually according to their GO terms (Fig. 3.7). Data reveals that the proteins identified by A3D as the 20% most soluble are involved in two essential cellular processes, protein folding and cell cycle (Fig. 3.7A), and are all located in the bacterial cytoplasm. These two functions are key for protein homeostasis as well as for the cell division and multiplication and it makes sense that nature has evolved molecules with a reduced risk to be inactivated by the anomalous interactions for these activities. On the other hand, proteins identified as the most aggregation-prone are involved in the establishment of the virus and symbiotic interactions, catabolic processes and ion transportation (Fig. 3.7B), and reside in the cytoplasm, as well as, in the cell wall and membrane. These proteins display in general large interaction interfaces used either to contact with exogenous entities, to form macromolecular transport complexes or to interact with biological membranes. Thus, there is a clear separation of ontology terms and functions according to proteins SAP.

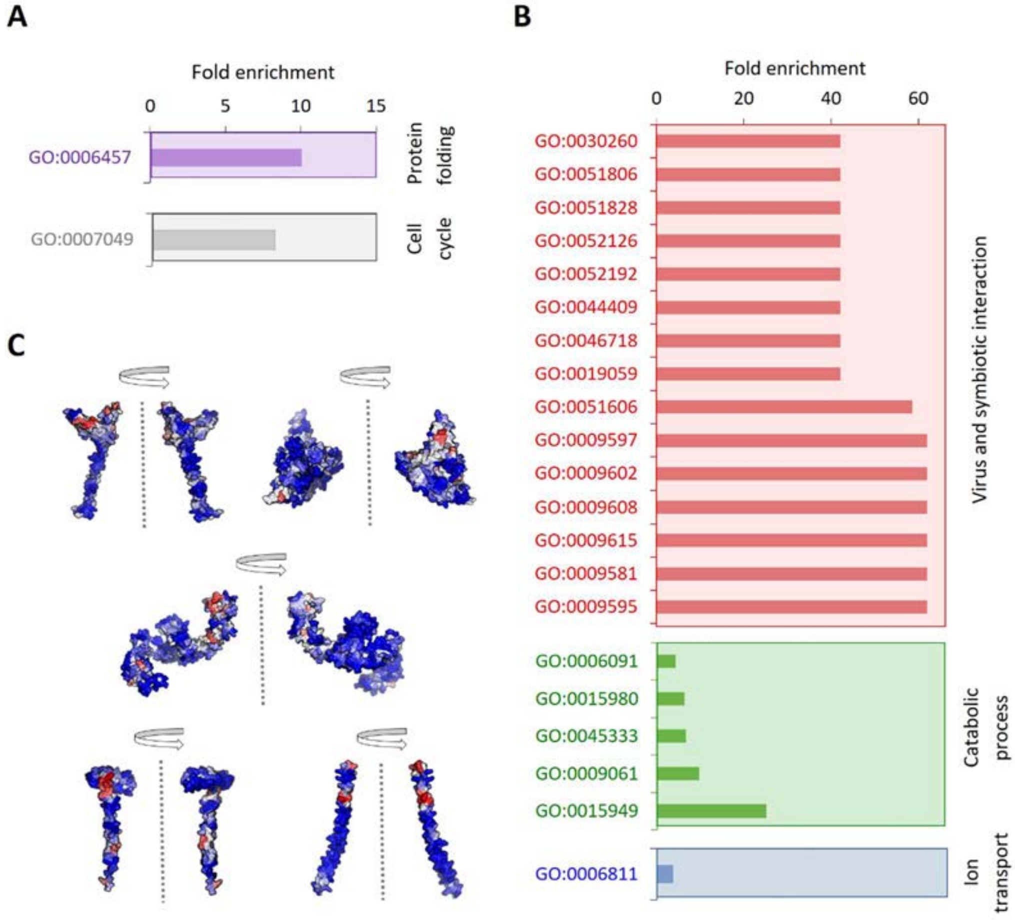
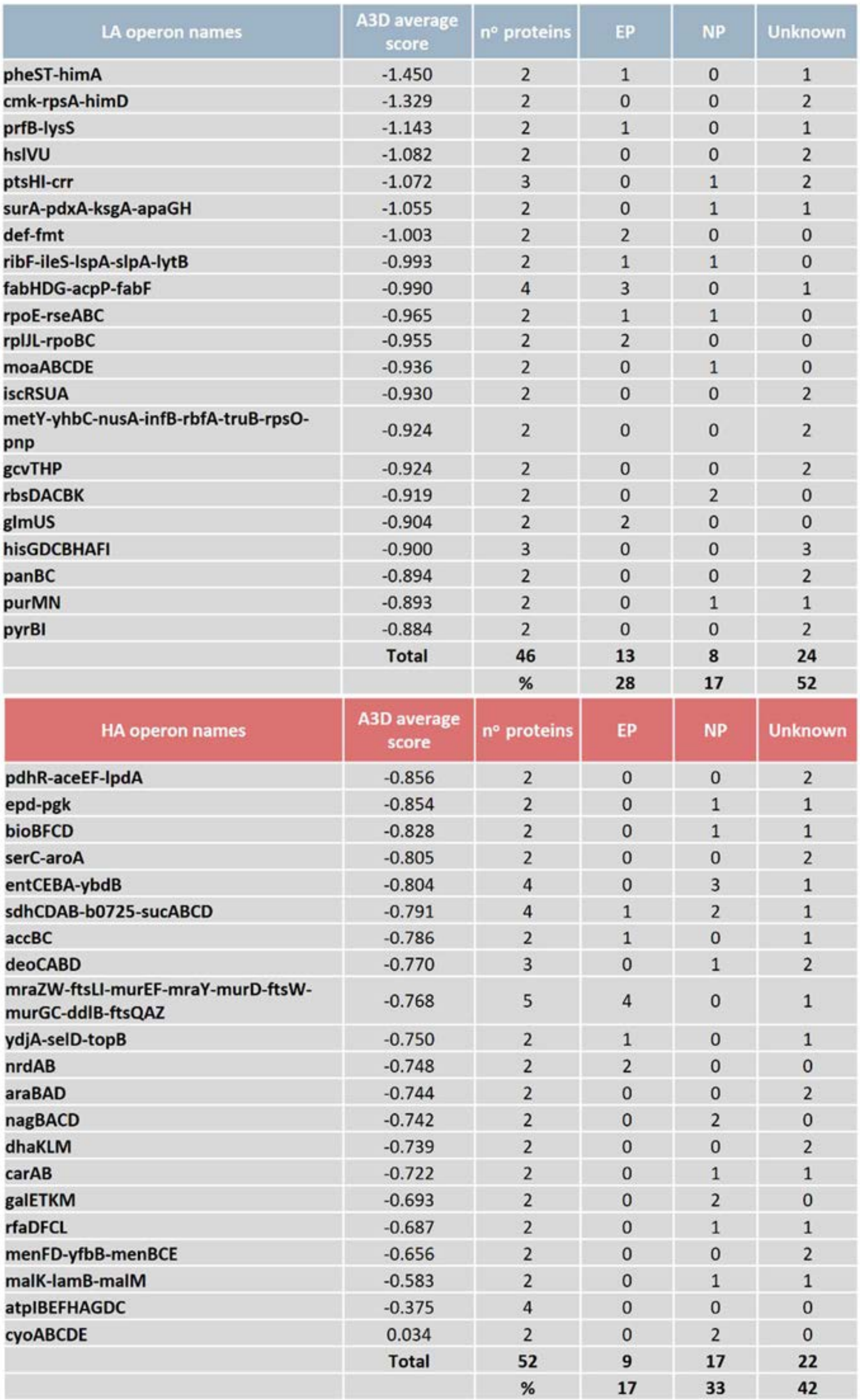


Figure 3.7. The relationship between SAP and biological function for bacterial proteins from the complete set. (A) Clustering of 20% proteins with the lowest SAP; (B) Clustering of 20% proteins with the highest SAP according to gene ontology (GO) terms. Clusters were generated manually based on initially proposed clusters by the functional annotation tool DAVID. (C) A3D analysis of representative proteins with the lowest SAP scores: GrpE protein (HSP-70 cofactor) (PDB: 1DKG:A; A3D average score = -1.402); Chaperone protein DnaK (PDB: 4JN4:A; A3D average score = -1.144); Trigger factor (TF) (PDB: 1W26:A; A3D average score = -1.598); Cell division proteins ZapA (PDB: 4P1M:A; A3D average score = -1.689) and ZapB (PDB: 2JEE:A; A3D average score = -2.373). Colour code is as in Figure 3.1.

Table 3.1. The list of operons regulating proteins with SAP lower than the mean aggregation propensity of the complete operon protein set (A3D average score = -0.880) (blue), together with the list of operons regulating proteins with SAP higher than the mean aggregation propensity of the complete operon protein set (red).



LA operon names	A3D average score	nº proteins	EP	NP	Unknown
pheST-himA	-1.450	2	1	0	1
cmk-rpsA-himD	-1.329	2	0	0	2
prfB-lysS	-1.143	2	1	0	1
hslVU	-1.082	2	0	0	2
ptsHI-crr	-1.072	3	0	1	2
surA-pdxA-ksgA-apaGH	-1.055	2	0	1	1
def-fmt	-1.003	2	2	0	0
ribF-ileS-lspA-slpA-lytB	-0.993	2	1	1	0
fabHDG-acpP-fabF	-0.990	4	3	0	1
rpoE-rseABC	-0.965	2	1	1	0
rplJL-rpoBC	-0.955	2	2	0	0
moaABCDE	-0.936	2	0	1	0
iscRSUA	-0.930	2	0	0	2
metY-yhbC-nusA-infB-rbfA-truB-rpsO-pnp	-0.924	2	0	0	2
gcvTHP	-0.924	2	0	0	2
rbsDACBK	-0.919	2	0	2	0
glmUS	-0.904	2	2	0	0
hisGDCBHAFI	-0.900	3	0	0	3
panBC	-0.894	2	0	0	2
purMN	-0.893	2	0	1	1
pyrBI	-0.884	2	0	0	2
	Total	**46**	**13**	**8**	**24**
		%	**28**	**17**	**52**

HA operon names	A3D average score	nº proteins	EP	NP	Unknown
pdhR-aceEF-lpdA	-0.856	2	0	0	2
epd-pgk	-0.854	2	0	1	1
bioBFCD	-0.828	2	0	1	1
serC-aroA	-0.805	2	0	0	2
entCEBA-ybdB	-0.804	4	0	3	1
sdhCDAB-b0725-sucABCD	-0.791	4	1	2	1
accBC	-0.786	2	1	0	1
deoCABD	-0.770	3	0	1	2
mraZW-ftsLI-murEF-mraY-murD-ftsW-murGC-ddlB-ftsQAZ	-0.768	5	4	0	1
ydjA-selD-topB	-0.750	2	1	0	1
nrdAB	-0.748	2	2	0	0
araBAD	-0.744	2	0	0	2
nagBACD	-0.742	2	0	2	0
dhaKLM	-0.739	2	0	0	2
carAB	-0.722	2	0	1	1
galETKM	-0.693	2	0	2	0
rfaDFCL	-0.687	2	0	1	1
menFD-yfbB-menBCE	-0.656	2	0	0	2
malK-lamB-malM	-0.583	2	0	1	1
atpIBEFHAGDC	-0.375	4	0	0	0
cyoABCDE	0.034	2	0	2	0
	Total	**52**	**9**	**17**	**22**
		%	**17**	**33**	**42**

The set of genes in an operon share a common gene expression regulation and are generally connected by their biological function. As a result, proteins encoded by the same operon are suggested to be present in similar amounts in the cell (Ishihama *et al*., 2008). The observed association between SAP and abundance would imply that polypeptides in the same operon should have related aggregation properties. To test this hypothesis, we first ascribed the proteins in our complete dataset to a particular operons. 234 out of 619 proteins could be ascribed to a particular *E. coli* operon (Table 3.1). We calculated A3D average scores and found out that the standard deviation of the A3D average score between proteins regulated by the same operon is lower in 93% of the cases 38 out of 41 than the standard deviation in the complete set of proteins (0.4) that could be ascribed to a particular operon (Fig. 3.8). This suggests again a link between protein aggregation propensities and the rates of transcriptional initiation.

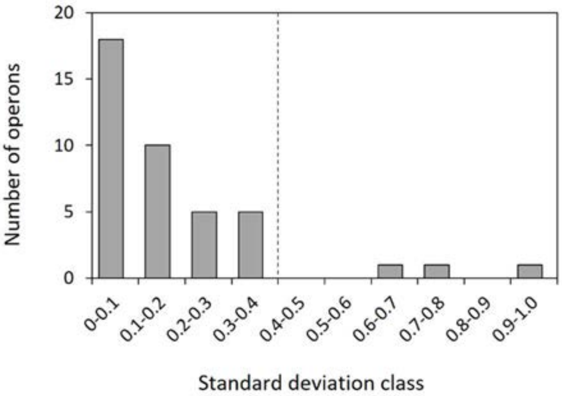


Figure 3.8. The variation in A3D average scores for known bacterial operons. The standard deviation of A3D average scores in 41 analysed operons was calculated and operons were divided into 10 classes. The dashed line indicates the standard deviation in the complete set of proteins (0.4). Low standard deviation within an operon indicates that the aggregation propensity of its proteins is similar.

The impact of anomalous protein interactions on cellular fitness would be ultimately associated with proteins functions. Therefore, it is conceivable that evolution would select for an overall decreased aggregation propensity in operons performing essential cellular functions. To explore this possibility, the bacterial operons were divided into two groups according to their A3D average scores, those with the lower and those with the higher aggregation propensity than the mean propensity of the complete operon protein set (A3D average score = -0.880). The essentiality of approximately half of the proteins in each subset has been annotated via genetic footprinting or knockout experiments (Gerdes *et al*., 2003; Baba *et al*., 2006). Importantly, considering only the annotated polypeptides, operons with low aggregation tendency regulate 62% of essential proteins (EP) and 36% of non-essential ones (NP). Instead, operons with high aggregation propensity, regulate 35% of essential and 65% of non-essential proteins (Table 3.1). This supports the view that the structures of essential bacterial proteins suffer a stronger selection against misassembly and aggregation than those of non-essential ones. Because smaller protein structures tend to be more soluble than the longer ones, it could be that the relationship between essentiality and aggregation results simply from the fact that essential proteins tend to be smaller.

To discard this possibility, we compared the length of essential and non-essential proteins in the complete set. 93 proteins were identified as essential and 153 proteins were identified as non-essential in the available databases. No significant differences in length were observed between these two groups ($p \geq 0.05$) (Fig. 3.9A). The average size of the conjunct of EP and NP proteins is 400 amino acids.

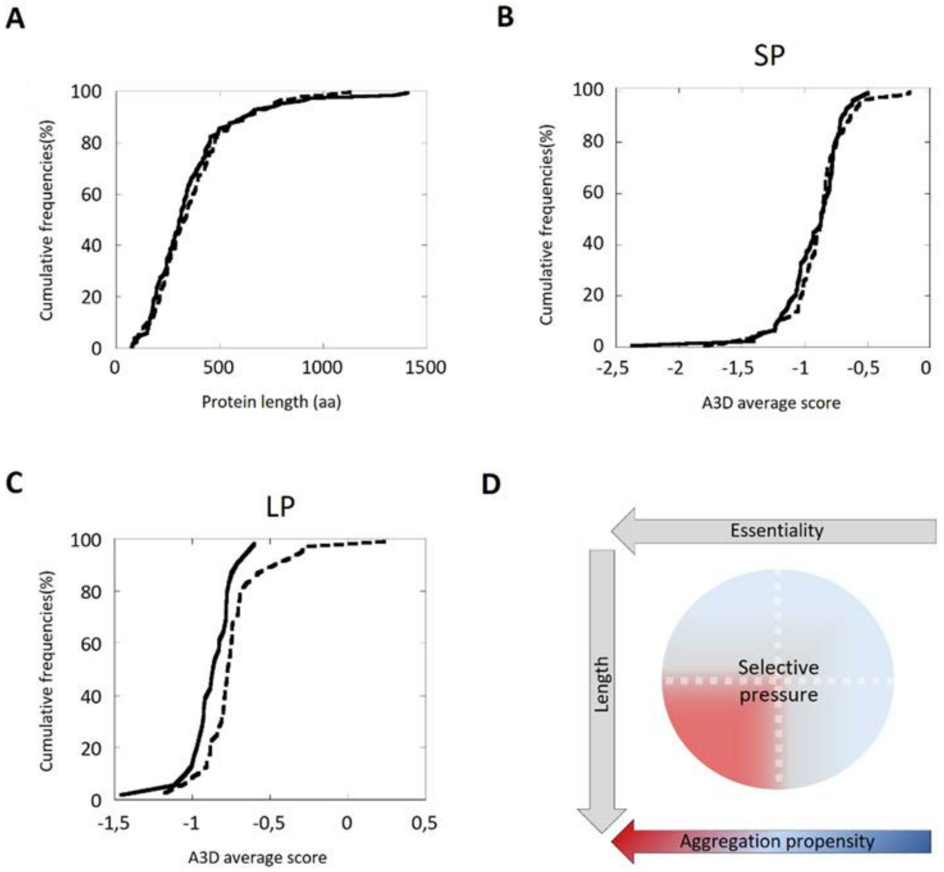


Figure 3.9. The relationship between SAP, protein essentiality and protein length. (A) A cumulative distribution of protein lengths for essential (full line) and non-essential proteins (dashed line) ($p \geq 0.05$). (B) A comparison between cumulative distributions of SAP for short (SP) essential and non-essential proteins ($p \geq 0.05$). (C) A comparison between cumulative distributions of SAP for long (LP) essential and non-essential proteins ($p \leq 0.05$). (D) The scheme indicating the group of proteins upon which the evolution imposed restrictions aimed to minimize SAP, due to their length and essentiality.

These two groups were divided into subgroups regarding their length: a group of short proteins (SP) with length < 400 amino acids and a group of long proteins (LP) with length ≥ 400 amino acids, and the aggregation propensities of the corresponding groups were compared according to their essentiality. The cumulative distribution of A3D average scores for SP group shows that there is no difference in the aggregation propensities of short essential and non-essential proteins ($p \geq 0.05$) (Fig. 3.9B). Unlike SP, the LP group exhibits significantly different A3D scores depending on the protein essentiality, with EP proteins being overall more soluble ($p \leq 0.05$) (Fig. 3.9C). Thus, it seems that the evolutionary constraints opposing protein misassembly are especially strong in long proteins (Fig. 3.9D), likely due to their higher propensity to establish non-functional interactions leading to protein inactivation and/or aggregation.

3.2.5. Effect of subcellular location on the structural aggregation propensity of bacterial proteome

Bacterial proteins can populate different subcellular compartments characterised by specific physiological conditions. It's of great importance that the protein maintains its biological function in its native state no matter in which subcellular location this protein resides (Monsellier *et al.*, 2008; Linding *et al.*, 2004). Bacterial proteins are found in other compartments apart from the cytosol, like the periplasm and the inner and outer membranes. Presumably, their aggregation properties would be adapted to the specific environment in those subcellular locations. Our A3D analysis data shows that the proteins that reside in the bacterial cytoplasm and periplasm possess the lowest SAP of the complete dataset (Fig. 3.10A and Fig. 3.10B). This result fits perfectly with the fact that the proteins in these subsets also display protein sequences of similar high solubility when analysed with linear algorithms (de Groot and Ventura, 2010).

Proteins with transmembrane domains were excluded and the location of the 20% proteins with the lowest A3D average scores and 20% proteins with the highest A3D average scores were analysed in terms of gene ontology (GO) to assess if any of these subsets were enriched in any particular cell compartment. We found out that in addition to the cytosol, soluble proteins were enriched in the periplasm, whereas enrichment in this last compartment was not observed among the more aggregation-prone protein structures (Fig. 3.10C). This is consistent with the fact that the periplasm, in contrast to the cytosol, lacks a sophisticated cellular system to control protein quality and avoid aggregation (Dougan *et al.*, 2002), and is separated from the outside solution by a highly permeable outer membrane that provides limited protection against environmental variations.

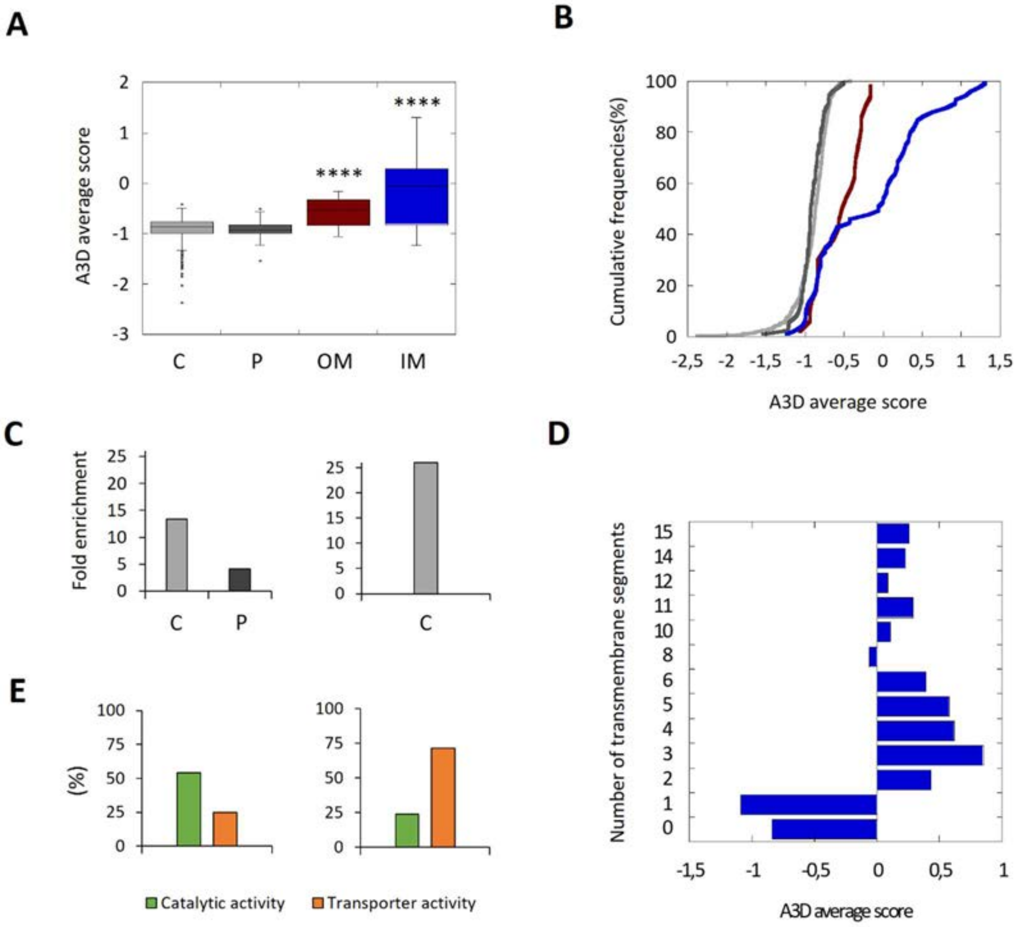


Figure 3.10. The relationship between SAP and subcellular location of bacterial proteins. (A) The complete bacterial dataset was divided into 4 groups based on the protein localization according to UniProt. (B) A cumulative distribution of SAP of proteins located in the cytoplasm (C; light grey), periplasm (P; dark grey), outer membrane (OM; dark red) ($p \leq 0.0001$) and inner membrane (IM; blue) ($p \leq 0.0001$). (C) Clustering of 20% proteins with the lowest- (*left panel*) and 20% proteins with the highest SAP (*right panel*) belonging to the dataset in which proteins with TS were excluded, according to GO terms. Clusters were generated manually based on initially proposed clusters by the functional annotation tool DAVID. (D) Diagram for the analysed IM proteins showing the number of TS and SAP. (E) An analysis of the molecular functions associated to IM proteins with lower (A3D average score < 0) and higher SAP (A3D average score ≥ 0) according to GO terms proposed by PANTHER.

Inner membrane proteins exhibit the highest aggregation propensities of all bacterial proteins according to A3D analysis (Fig. 3.10A and Fig. 3.10B). The gram-negative bacterial inner membrane is a semipermeable shield that preserves the cytoplasm environment. The proteins associated with this membrane can have a variable number of transmembrane segments (TS) per protein (Santoni *et al.*, 2000). These regions are stable in the hydrophobic environment of this lipid bilayer due to their enrichment in apolar residues. A visual inspection of their structures exhibits that the highly aggregation-prone surfaces identified by A3D in many of these proteins sharply coincide with regions embedded in the membrane, since the membrane width can be traced in their A3D coloured structure (Fig. 3.11).

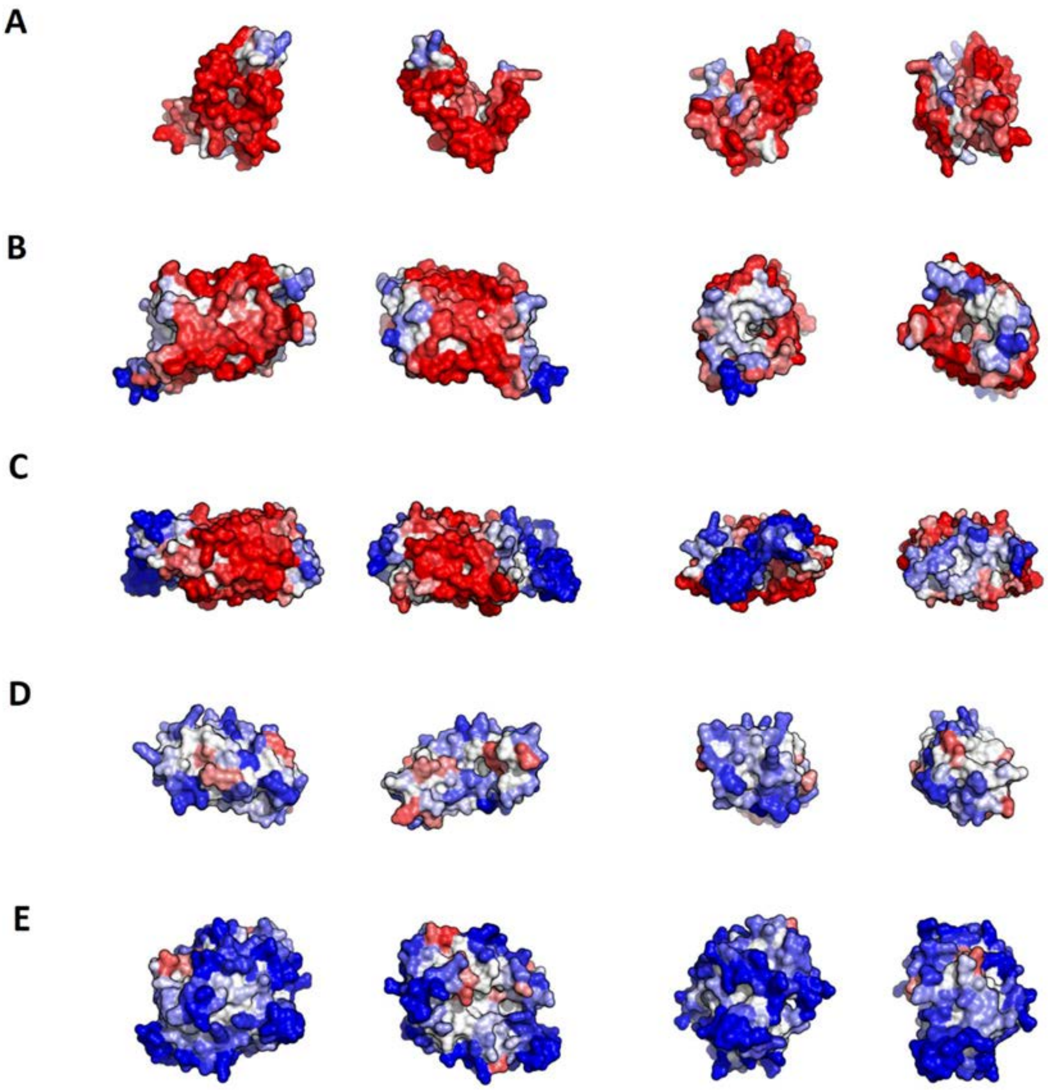


Figure 3.11. A3D analysis of selected representative inner membrane proteins. (A) Disulfide bond formation protein B (PDB: 2HI7:B; A3D average score = 1.233); (B) Aquaporin Z (PDB: 1RC2:A; A3D average score = 0.417); (C) Nitrate/nitrite transporter NarK (PDB: 4JR9:A; A3D average score = 0.051); (D) Acyl-CoA thioester hydrolase YbgC (PDB: 1S5U:A; A3D average score = -0.664); and (E) Septum site-determining protein MinD (PDB: 3Q9L:A; A3D average score = -1.031). Colour code is as in Figure 3.1.

We found that the main difference between these two groups is the number of TS (Fig. 3.10D). In fact, the inner membrane group contains a number of proteins without TS and a low SAP. An analysis of their structures indicates that they are devoid of large aggregation-prone patches suggesting that they associate to the membrane transiently (Fig. 3.11D and Fig. 3.11E).

Proteins with a single TS also display significant solubility since apart from the short TS anchored to the membrane (usually an α-helix) they also exhibit large globular domains facing either the cytosol or the periplasm (Fig. 3.10C). Proteins exhibiting two or more TS account for the more aggregation prone structures in the analysed sub-proteome. Indeed, an analysis of the molecular functions associated to IM proteins with lower and higher aggregation propensity clearly indicates that they play different cellular roles, the low aggregation-prone ones displaying preferentially catalytic activity, and the high aggregation-prone proteins possessing mainly the role of transporters (Fig. 3.10E). Therefore, here we found again a significant association between the function of the protein and its associated SAP.

As IM proteins, outer membrane proteins are thought to be located in a hydrophobic environment, and consequently, they are expected to have a high aggregation tendency. However, their SAP lies between those of cytoplasmic and IM proteins (Fig. 3.10A and Fig. 3.10B). In fact, the outer membrane acts as a permeable barrier to hydrophobic substances. In general, outer membrane proteins display a β-barrel structure that encloses a hydrophilic cavity covered by a hydrophobic outer layer embedded in the membrane. A3D structural predictions are able to capture this particular architecture (Fig. 3.12). In the predictions it can be easily seen a high aggregation-prone fringe in the outside, restrained to the exact boundary embedded in the membrane, flanked by low aggregation-prone regions that protrude out of the membrane. Seen from above, or below, it can be observed that the residues flanking the cavity at both inner and outer sides conform surfaces of low aggregation propensity.

This particular assembly is achieved by alternating hydrophobic and hydrophilic segments in the sequence (Cowan *et al*., 1992; Schirmer, 1998). This explains why when linear predictors were used to analyse the aggregation propensities of these proteins, they were predicted as being even more soluble than the cytoplasmic proteins. In reality, in these proteins, the scattered aggregation-prone regions in the sequence come together in the structure to allow their stable insertion in the membrane. This apparent duality between the aggregation properties of the sequence and the structure might be important for their biogenesis, since the folding of proteins into the outer membrane presents important challenges to Gram-negative bacteria because they must migrate from the cytosol, through the inner membrane and into the periplasm before they could be recognized by the β-barrel assembly machinery and inserted into the outer membrane (Knowles *et al*., 2009). In most of these steps and compartments, the protein is unfolded and accordingly, sequences with reduced aggregation propensities would represent a selective advantage.

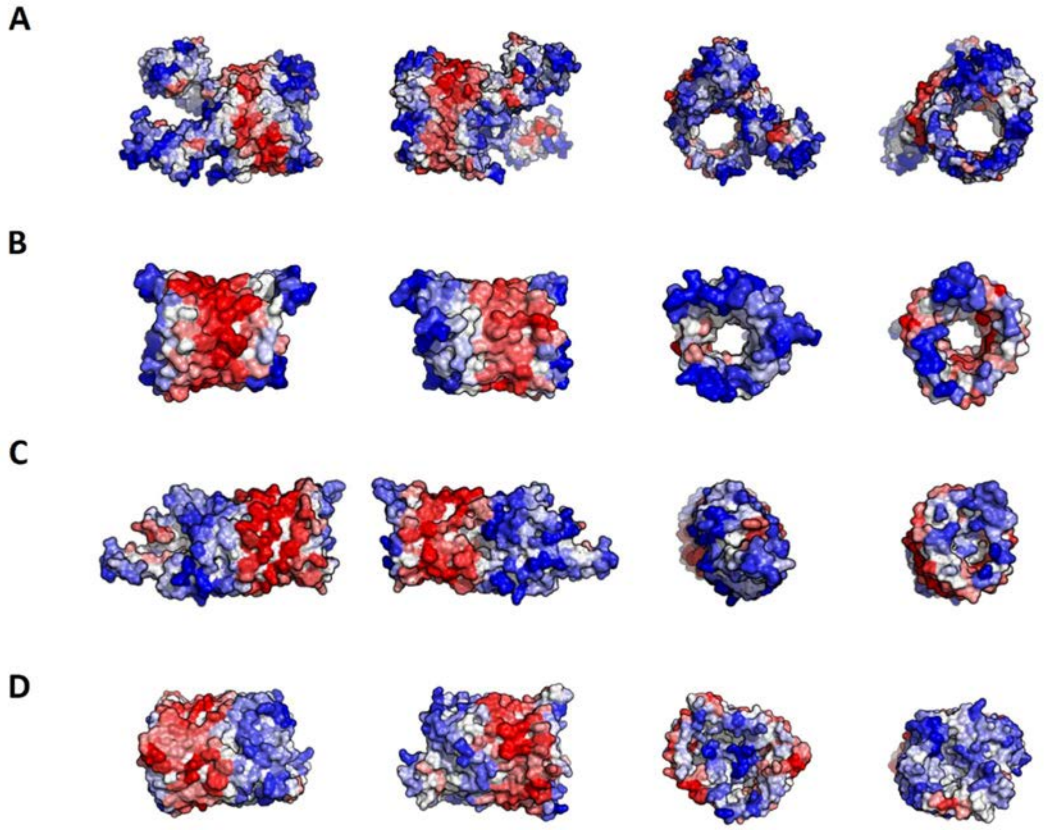


Figure 3.12. A3D analysis of selected representative outer membrane proteins. (A) Usher protein FimD (PDB: 3RFZ:B; A3D average score = -0.565); (B) Outer membrane protein G (PDB: 2F1C:X; A3D average score = -0.275); (C) Long-chain fatty acid transport protein (PDB: 1T16:A; A3D average score = -0.350) and (D) Fe(3+) dicitrate transport protein FecA (PDB: 1KMO:A; A3D average score = -0.296). Colour code is as in Figure 3.1.

3.2.6. Compositional, structural and thermodynamic determinants of protein structural aggregation propensity in the bacterial proteome

To understand whether there is any compositional bias in the surfaces of the 10% most- and 10% least aggregation-prone structures, we monitored the frequency of the 20 natural amino acids at their surfaces and compared. Eleven out of twenty amino acids did not show any significant enrichment in any of these two groups (Fig. 3.13A). However, we found out that the charged residues (glutamic acid, lysine and arginine) were significantly enriched in the soluble structures subset (Fig. 3.13A). It is generally accepted that charge−charge interactions are important for protein solubility by inducing long-range repulsion between alike-charged species. This has been clearly demonstrated by introducing multiple charged side chains, a strategy that provides an increased resilience toward aggregation, particularly at elevated temperatures (Lawrence *et al*., 2007). Conversely, neutralization of surface charges is suggested to be required for efficient amyloid fibril formation (Jeppesen *et al*., 2010) and chemical and mutational neutralization of charges has indeed been shown to promote the aggregation.

The aromatic residues (phenylalanine, tyrosine and tryptophan) were enriched in aggregation-prone structures (Fig. 3.13A). This can be related to the fact that exposed aromatic residues facilitate protein-protein interactions and indeed these three residues are more frequent at protein interfaces than at the surfaces. This property can be partially explained by their ability to establish both π-π or π-cation interactions, their flat surfaces and the entropic benefit of hiding them from water inside interfaces. However, these properties also imply an increased probability of establishing competing non-functional interactions. The trade-off between proper and anomalous interactions will explain why we did not find any hydrophobic aliphatic residue enriched in aggregation-prone structures, since they will increase the aggregation potential without providing a counteracting functional advantage. We also identified glycine and especially proline enriched in the surface of aggregation-prone proteins (Fig. 3.13A). These two residues are bad β-sheet formers and their presence might counteract the enrichment in aromatic residues by diminishing the protein probability to form wrong intermolecular β-sheets contacts leading to protein aggregates (Villar-Pique and Ventura, 2012).

Next, we studied the secondary structure content of proteins displaying differential SAP. To this aim, we identified and quantified the α-helical and β-sheet elements in all the structures in our database, without taking into account proteins displaying transmembrane segments. The low aggregation-prone structures were significantly enriched in α-helices and depleted in β-sheets relative to most aggregation-prone and the other way around (Fig. 3.13B). This is consistent with the observation that proteins that have undergone non-native aggregation often are observed to have increased levels of β-sheet secondary structures (Weiss *et al*., 2009). This indicates a redundant strategy to protect certain protein structures from aggregation since they have at the same time less solvent-exposed hydrophobicity, lower β-sheet content and are enriched in charged residues, three properties that would fight the formation of toxic/non-functional β-sheet enriched aggregates.

The flexibility of protein structures is linked both to protein stability and protein aggregation, in such a way that destabilizing mutations in proteins usually result in increased structural fluctuations and aggregation tendencies, without a need of a large polypeptide unfolding (Chiti and Dobson, 2009). Conversely, the stabilization of protein structures by mutations or co-solvents is a common strategy to prevent the aggregation of therapeutic proteins in the biotechnological industry. Recently, Leuenberger and co-workers assessed experimentally the *in vivo* thermostability of almost the complete *E. coli* proteome (Leuenberger *et al*., 2017). We used these data to test whether the most soluble and aggregation-prone protein structures in our dataset differ in their stability (Fig. 3.13C). Aggregation-prone proteins turned to be significantly more stable than soluble ones ($p \leq 0.0001$). This might reflect a safeguarding strategy to prevent that partial unfolding might further increase their already high propensity to establish non-functional interactions.

Translation is an error-prone process (Drummond and Wilke, 2009). Therefore, selection for proper protein structure and function should cause codons with reduced error rates to be selected at positions in which translation errors would be more disruptive. Accordingly, it has been found that in *E. coli*, translationally optimal codons associate with aggregation-prone sites (Zhou *et al*., 2009).

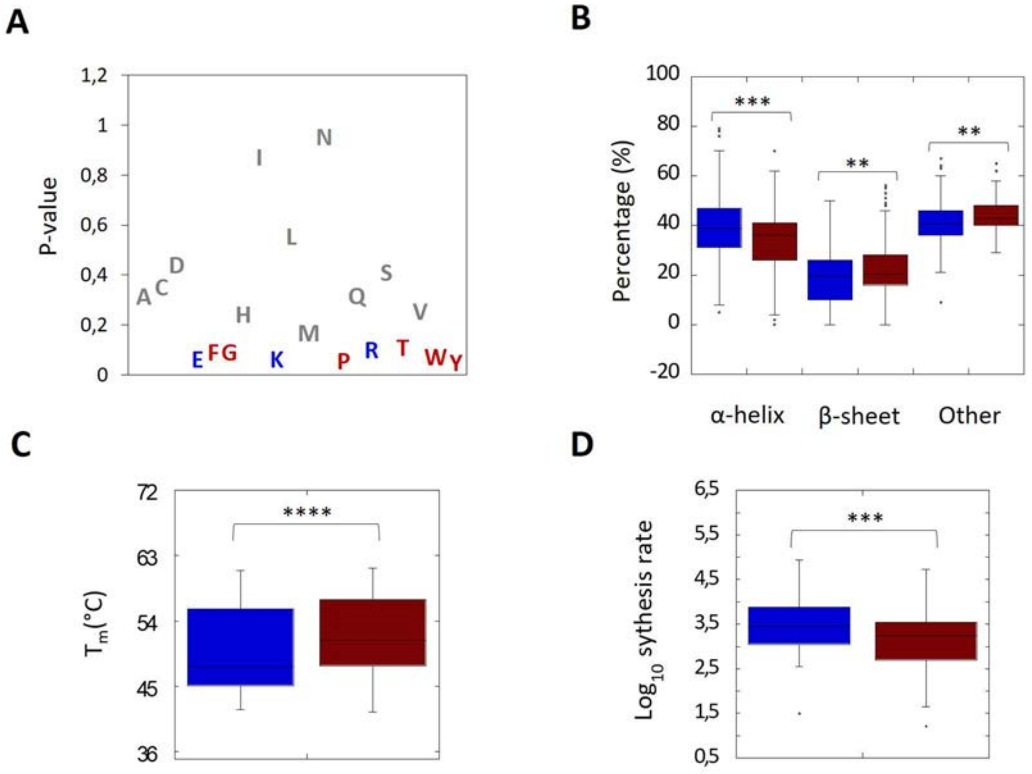


Figure 3.13. The comparison between SAP and structural features of bacterial proteins. (A) Plot of p-values of correlating surface amino acid composition with aggregation propensity for 2 groups: 10% least- (blue) and 10% most aggregation-prone proteins (red). In grey: amino acids that do not show significant differences. In blue: amino acids that are more abundant in soluble proteins. In red: amino acids that are more abundant in aggregation-prone proteins. (B) Comparison between SAP and protein secondary structure. Secondary structure was determined for 20% least- and 20% most aggregation-prone proteins. (C) Graphical representation of experimentally determined melting temperatures (T_m), indicating protein stability, for 10% least- and 10% most aggregation-prone proteins. The T_m median for group of soluble proteins is 46°C and for group of aggregation-prone proteins is 51°C. (D) Graphical representation of $\log_{10}$ values of protein synthesis rates under minimal growth conditions determined for 10% least- and 10% most aggregation-prone proteins.

At least two different explanations may account for this observation: (i) optimal codons may be required to reduce the frequency of translation errors at aggregation-prone sites that coincide with functional sites, such as protein–protein interfaces. (ii) Alternatively, optimal codons may be required for rapid translation of aggregation-prone protein regions. Importantly, the association between optimal codon and aggregation-prone residues seems to occur both at buried and at exposed sites, suggesting that burial and aggregation propensity to influence codon usage independently of each other. This would imply that this translation/aggregation link might be related not only to the protein sequence propensity but also to the SAP. To test this possibility, we compiled the synthesis rates of the proteins in our dataset from ribosome profiling experiments (Li *et al*., 2014) and found that, effectively, aggregation-prone proteins are synthesized at significantly higher rates than those having more soluble structures ($p \leq 0.001$) (Fig. 3.13D).

3.2.7. The relationship between evolutionary conservation and surface aggregation propensity in bacterial proteins

We asked whether SAP might be an evolutionary conserved feature. We first identified orthologous and paralogous genes of those genes encoding the proteins in our *E. coli* sub-proteome. Orthologs are genes from different species that evolved from a common ancestral gene through speciation and typically retain the same function. Paralogs arise from genomic duplication within the same species and typically evolve different functions. Paralogous proteins are almost always younger and arise from gene duplication events that occur after the speciation. Comparison of the variation in the SAP of proteins from orthologous and paralogous gene pairs indicates that, in both cases, the differences in SAP are larger between aggregation-prone structures than between soluble ones, indicating that the evolution of this last protein group is more constrained (Fig. 3.14A). However, only in the case of orthologous proteins, the difference is statistically significant ($p \leq 0.01$) which might result of their older evolutionary age compared with paralogous proteins (Fig. 3.14B). However, these differences do not respond, or at least not only, to overall differences in the sequences identities of the respective protein groups, suggesting that they can arise from the adoption of specific new functions, especially those mediated by protein interactions, in the case of paralogs or from the adaptation of novel cellular environments in the case of orthologs.

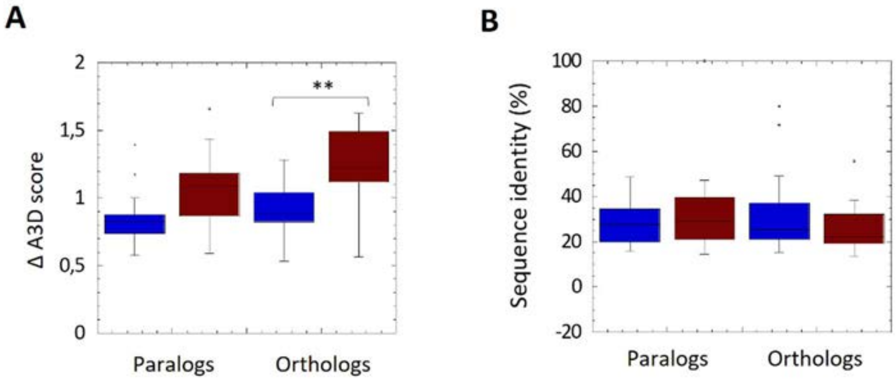


Figure 3.14. The relationship between evolutionary conservation and SAP for bacterial proteins. (A) The differences in A3D average scores for paralogous and orthologous protein pairs identified for 30 least- (blue) and 30 most aggregation-prone proteins (red) are shown. (B) The determined sequence identity for analysed paralogous and orthologous protein pairs.

3.2.8. The relationship between protein structural aggregation propensity and oligomeric assembly in bacteria

We decided to assess if the aggregation propensities of proteins that need to assemble into functional complexes in order to fulfil their biological functions would differ compared to those that perform their biological role as monomers. For that reason, we divided the analysed protein set into two groups regarding the protein active form: proteins active as monomers (M) and proteins active as oligomers (O), and the aggregation propensities of the monomers/subunits in the corresponding groups were compared ($p \leq 0.0001$) (Fig. 3.15A). The structures of M proteins were significantly less aggregation-prone than that of O proteins. This is not surprising if one takes into account that the physico-chemical properties leading to the formation of functional and non-functional assemblies overlap significantly, in such a way that the protein interfaces tend to be stickier than the surfaces, providing a driving force for the formation of the functional complexes but at the same time potentially leading to nonspecific interactions and aggregation. We explored the relationship between the number of units in the functional assembly and the aggregation propensities of these components (Fig. 3.15B). Not surprisingly the SAP of the M proteins increases almost linearly with the number of units needed to become functional, which is consistent with the need for a larger number of interactions sites in the subunits embedded in complexes containing more polypeptide chains.

As we described above, there is a negative correlation between protein abundance and SAP. Indeed, it has been shown that “supersaturated” proteins, which are maintained at a high concentration relative to their solubility are among the most aggregation susceptible proteins in different proteomes (Ciryam *et al.*, 2013). The presence of supersaturated proteins in the cell may respond to the requirements to exert their biological functions. The stoichiometry of multimeric proteins might imply that their monomeric subunits would be supersaturated relative to proteins that are active in the cell as monomers. In addition, high concentrations may also increase the probability of self-assembly. We calculated the structural supersaturation index (SSI) for both oligomeric and monomeric proteins in our dataset (Fig. 3.15C).

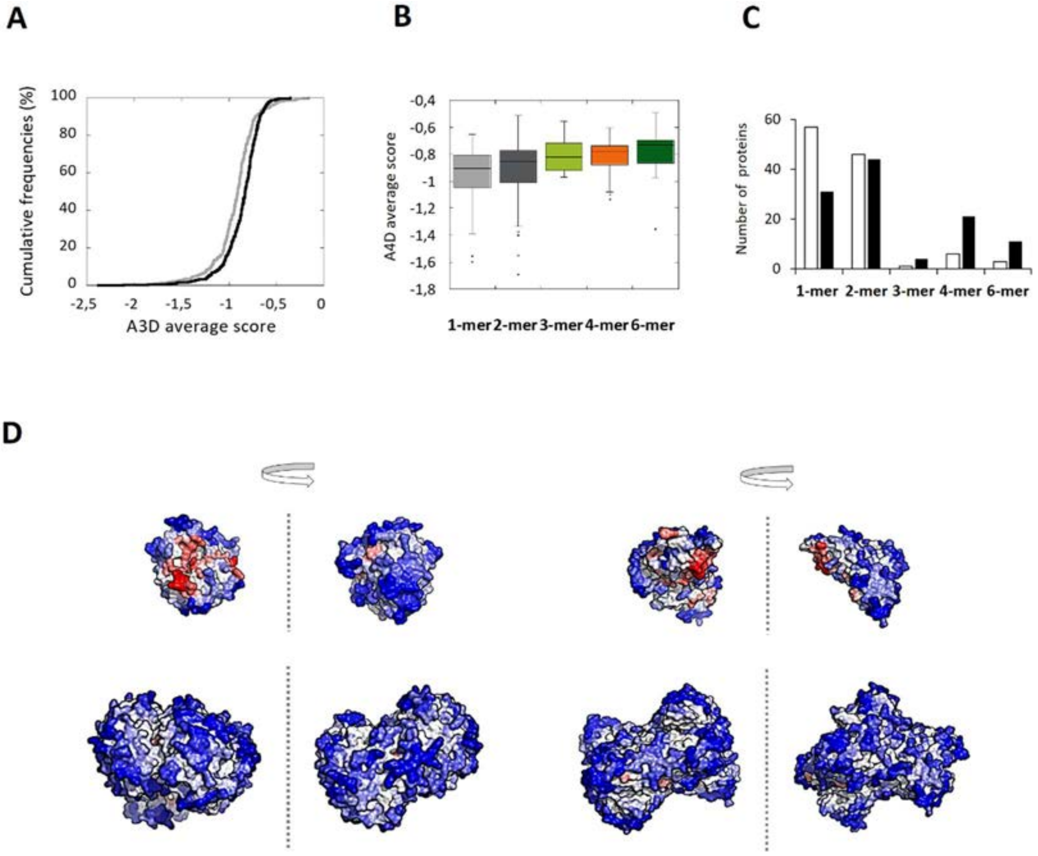


Figure 3.15. The relationship between SAP and protein active form. (A) A comparison between cumulative distributions of A3D average scores for proteins active as monomers (M; grey line) and proteins active as oligomers (O; black line) ($p \leq 0.0001$). (B) A bar graph representing the mean and the distribution of A3D average scores for the proteins that acquire M or O form in order to perform their biological role. (C) A comparison between protein active form and structural supersaturation index (SSI) identified for 20% of proteins with the lowest SSI (white bars) and 20% of proteins with the highest SI (black bars). (D) A3D analysis of representative proteins biologically active as tetramers. Left panel: analysis of one chain (PDB: 1FDY:A) and of biologically active tetrameric form (PDB: 1FDY). Right panel: analysis of one chain (PDB: 1CS1:A). and of biologically active tetrameric form (PDB: 1CS1). Colour code is as in Figure 3.1.

In monomeric proteins, polypeptides having a low SSI are predominant, indicating that they display generally low risk to aggregate. In dimers, we found a similar trend, whereas for oligomeric proteins displaying more subunits, high SSIs are predominant. Therefore, it becomes clear that the formation of multimeric proteins comes at the risk of a misassembly. Why these proteins do not aggregate *in vivo* under physiological conditions? Mainly, because the sticky interfaces needed for the assembly of the quaternary structure are protected in the native state (Fig. 3.15D) and accordingly the aggregation-risk is only transitory. This implies, however, that, cellular conditions or genetic changes that favour multimers dissociation would also favour non-functional interactions, as it occurs in the case of human TTR or SOD1, whose dissociation leads to their aggregation and the onset of amyloidosis.

3.3. Material and methods

Selection criteria for analysis of bacterial proteome

The protein dataset including 1103 different proteins whose presence could be experimentally detected in bacteria (Ishihama *et al*., 2008) was used to search through Protein Data Bank (PDB) (Berman *et al*., 2002) in order to identify proteins from the dataset which have available solved crystal structures. *E.coli* proteins with available three-dimensional structure in PDB format, with X-ray resolution ≤ 3.5, were considered in this analysis. Furthermore, available three-dimensional structures covering < 90% of protein primary sequence were excluded, as well as proteins displaying sequence homology > 40%, resulting in a set of 619 bacterial proteins.

Structural and sequential aggregation propensity predictions

Structural aggregation propensity (SAP) predictions for the resulting polypeptide set were performed using AGGRESCAN3D (A3D) algorithm which unlike other algorithms that rely on linear sequences, implements structure-based approach and predicts aggregation propensity of initially folded states (Zambrano *et al*., 2015b). Selected proteins were submitted to A3D as single protein chains in the 'Static Mode' and 10 Ǻ was selected as a distance for aggregation analysis (default sphere radius). For the study of the aggregation propensity of symmetric complexes, multimeric structures were analysed in the same way. The Average scores for each protein in the set were obtained from the A3D output interfaces. The Average score allows comparing the solubility of protein structures differing in size. It also allows assessing changes in solubility promoted by amino acid substitutions in a particular protein structure. To perform sequential aggregation analysis, AGGRESCAN (Conchillo-Sole *et al*., 2007) was implemented to assess the global protein aggregation propensities for the sequences corresponding to analysed protein structures. Na4vSS (Normalised a4v Sequence Sum for 100 residues Na4vSS) values were obtained.

Datasets

Protein abundance data were taken from PaxDb (Wang *et al*., 2012). The obtained values *were* then $\log_{10}$ transformed for statistical analysis. Information about protein length, protein subcellular location, number of transmembrane segments and protein active form were obtained from Swiss-Prot Protein knowledgebase (Boeckmann *et al*., 2003). Gene Ontology (GO) enrichment analysis was performed with the Database for Annotation, Visualization and Integrated Discovery (DAVID) (Dennis *et al*., 2003). Automatic GO clusters for three ontologies were retrieved and manually curated. The essentiality of bacterial proteins for cellular fitness was derived from the reported data (Gerdes *et al*., 2003; Baba *et al*., 2006) and updated with the data available at server EcoGene 3.0. (Zhou and Rudd, 2013). The RegulonDB was used to obtain the known *E.coli* operon structure set (Huerta *et al*., 1998). Amino acids present on the surface of 10% least- and 10% most aggregation-prone proteins from the dataset (proteins that do not possess transmembrane segments) were identified by A3D, and the frequency of appearance of each individual amino acid among the total set of surface amino acids belonging to the two mentioned groups was calculated. Amino acid frequencies for aggregation-prone and soluble *E.coli* proteins were compared using Wilcoxon test. The information about secondary structure content was obtained from the PDB files for the analysed proteins. The information about the protein synthesis rate was obtained from the ribosome profiling experimental data (Li *et al*., 2014) and the obtained values were then $\log_{10}$ transformed for statistical analysis. The reported experimentally determined melting temperatures (Tm), indicating protein stability (Leuenberger *et al*., 2017) were used to compare the stability of the aggregation-prone and soluble proteins. The Pfam database (Finn *et al*., 2014) was used to identify ortholog and paralog candidates for the 10% most aggregation-prone and 10% most soluble proteins from the dataset according to A3D, together with the Clustal Omega algorithm (Sievers and Higgins, 2014) to determine the percentage of sequence identity for orthologs and paralogs.

Definition of the supersaturation index (SSI)

The structural supersaturation index (SSI), calculated only for proteins without transmembrane domains was calculated as the sum:

$$SSI = \frac{(C + A)}{2}$$

where C represents normalised protein concentration and A is the normalised structural aggregation propensity score.

The logarithm of the protein abundance levels derived from PaxDb (Wang *et al*., 2012) was used in order to determine C of each polypeptide from the dataset. These values were then normalised by rescaling them between 0 and 1.

$$C = \frac{(Ci - \min(Ci \ldots Cn))}{(\max(Ci \ldots Cn) - \min(Ci \ldots Cn))}$$

Cmin = the minimum value of protein concentration from the dataset
Cmax = the maximum value of protein concentration from the dataset

The propensity of proteins to aggregate represented by A3D average score was normalised in the same manner by rescaling the values between 0 and 1.

$$A = \frac{(Ai - \min(Ai \ldots An))}{(\max(Ai \ldots An) - \min(Ai \ldots An))}$$

Amin = the minimum A3D score from the dataset
Amax = the maximum A3D score from the dataset

Statistical analysis

Statistical significance was determined using Wicoxon test. P values <0.05 were considered statistically significant. For Venn Diagrams, p values were calculated using hypergeometric test to determine whether there is a significant overlap and the complete set of bacterial proteins was used as a background.

4. Protein aggregation into insoluble deposits protects from oxidative stress

4.1. Introduction

Protein aggregation and the long term accumulation of amyloid protein inclusions (PI) in the central nervous system are pathognomonic signs of devastating neurodegenerative disorders like Alzheimer's (AD), Parkinson's (PD) and Huntington's (HD) diseases (Dobson, 2002; Aguzzi and O'Connor, 2010). In spite of this, it is becoming evident that the coalescence of proteins and other molecules into dynamic liquid-like assemblies is essential for an increasing number of cellular processes (King *et al*., 2012; Mitrea and Kriwacki, 2016). In contrast, the beneficial or detrimental effects associated with the formation of insoluble protein aggregates are still under debate.

Cells have evolved two main mechanisms against protein aggregation: (i) refolding of misfolded conformers by chaperones and (ii) targeted destruction of damaged proteins by the ubiquitin-proteasome and/or autophagy systems (Frydman, 2001; Hartl *et al*., 2011; Wong *et al*., 2012). The presence of PI in neurodegenerative diseases has been traditionally assumed to reflect the insufficiency of this first line of defence (Ross and Pickart, 2004). However, it was proposed later that the formation of PI can instead act as a protective strategy (Arrasate *et al*., 2004; Ross and Poirier, 2005), triggering an intense dispute on the specific protein species responsible for the cellular damage (Evangelisti *et al*., 2016; Forloni *et al*., 2016). It has been suggested that earlier aggregation intermediates, such as oligomeric and/or protofibrillar assemblies, would disrupt the cellular homeostasis by promoting aberrant interactions with different cellular components (Deas *et al*., 2016), including the protein quality surveillance machinery, whereas the formation of inclusions at specific cellular sites might be indeed a detoxifying mechanism against the presence of these promiscuous species (Ross and Poirier, 2005; Cohen *et al*., 2006). In agreement, studies in AD, PD and HD models report that the presence (Ross and Poirier, 2005; Sydow *et al*., 2011).

Oxidative stress plays a central role in the pathophysiology of the different neurodegenerative diseases linked to protein aggregation (Facchinetti *et al*., 1998). Cumulative oxidative stress induces membrane damage, impairment of the DNA repair system and mitochondrial dysfunction, which ultimately can lead to cell death (de Groot and Burgas, 2015; Wang *et al*., 2014). Accordingly, there is clear evidence of oxidative damage in post-mortem AD, PD and HD brains (Ansari and Scheff, 2010; Tabner *et al*., 2005). The mechanism by which aggregating proteins may induce oxidative stress in the brain is not completely understood, but the interaction between these two phenomena is bidirectional, in such a way that aggregation-prone conformers increase the production of reactive oxygen species (ROS), while at the same time oxidative stress exacerbates protein aggregation (Tabner *et al*., 2005). Again, there is discrepancy on the nature of the aggregating species causing the oxidative damage, some reports attributing this property to early oligomers (Deas *et al*., 2016) and others indicating that PI constitute centres of oxidative events, questioning their cytoprotective role (Firdaus *et al.*, 2006).

The use of simple cellular models has led to great advances in the study of neurodegenerative disorders. The budding yeast Saccharomyces cerevisiae shares with higher eukaryotes numerous fundamental cellular pathways involved in neurodegeneration, such as protein quality control, autophagy or the regulation of the cellular redox state regulation (Khurana and Lindquist, 2010; Tenreiro and Outeiro, 2010). Thanks to this, humanized yeast models for AD, PD and HD have been successfully developed, recapitulating some of the pathological features associated with these disorders, including oxidative stress (Treusch *et al*., 2011; Cooper *et al*., 2006; Willingham *et al*., 2003). We have previously exploited yeast to dissect the relationship between proteins aggregation *in vivo* and their physicochemical properties. For this purpose, we designed and expressed intracellularly 20 different variants of the amyloid-β-peptide (Aβ42) fused to GFP (Villar-Pique and Ventura, 2013). Later, we investigated the proteomic response caused by the expression of two individual proteins from this collection displaying different *in vivo* aggregation properties, one of them remaining diffusely distributed in the cytosol and the other one forming large PI in yeast cells. The main differences between the proteomic response elicited by these two Aβ42 variants were related to redox homeostasis (de Groot *et al*., 2015). This anticipated that the complete Aβ42-GFP set might be employed as a tool to dissect quantitatively the contribution of different aggregating species to oxidative damage. To this aim, in the present study we performed a systematic investigation of the relationship between the intrinsic aggregation propensities of 21 Aβ42-GFP variants, their ability to be recruited into PI and the induced intracellular oxidative stress. With these analyses, we demonstrate the existence of a conspicuous inverse relationship between Aβ42 peptides aggregation propensities and oxidative stress levels, distinguishing peptide variants unable to be recruited into inclusions as those causing the highest levels of oxidative damage. In this way, we identified as the elicitors of the oxidative insult those species that despite being in an aggregated state remain diffusely distributed in the cytosol, available for establishing disrupting non-functional interactions. Overall, our data provide strong support for the detoxifying role against oxidative stress played by PI.

4.2. Results and discussion

4.2.1. The Aβ42-GFP protein collection

In the present study we used a set of 20 Aβ42-GFP variants that correspond to all possible mutations for residue 19 in the Aβ42 peptide (Treusch *et al*., 2011). This residue is located in the central hydrophobic cluster (CHC) of the peptide and has an important influence in the peptide's folding, self-assembly and structure in the fibrillar state (Bitan *et al*., 2003; Morimoto *et al.*, 2004). Predictions of the aggregation propensities of the resulting sequences using the AGGRESCAN algorithm (de Groot *et al.*, 2012) anticipated that this protein dataset would display a continuum of aggregation propensities (Fig. 4.1). Analysis with predictors based on different physicochemical and structural protein properties such as TANGO (Fernandez-Escamilla *et al.*, 2004) and Foldamyloid (Garbuzynskiy *et al.*, 2010) rendered very similar aggregation gradation, ascribing in all cases the highest aggregation propensities to the natural Aβ42 sequence (Phe19) and those variants in which this residue is substituted by other aromatic or aliphatic residues (Fig. 4.1).

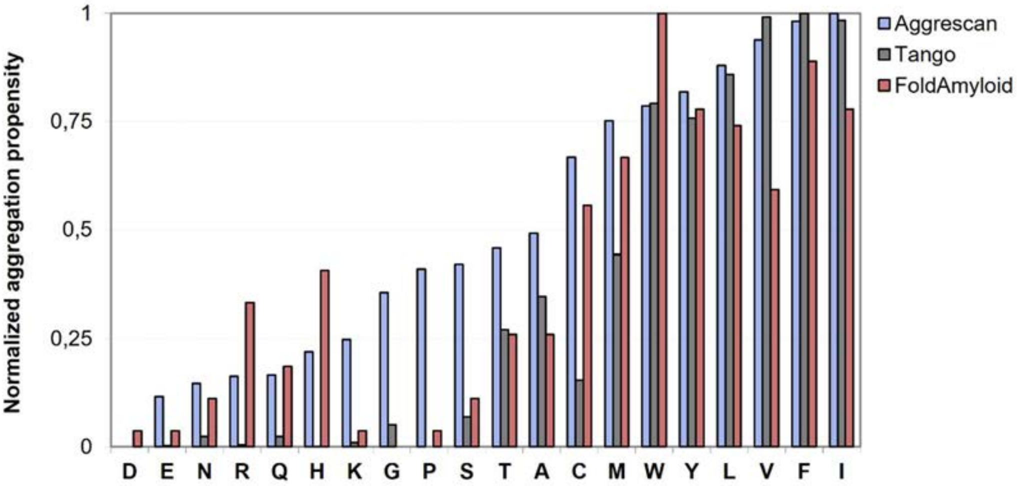


Figure 4.1. Comparison of the predicted aggregation propensities of 20 Aβ42-GFP variants. Bar graphs represent the aggregation propensities obtained by three bioinformatic predictors: AGGRESCAN, TANGO and FoldAmyloid. Values have been normalized for each predictor. The variants represented on the x-axis have been ordered according to their AGGRESCAN predicted aggregation propensities.

4.2.2. PI formation propensity of Aβ42-GFP variants

The 20 Aβ–GFP proteins were expressed individually in yeast under the control of the galactose-inducible GAL1 promoter to allow a rapid, strong and synchronous induction of protein expression. All yeast cells exhibited fluorescence 16 h after induction of protein expression. However, the formation of PI was observed only for 10 out of the 20 assayed variants. Representative images of selected PI-forming and PI–non-forming variants are shown in Fig. 4.2A. The 10 PI-forming variants correspond precisely with those proteins displaying the highest aggregation propensity in the AGGRESCAN scale (Fig. 4.2B). This points to a defined protein aggregation propensity threshold governing the PI formation in *S.*

cerevisiae. In the conditions of our assay, this limit corresponds to ∼ 45% of the aggregation potential of wild type Aβ.

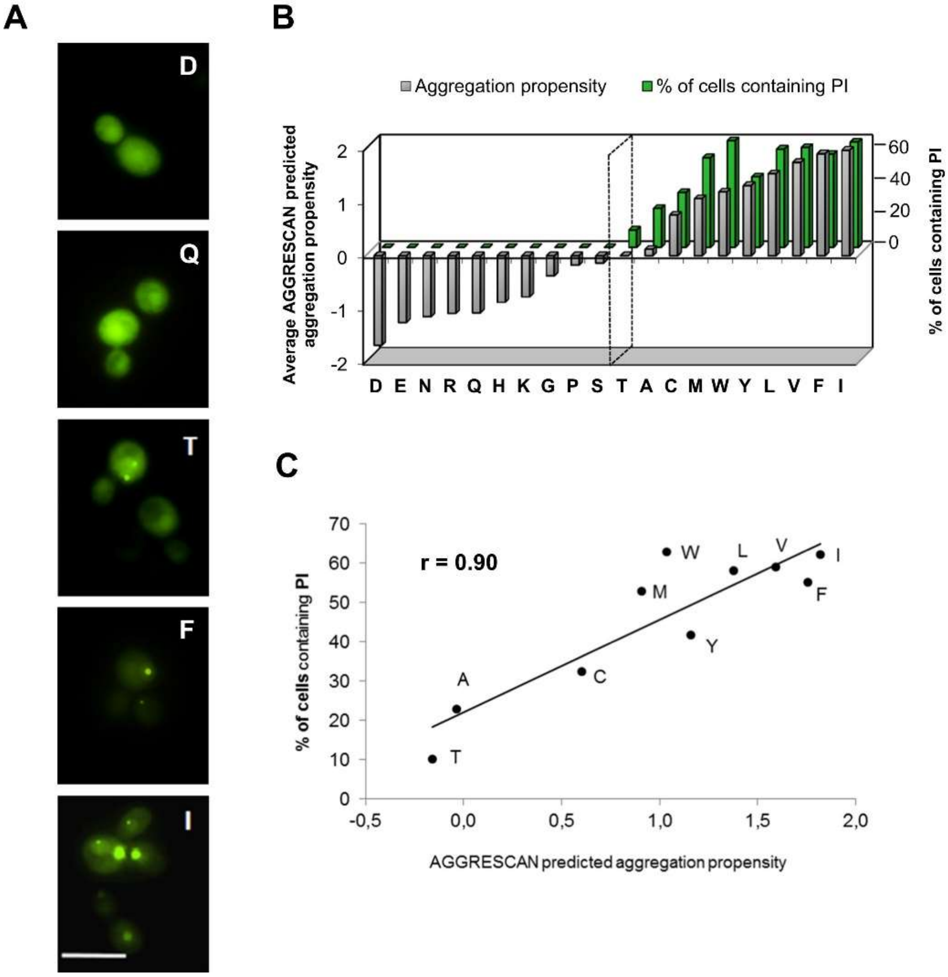


Figure 4.2. GFP fluorescence determined in *S. cerevisiae* cultures expressing Aβ42-GFP mutants for 16 h. (A) Representative fluorescence microscopy images of yeast cells expressing selected Aβ42-GFP variants (Asp, Gln, Thr, Phe and Ile). Scale bar represents 10 μm. (B) 3D bar graph representing the relative aggregation propensities of 20 natural amino acids, derived from the analysis of mutants in central position of the central hydrophobic cluster in amyloid-β-protein previously described by de Groot et al. (de Groot *et al.*, 2006), represented on the primary y-axis, together with the percentage of cells containing intracellular visible fluorescent protein inclusions (PI) for Aβ42-GFP variants represented on the secondary y-axis. The dashed black line represents defined threshold of PI formation. Values were obtained by determining the number of cells with PI among the ~ 500 cells from two different cultures for each variant. (C) Correlation between the percentage of cells containing visible intracellular PI and the intrinsic aggregation propensity predicted by AGGRESCAN (r = 0.90, p ≤ 0.0001).

Below this threshold, and independently of the considered mutant, none of the cells in the population forms PI, whereas above this value the proportion of cells displaying PI sharply depends on the identity of the mutant, varying from 10% for Thr to >60% for the wild type or the Ile mutant. Then by focussing on the set of PI forming mutants, we observed that the fraction of cells containing PI and the predicted aggregation propensities according to AGGRESCAN significantly correlate ($r = 0.90$, $p \leq 0.0001$) (Fig. 4.2C). The slope of the regression line (0.85) indicates that, above the aggregation threshold, a 10% increase of protein aggregation propensity of the peptide results in an 8.5% increase of the proportion of cells displaying protein inclusions.

4.2.3. The oxidative stress induced by Aβ42-GFP variants has a negative correlation with their aggregation propensity

In order to decipher whether the formation of intracellular PI causes or protects from oxidative stress, we selected three different Aβ–GFP variants for further study: (i) Ile as a representative of high PI formation, (ii) Thr as a representative of low PI formation and (iii) Gln as a variant that does not generate PI. In Aβ–GFP fusions, the GFP fluorescence acts as a reporter of the aggregation propensity, in such a way that more fluorescence indicates less aggregation and vice versa (de Groot *et al.*, 2006). In agreement with the proportion of PI forming cells, the median of the population fluorescence measured by flow cytometry (FC) indicates that Gln is the most fluorescent variant, Ile is the less fluorescent and Thr has a fluorescence intensity in between the other two (Fig. 4.3A).

Then we monitored the oxidative stress caused by the expression of these variants by measuring four different parameters: ROS production, H_2O_2 levels, catalase activity and protein carbonylation. To monitor the ROS production we analysed the cells expressing the three Aβ–GFP variants with CellROX fluorogenic probe and FC. The Gln variant exhibited the highest levels of ROS, followed by Thr and Ile (Fig. 4.3B). Therefore, the more soluble variant (Gln) caused the highest level of oxidative stress and the most aggregation-prone mutant (Ile) turned out to be the most innocuous. These results were confirmed by the levels of H_2O_2 (Fig. 4.3C), again Gln exhibited the highest levels of oxidative stress and Ile the lowest, with Thr displaying intermediate H_2O_2 levels. In agreement, the enzyme catalase, which protects against ROS, showed higher activity in yeast strains expressing the Gln variant than in those expressing the Ile one (Fig. 4.3D). Protein carbonyl formation is a marker for protein oxidation and can be detected by DNPH derivatization followed by western blot with anti-DNP antibodies (Sorolla *et al.*, 2011). We fractioned yeast cells and analysed the level of carbonylation in both the soluble and insoluble protein fractions by western blot. Whereas we did not observe any significant differences between soluble fractions, the level of protein oxidation was clearly higher in the insoluble fraction of cells expressing the Gln variant than in those expressing the Ile one (Fig. 4.3E).

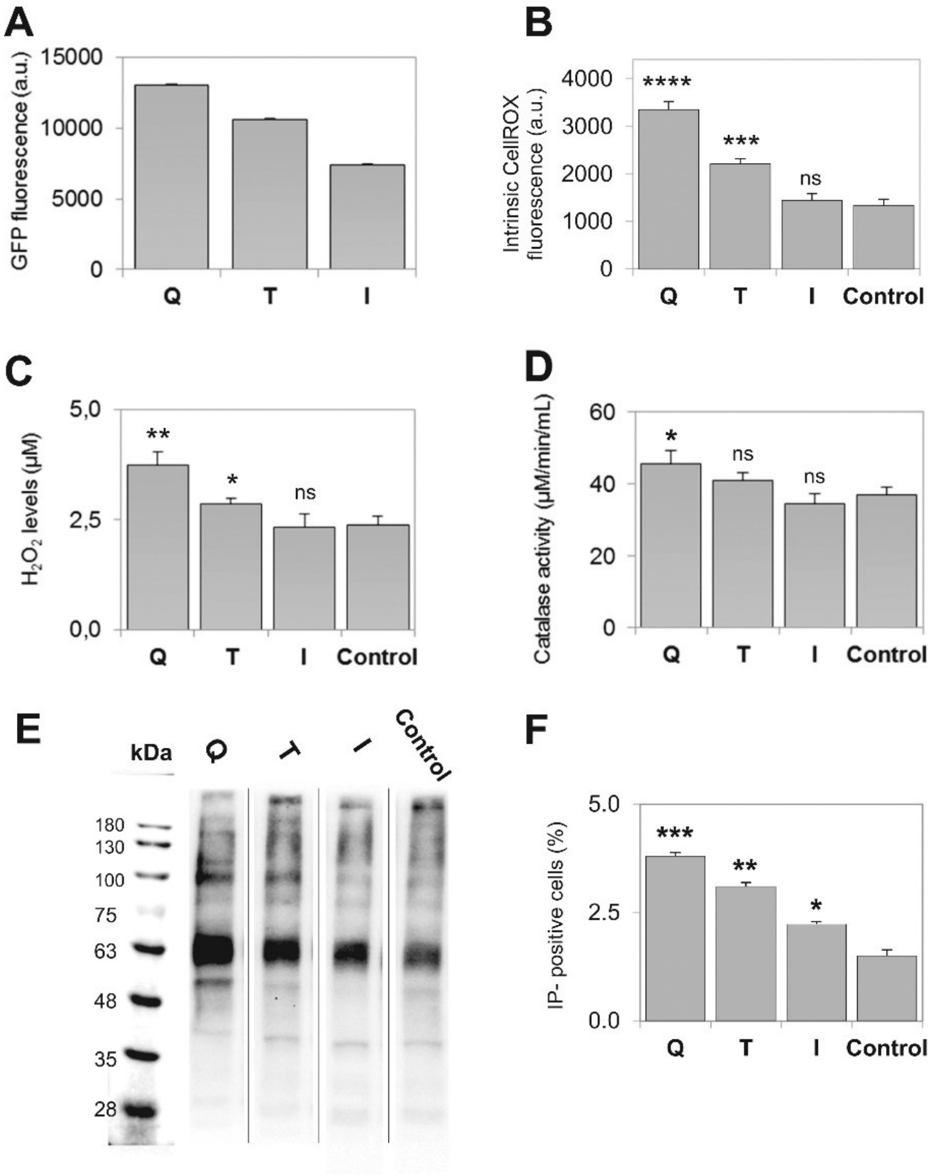


Figure 4.3. Analysis of *S. cerevisiae* cultures expressing Gln, Thr and Ile mutants for 16 h. (A) Median fluorescence values of expressed GFP from 10,000 cells determined by FC analysis (FITC-A) for selected Aβ42-GFP variants. (B) CellROX median fluorescence intensity (APC-A) (as indicator of oxidative stress levels) determined by FC for selected Aβ42-GFP variants. Error bars represent CV of the FC gated cells (P3 population). (C) H_2O_2 levels determined by monitoring the fluorescence at 595 nm for selected Aβ42-GFP variants. Error bars represent ±SE (n=3). (D) Catalase activity determined by absorbance measures at 550 nm for selected Aβ42-GFP variants. Error bars represent ±SE (n=3). (E) Immunodetection of protein carbonylation levels of insoluble cell fractions for selected Aβ42-GFP variants. All samples were run in the same gel. The places where the membrane was trimmed are indicated by a thin line. The normalised densities of the bands are: 135, 124, 88 and 100 a.u. for variants ordered from left to right, respectively. (F) The percentage of IP positive cells (as indicator of cell mortality) was obtained by FC analysis (PerCP-Cy5-A) for selected variants. Error bars represent ±SE (n=3). Cells expressing GFP alone were used as control.

In order to assess whether the recorded levels of oxidative stress may impact cell fitness, we analysed the viability of cells expressing the different variants by FC using propidium iodide (IP) staining. IP is a DNA-binding dye that is excluded by live cells, but enters dead cells with damaged membranes. In good agreement with their relative ROS levels, the viability of the cell population was more compromised for the Gln variant than for Ile (Fig. 4.3F).

The above described results suggest that PI formation is not the main elicitor of cellular oxidative stress. To confirm this, we used confocal fluorescence microscopy and dihydroethidium fluorescence, a marker of intracellular oxidative stress, to detect whether the cellular distribution of Gln and Ile protein variants correspond with sites of ROS production. In good agreement with the above described data, the levels of ROS are higher in the more soluble Gln variant (Fig. 4.4). In both cases, the oxidative stress is distributed in the cytosol and does not co-localise with IP, suggesting that they are not centres of oxidative events as proposed by other studies (Cohen *et al*., 2006).

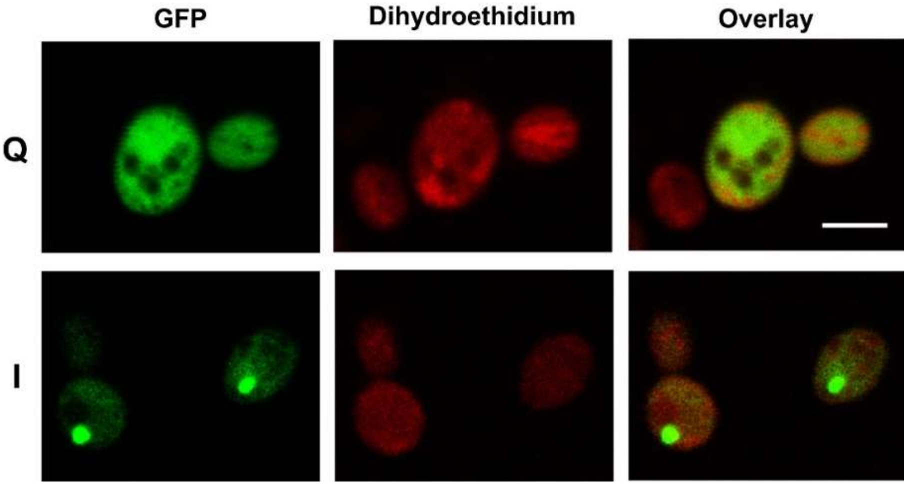


Figure 4.4. Characterisation of GFP fluorescence and oxidative stress levels for Gln and Ile variant. Representative confocal microscopy images of yeast cells expressing Gln and Ile variants showing the levels of GFP fluorescence (*left*), cellular oxidative stress (*middle*) and overlay image (*right*). Induced cultures were incubated with 10 µM dihydroethidium prior to image acquisition. Scale bar represents 5 µm.

To address the relationship between protein aggregation properties and oxidative stress in a more quantitative manner, we monitored the levels of ROS for the rest of mutants belonging to the Aβ–GFP collection using CellROX and FC. Then, we compared the levels of ROS with the aggregation propensity predicted by three different algorithms (Fig. 4.5A, Fig. 4.6A and Fig. 4.6B) and with all of them we obtained a significant negative correlation (AGGRESCAN: $r = 0.91$, $p \leq 0.0001$; TANGO: $r = 0.82$, $p \leq 0.0001$; FoldAmyloid: $r = 0.84$, $p \leq 0.0001$).

The aggregation propensity of a polypeptide sequence in a defined environment depends on its physicochemical properties (Chiti *et al*., 2003). In this context, it has been found that β-sheet propensity and hydrophobicity favour the self-assembly of proteins into aggregates. Consistently, our results show that these two protein properties correlate negatively with oxidative stress levels (β-sheet propensity: r = 0.72, $p \leq 0.0001$, Fig. 4.5B; hydrophobicity: r = 0.78, $p \leq 0.0001$, Fig. 4.5C).

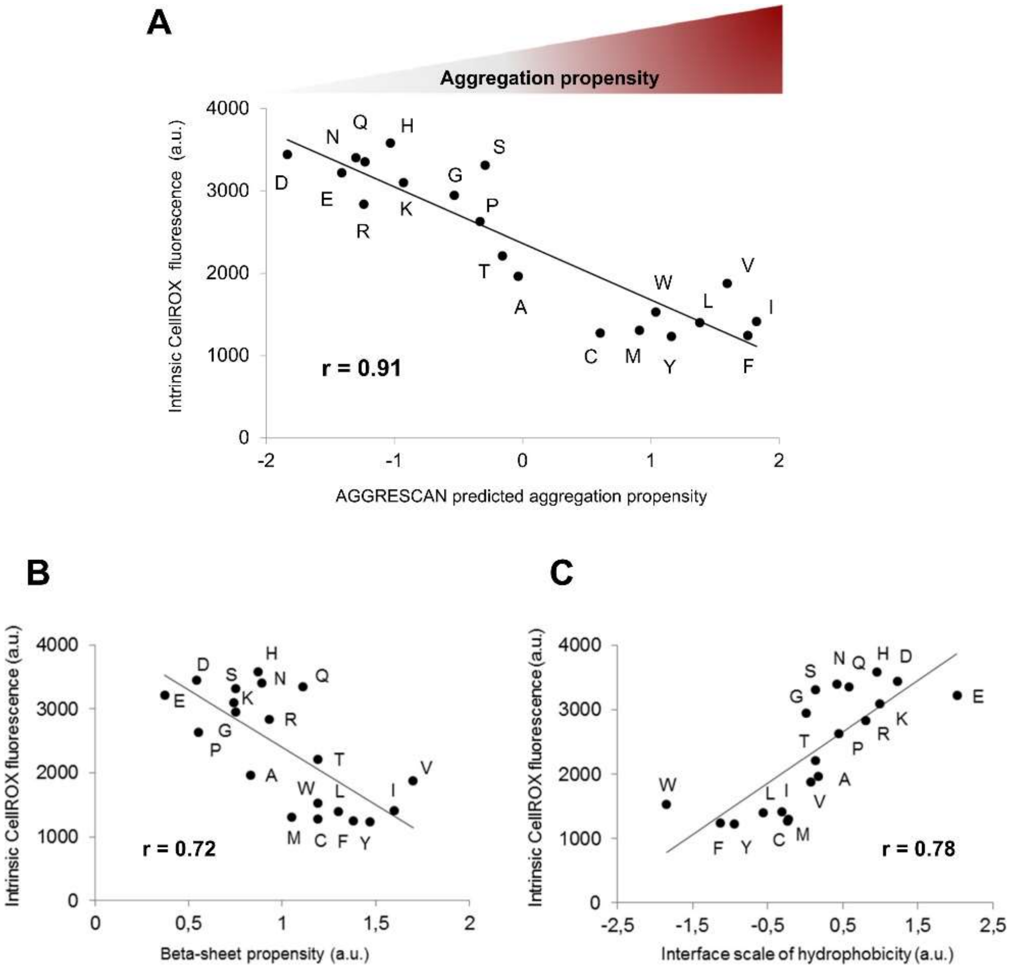


Figure 4.5. Relationship between intrinsic oxidative stress levels of 20 Aβ42-GFP mutants with AGGRESCAN predicted aggregation propensities and amino acid physicochemical properties. Correlation between CellROX intrinsic fluorescence and: (A) intrinsic aggregation propensity predicted by AGGRESCAN (r = 0.91, $p \leq 0.0001$); (B) β-sheet propensity from Chou and Fasman scale (r = 0.72, $p \leq 0.0001$) and (C) hydrophobicity derived from the interface scale (r = 0.78, $p \leq 0.0001$).

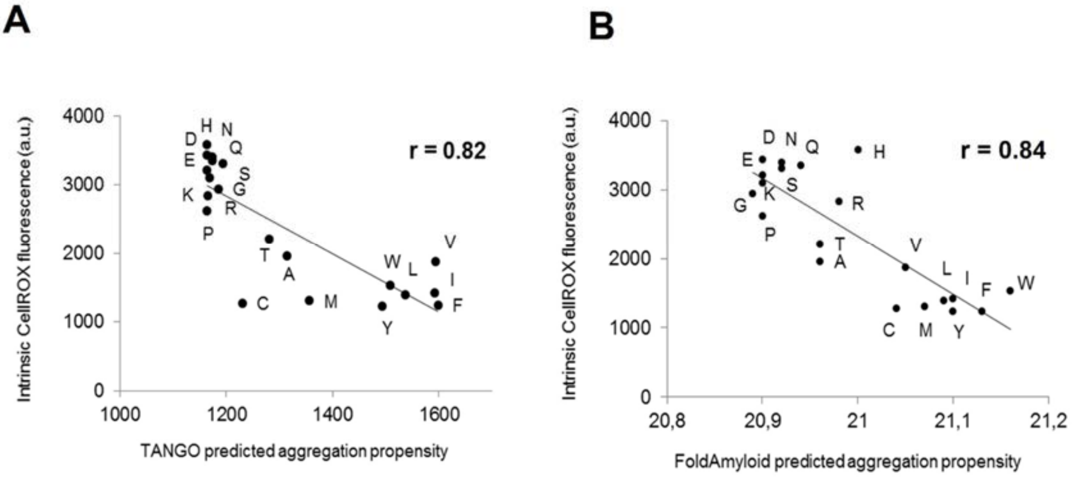


Figure 4.6. Correlation between the intrinsic oxidative stress levels and alternative prediction algorithms: (A) TANGO, with a correlation coefficient r = 0.82, p ≤ 0.0001; (B) FoldAmyloid, with a correlation coefficient r = 0.84, p ≤ 0.0001.

4.2.4. Both PI formation and aggregation-prone protein clearance prevent oxidative stress

The comparison between GFP fluorescence, an indication of aggregation, and the oxidative stress shows a significant positive relationship (r = 0.68, p ≤ 0.0001) though the correlation coefficient is lower than the one observed for predicted aggregation propensities (Fig. 4.6A). In this distribution, Trp and Phe are clear outliers, essentially because they display lower fluorescence than expected. This matches with the smaller and less fluorescent PI they exhibit, in comparison with the other aggregation-prone mutants. In addition, we have previously shown that these observations are linked to the fact that Trp and Phe mutants are recognised by the protein quality machinery and degraded through autophagy (Villar-Pique and Ventura, 2013), resulting in very low protein levels (Fig. 4.8A). All these observations suggest that the degradation of an aggregation-prone variant or sequestering it in PI may result in a similar prevention of oxidative stress. However, Trp and Phe are not the only responsible for the observed lower correlation, since removing them from the dataset only increases the correlation coefficient to r = 0.72 (p ≤ 0.0001). We must consider the observation that PI contribute much more to GFP fluorescence than they do to oxidative stress levels (Fig. 4.4). Thus, taking only into consideration the fluorescence of non-PI forming cells, the correlation coefficient raises to r = 0.87 (p ≤ 0.0001; Fig. 4.7B), supporting that PI have a low impact on the generated oxidative stress levels.

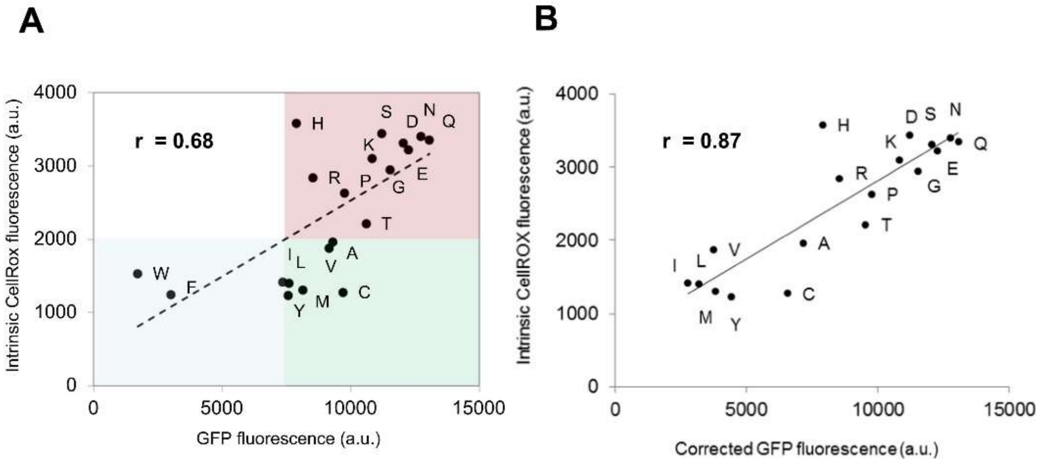


Figure 4.7. Correlation between GFP fluorescence and intrinsic levels of oxidative stress of the expressed 20 mutants from Aβ42-GFP set. (A) Spatial distribution of Aβ42-GFP mutants depending on GFP fluorescence *vs*. intrinsic levels of oxidative stress is represented. In blue: mutants with low levels of oxidative stress presenting low levels of GFP fluorescence. In green: mutants with low levels of oxidative stress presenting high levels of GFP fluorescence. In red: mutants with high levels of oxidative stress presenting high levels of GFP fluorescence (r = 0.68, p ≤ 0.0001). (B) Correlation between corrected GFP fluorescence, after removing the variants that behave as outliers from the dataset (Phe and Trp), and intrinsic levels of oxidative stress for mutants expressing Aβ42-GFP, (r = 0.87, p ≤ 0.0001).

4.2.5. Aggregated but diffusely distributed protein species are the elicitors of oxidative damage

To further study the link between cellular oxidative stress and the soluble/aggregated state of the Aβ42-GFP variants, we measured their distribution between the soluble and insoluble cell fractions using western blotting (Fig. 4.8A). Irrespective of whether the proteins can form PI, the 20 variants were found to be mainly located in the insoluble fraction, consistently with previous data obtained by our group and others (Villar-Pique and Ventura, 2013; Bharadwaj *et al*., 2012). However, the levels of soluble protein are higher for polar and charged residues compared to the hydrophobic ones (Fig. 4.8B). The fact that mutants diffusely distributed in the cytosol displayed levels of insoluble protein equivalent or even higher than those forming PI was surprising (Fig. 4.8B). This argues that the insoluble fraction might contain at least two types of aggregates, one corresponding to PI and the other to proteins that, despite being aggregated, do not coalesce in visible inclusions, remaining homogenously distributed through the cytosol; we will call this last type diffuse aggregates.

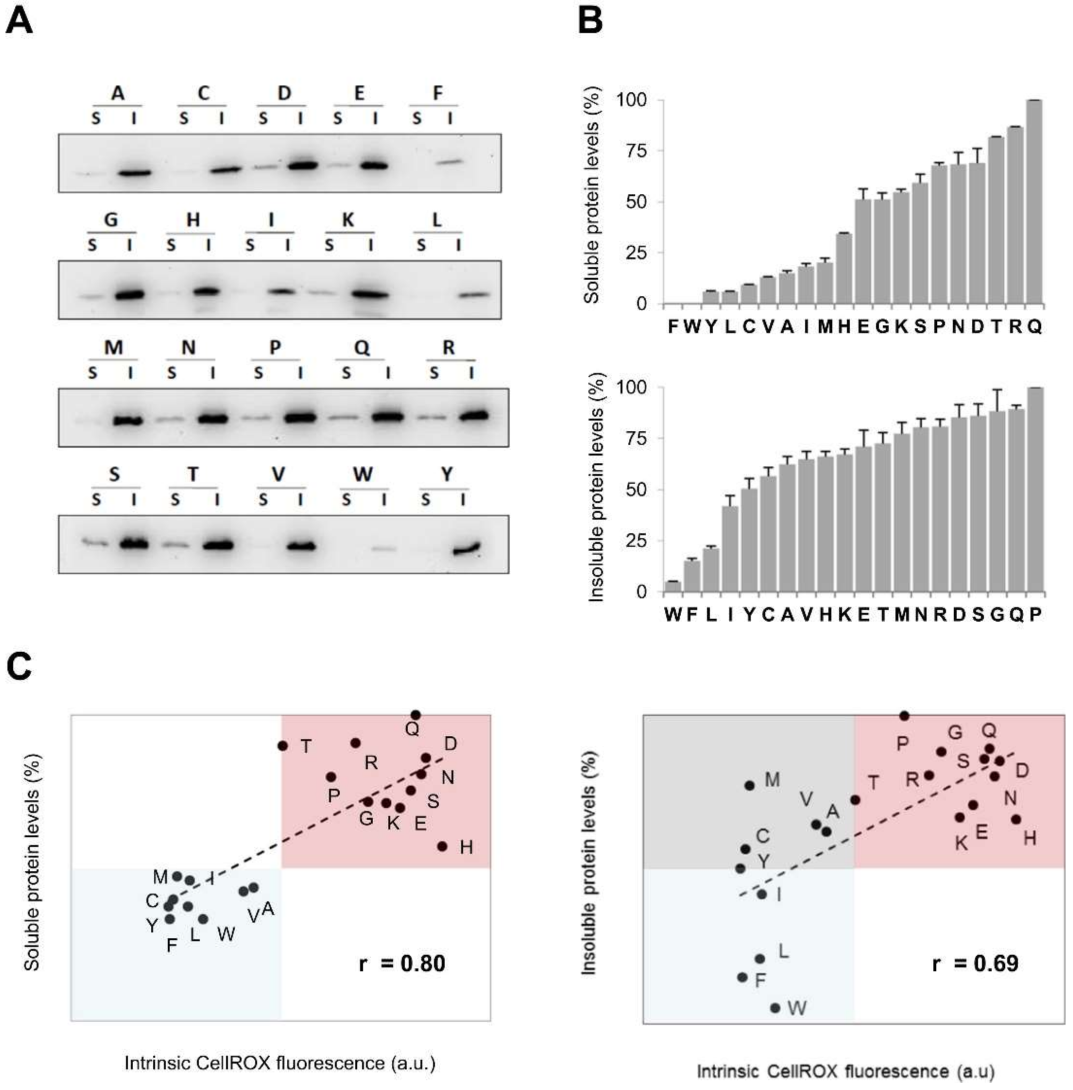


Figure 4.8. Characterisation of 20 Aβ42-GFP mutants expression after 16 h. (A) Immunodetection of soluble (S) and insoluble (I) cell fractions of the 20 Aβ42-GFP variants. Detected bands correspond to MW = 35 kDa. (B) Bar graph with the percentage values of the soluble and insoluble protein fractions quantified using Quantity One software. Error bars represent ±SE (n=3). (C) Relation between the intrinsic levels of oxidative stress *vs*. soluble and insoluble protein fractions of Aβ42-GFP variants. The spatial distribution of Aβ42-GFP mutants depending on CellROX intrinsic fluorescence and the corresponding protein levels: (blue) mutants with low levels of corresponding protein fraction of Aβ42-GFP displaying low oxidative stress levels; (red) mutants with high levels of corresponding protein fraction of Aβ42-GFP presenting high oxidative stress levels and; (grey) mutants with high levels of corresponding protein fraction of Aβ42-GFP showing low oxidative stress levels. Correlation index for soluble protein levels: r = 0.80, $p \leq 0.0001$; for insoluble protein levels: r = 0.69, $p \leq 0.0001$.

With the aim to localize the protein species generating the oxidative stress, we analysed whether the measured CellROX fluorescence correlates with the detected protein levels in either the soluble or the insoluble fraction. The correlation with the levels of soluble fraction was higher than with the levels of insoluble fraction (Fig. 4.8C). From these results we discarded PI as centres of oxidative stress and considered two possible interpretations: (i) the soluble species themselves correspond to the reactive forms, or (ii) the diffuse aggregates are the reactive forms and the levels of soluble protein are just a reflection of the levels of these disseminated assemblies in the insoluble fraction.

In order to discriminate between these two possibilities, we sought to design a novel Aβ42 variant with enhanced solubility, in such a way that the amount of protein in the soluble fraction would be higher than the amount in the insoluble one, while keeping sequence changes to a minimum. If the soluble species are the ROS promoting forms, we expected oxidative stress levels being higher in this mutant than in any other variant. Different aggregation prediction algorithms coincide to indicate that Aβ42 peptide consists of two main aggregation-prone stretches, comprising residues 17–21 and 30–41 (Fig. 4.9A). In the solid-state NMR, the structure of Aβ amyloid fibrils demonstrates that the residue Phe19 in the first β-strand interacts with Leu34 in the second β-strand (Lührs *et al*., 2005). Based on these data, we introduced in one of the most soluble and fluorescent variants (Aβ42-F19D) a second mutation by changing the original Leu at position 34 to Pro (a β-breaking residue), a combination (19Asp/34Leu) that we have already shown to be highly soluble in bacteria (Navarro *et al*., 2014).

In good agreement with the design, 16 h after induction, cells expressing F19D/L34P variant exhibit a highly fluorescent cytosol devoid of any PI (Fig. 4.9B) Fractionation of cell content and immunoblotting indicated that, as intended, in 19Asp/34Leu the majority of the protein was now located in the soluble fraction (Fig. 4.9C). Quantitative analysis of the GFP fluorescence of cells expressing the double mutant using FC indicates that this mutant is by far the most fluorescent variant in the protein set (Fig. 4.9E). Despite these properties, the levels of protein carbonylation (Fig. 4.9D) and oxidative stress (Fig. 4.9F) that it induced were moderate, which discards the possibility of soluble species being the main elicitors of oxidative damage. To asses if this effect is caused instead by diffuse aggregates, we took profit of the kinetics of protein aggregation. Therefore, we expressed the double mutant for 48 h. At this time point, even if the fluorescence remains diffusely distributed in the cytosol (Fig. 4.9B), the large majority of the protein is now located in the insoluble fraction (Fig. 4.9C).

Thus, in these conditions, the double mutant resembles Gln variant after 16 h of induction of protein expression. We also expressed Gln, Thr and Ile mutants for 48 h and monitored the distribution of GFP fluorescence in the cytosol at this time point (Fig. 4.10A). A proportion of Gln expressing cells begin to display PI, the inclusions of Thr become similar to those of Ile at the 16 h, and in yeast cells expressing Ile variant all the fluorescence is concentrated in PI at 48 h. Comparison of the relative levels of oxidative stress for the four mutants at 16 h and 48 h time points, as determined by using FC (Fig. 4.10B), allows to visualise that the highest oxidative stress is attained when the protein is diffusely distributed in the cytosol but located in the insoluble fraction like in the case of Gln at 16 h or the double mutant at 48 h, whereas the formation of PI inclusions results in reduced levels of oxidative stress. After 48 h, cells expressing the double mutant display slightly lower viability than the other variants (Fig. 4.10C), whereas at 16 h it is the Gln variant that induces the highest mortality (Fig. 4.10C).

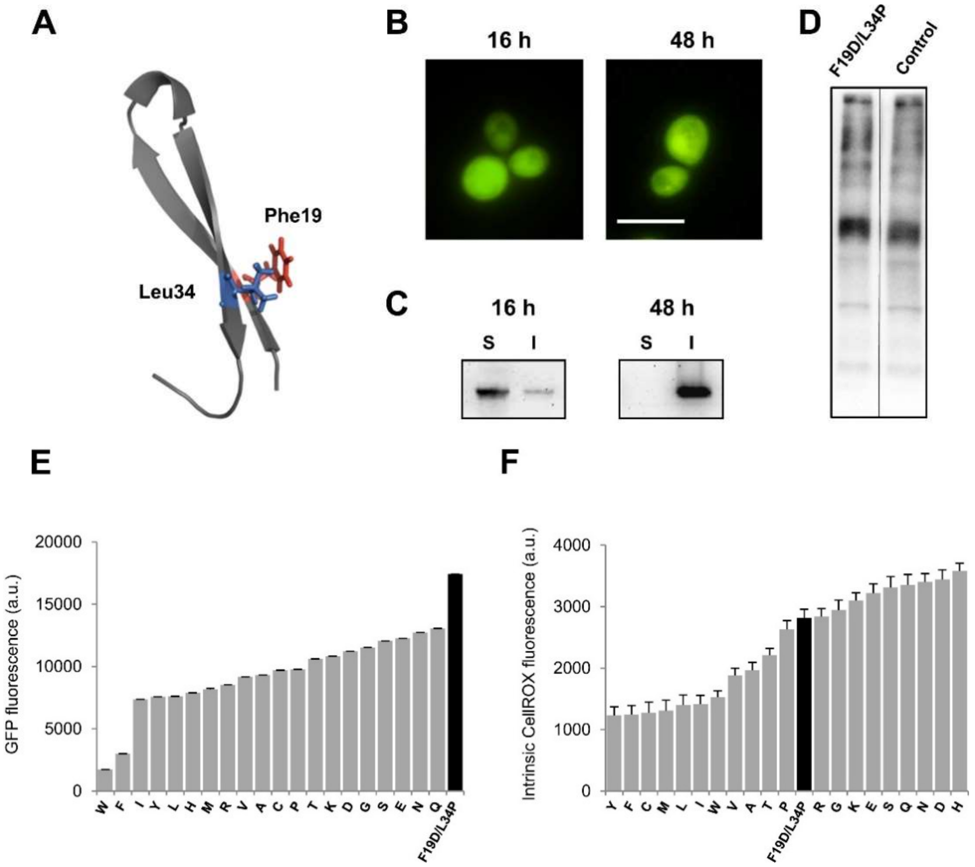


Figure 4.9. Characterisation of Aβ42-F19D/L34P variant. (A) Structure of Aβ42 peptide (PDB: 2OTK) showing the side chains of double mutant, F19 in red and L34 in blue. (B) Representative fluorescence microscopy images of yeast cells expressing Aβ42-F19D/L34P-GFP variant for 16 and 48 h showing GFP fluorescence. Scale bar represents 10 µm. (C) Western blotting of the soluble (S) and insoluble (I) cell fractions of the mentioned variant after 16 and 48 h of expression. Detected bands correspond to MW = 35 kDa. (D) Immunodetection of protein carbonylation levels of insoluble cell fractions for Aβ42-F19D/L34P and pESC-URA-GFP. The normalised densities of the bands are: 103 for Aβ42-F19D/L34P and 100 a.u. for pESC-URA-GFP. Both samples were run in the same gel. The place where the membrane was trimmed is indicated by a thin line. (E) Bar graph representing the median GFP fluorescence (FITC-A channel) of P2 population, for each expressed Aβ42-GFP variant, together with Aβ42-F19D/L34P, acquired by FC. (F) Bar graph representing median fluorescence intensity of CellROX probe (APC-A channel) (levels of oxidative stress) of P3 population determined by FC. Error bars represent CV of the FC gated cells.

To further understand the properties of these diffuse aggregates we addressed whether they might display a compact structure by monitoring their accessibility to proteinase K (PK) (Fig. 4.10D). Despite its high activity for cleaving peptide bonds, PK cannot attack the highly packed backbones in an amyloid β-sheet structure. The insoluble fractions of cells expressing Ile, Thr, Gln and 19Asp/34Leu variants were treated with PK up to 10 min.

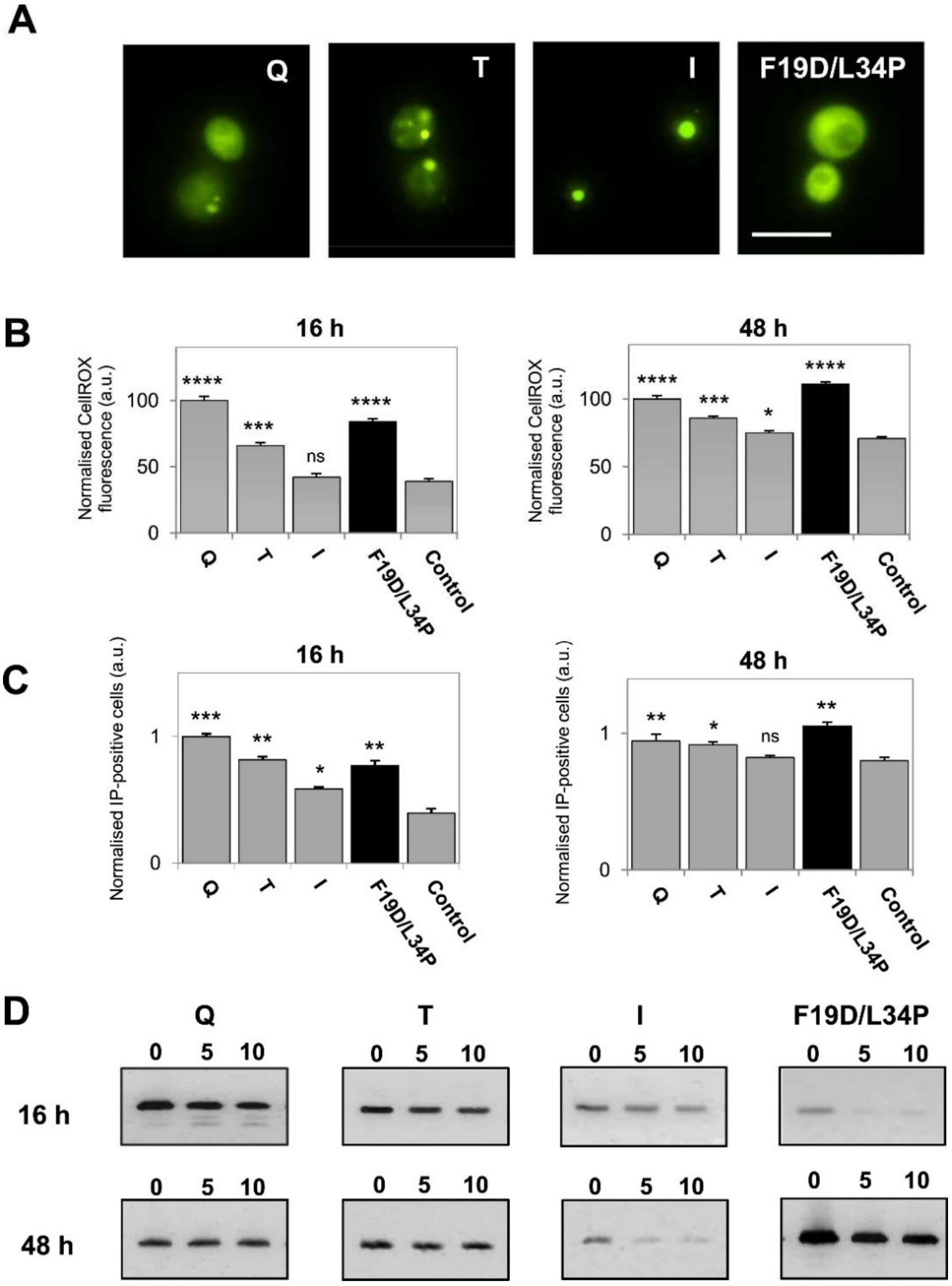


Figure 4.10. Comparison between GFP fluorescence, oxidative stress levels, cell viability and proteinase K (PK) resistance for Aβ42-GFP mutants after 16 and 48 h of expression. (A) Representative fluorescence microscopy images showing GFP signal of yeast cells expressing selected Aβ42-GFP variants (Gln, Thr, Ile and F19D/L34P) for 48 h. Scale bar represents 10 µm. (B) CellROX median fluorescence bar graph of yeast cultures expressing selected Aβ42-GFP variants induced for 16 and 48 h determined by FC analysis. Error bars represent CV from 10,000 events. (C) Percentage of IP positive cells in cultures expressing selected Aβ42-GFP variants induced for indicated times. Error bars represent ±SE (n=3). (D) Time-lapse PK resistance of insoluble fractions of selected Aβ42-GFP variants, induced for 16 and 48 h, detected by western blotting. Detected bands correspond to MW = 35 kDa. Cells expressing GFP alone were used as control.

Interestingly, at 16 h the aggregated fraction of Gln cells was essentially resistant to proteolysis. The same was true for the insoluble fraction of 19Asp/34Leu after 48 h, which indicates that diffusely distributed aggregates already possess an amyloid-like conformation able to elicit the oxidative stress response. The lower levels of insoluble Ile protein at 48 h suggest that this mutant resembles now Trp or Phe at the 16 h, becoming a substrate for active degradation.

The above observed influence of the aggregation kinetics on ROS promoted levels, predicts that at early times, when aggregation-prone mutants still haven't been compacted in large and localised inclusions, these variants would promote more oxidative stress and cytotoxicity than their more soluble counterparts. To validate this prediction, Gln and Ile mutants were analysed 8 h after induction of protein expression. At this time point, Gln fluorescence was homogenously distributed in the cytosol, whereas in the case of Ile, multiple small fluorescent foci were already observable in the cells (Fig. 4.11A). As expected, the levels of GFP fluorescence, measured by FC, were higher in Gln expressing cells (Fig. 4.11B), but contrary to what was observed at 16 h, after 8 h of expression it is the more aggregation-prone Ile variant that exhibits the highest levels of oxidative stress (Fig. 4.11C) and the lowest viability (Fig. 4.11D).

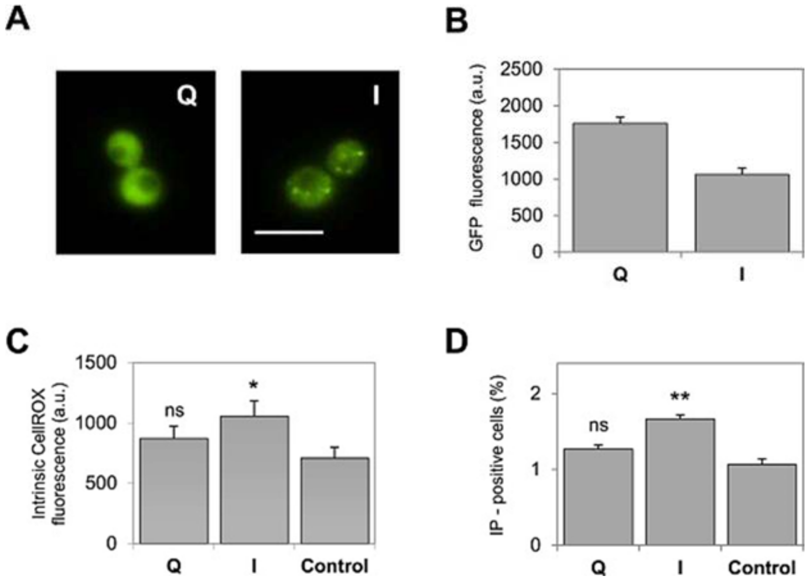


Figure 4.11. Characterisation of GFP fluorescence, oxidative stress levels and cell viability for *S. cerevisiae* cultures expressing Gln and Ile Aβ42-GFP mutants for 8 h. (A) Representative fluorescence microscopy images of yeast cells expressing selected variants for 8 h. Scale bar represents 10 µm. (B) Median fluorescence values of expressed GFP from 10,000 cells determined by FC analysis (FITC-A) for selected Aβ42-GFP variants. (C) CellROX median fluorescence intensity (APC-A) (as indicator of oxidative stress levels) determined by FC in selected Aβ42-GFP variants. Error bars represent CV of the FC gated cells (P3 population). (D) The percentage of IP positive cells (as indicator of cell mortality) was obtained by FC analysis (PerCP-Cy5-A) for selected variants. Error bars represent ±SE (n=3). Cells expressing GFP alone were used as control.

4.3. Material and methods

Protein expression in Saccharomyces cerevisiae

Yeast cells BY4741 (MAT a his3Δ1 leu2Δ0 met15Δ0 ura3Δ0) transformed as previously described (Villar-Pique and Ventura, 2013; Morell *et al.*, 2011) with pESC(-Ura) plasmid (Addgene) encoding for the Aβ42-GFP protein and 19 variants differing only in the residue in position 19th of the peptide, together with the variant F19D/L34P were grown overnight in glucose synthetic complete ura- dropout medium (SC-URA) medium at 30 °C, and 100 µL was used to inoculate 5 mL of fresh medium. At an OD590 of 0.5, cells were changed to a fresh raffinose SC-URA medium. After 30 min of incubation, cells were changed again to a fresh SC-URA medium containing 2% of galactose, as a carbon source, to induce the recombinant protein expression during 8 h, 16 h and 48 h. Cells were harvested and pellets were stored at −80 °C.

Prediction of aggregation propensities

The aggregation propensity of the generated variants was analysed using three different algorithms: (i) AGGRESCAN (Conchillo-Sole *et al.*, 2007), (ii) TANGO (Fernandez-Escamilla *et al.*, 2004) and (iii) FoldAmyloid (Garbuzynskiy *et al.*, 2010). Aggregation propensities were calculated using the default settings.

Fluorescence and confocal microscopy

Cells expressing selected Aβ42-GFP variants for 8 h, 16 h and 48 h were washed with PBS and visualised by fluorescence microscopy (Nikon Eclipse TE200E). Cells were washed with PBS before image acquisition at 100-fold magnification. GFP fluorescence was acquired with a 482/35 nm laser and emission was collected at 536/40 nm. The cultures expressing Gln and Ile variants of Aβ42-GFP during 16 h, were added 10 µM dihydroethidium (Life Technologies) and incubated for 10 min in dark at 30 °C. Cells were washed in PBS, and deposited on top of glass slides. Images were obtained under UV light using a filter for GFP excitation (450–500 nm) and collecting with an emission range (515–560 nm) in a Leica fluorescence microscope (Leica Microsystems, Germany), while the fluorescence of dihydroethidium was excited using a 521–565 nm filter and the emission was collected at 553–633 nm. Image acquisition and data analysis were performed using NIS elements imaging software.

Flow cytometry analysis

Yeast cells after 8 h, 16 h and 48 h of expression were washed and resuspended in sterile PBS to an OD590 of 0.1. Flow cytometry measurements were performed using a BD FACSCanto flow cytometer (BD Biosciences) equipped with a 488 nm blue laser (for GFP) and 635 nm red laser (for CellROX and propidium iodide). Gated cells (by means of FSC and SSC parameters) were analysed for green emission measured on 530/30 nm BP filter and for deep red emission on 660/20 BP filter. Obtained data from 10,000 events were analysed using BD FACSDiva software (BD Biosciences) and the median fluorescence intensities at the emission maximum were obtained.

Oxidative stress detection

5 µM CellROX Deep Red Reagent (Life Technologies) was added and cells were incubated at 30 °C, for 30 min in the dark. After incubation, cells were washed 3 times and resuspended in the same volume of PBS before their analysis with flow cytometry.

Propidium iodide staining

Cell pellets were resuspended in PBS to an OD590 of 0.1. and incubated for 20 min with 2 µg/mL propidium iodide (here abbreviated IP to avoid confusion with protein inclusions (PI)). After incubation, cells were washed 3 times and resuspended in the same volume of PBS before their analysis with flow cytometry.

ROS assay

Hydrogen peroxide levels were determined using Red Hydrogen Peroxide Assay Kit (Enzo Life Sciences) according to manufacturer's protocol. 10 mL of cells induced for 16 h were harvested and resuspended in lysis buffer (20 mM phosphate buffer, 5 mM EDTA, 0.2 mM PMSF, pH 7.2). Cells were broken using glass beads (0.5 mm diameter) by vortexing eight times for 1 min with intervals of 1 min on ice. Supernatant was removed to new tubes and the concentration of total protein was determined by Bradford method. The conversion of red peroxidase substrate was determined by monitoring the fluorescence increase with an excitation at 531 and an emission filter at 595 nm in a Victor 3 Plate Reader (Perkin-Elmer).

Antioxidant enzyme activity determination

The catalase (CAT) activity was determined using Catalase Assay Kit (Sigma-Aldrich) according to manufacturer's protocol. 10 mL of cells expressing for 16 h were harvested, resuspended in lysis buffer (10 mM HEPES, 1.5 mM MgCl2, 10 mM KCl, 0.5 mM DTT, 0.2 mM PMSF, pH 7.9) and broken as described above. The amount of remaining H_2O_2 after the action of catalase was determined by measuring the absorbance of red quinoneimine dye at 550 nm in a Victor 3 Plate Reader (Perkin-Elmer).

Protein carbonylation determination

Yeast cells after 16 h of expression were broken by repeated vortexing in cold conditions using glass beads (0.6 mm diameter). Total protein extracts were centrifuged at 10,000 rpm for 10 min in order to obtain soluble fractions and at 13,000 rpm for 30 min in order to obtain insoluble fractions. Obtained fractions were boiled in 6% SDS at 90 °C for 3 min and stored at

−20 °C for further analysis. Before adding SDS the concentration of the fractions was determined by Bradford method. The analysis of protein oxidation for soluble and insoluble fractions were performed by derivatization of carbonyl groups with dinitrophenylhydrazine (DNPH), followed by SDS–PAGE, transfer to PVDF membranes and western blotting, as described (Levine *et al*., 1994). The antibodies employed were: anti-DNP antibody (Dako) and goat anti-rabbit IgG-HRP conjugate (Pierce). Bands were quantified by densitometry analysis using ImageLab (Bio-Rad). The level of carbonylation detected for each individual Aβ42-GFP mutant was normalised relative to the one exhibited by control cells expressing GFP alone.

Immunoblotting analysis

Collected cells were resuspended in PBS. 200 µL of each mutant was prepared to an OD590 of 20. For soluble/insoluble fraction analysis, 100 µL of each mutant was harvested by centrifugation and resuspended in the same volume of Y-PER protein extraction reagent (Thermo Scientific), supplemented with 1 mM PMSF, and incubated for 20 min at room temperature under mild agitation. Then the protein extract concentration was determined by Bradford method. Samples were centrifuged at maximum speed for 15 min. Soluble fractions were removed to new tubes and insoluble fractions were resuspended in 100 µL of PBS containing PMSF. 5 µL of soluble and insoluble fractions of each sample were loaded on a 15% SDS-PAGE and blotted onto a PVDF membrane. The antibodies employed were: β-amyloid antibody 6E10 (BioLegend) and goat anti-mouse IgG-HRP conjugate (Bio-Rad). Membranes were developed with Luminata (Milipore) and bands were quantified by densitometry analysis using Quantity One software (Bio-Rad).

Proteinase K resistance assay

Insoluble fractions, obtained as described above, were incubated with 10 µg/mL proteinase K (PK) (Sigma-Aldrich) in PBS at 37 °C. Aliquots digested with PK were taken at 5 and 10 min and the reaction quenched by the addition of the same amount of 4 times concentrated denaturing sample buffer. Samples were heated at 99 °C for 10 min and 5 µL of each sample was loaded on SDS-PAGE. After blotting onto a PVDF membrane, the digested fractions were detected by western blotting as previously described.

Statistical analysis

Statistical significance was determined using the two-tail t-Test. P values <0.05 were considered statistically significant.

5. Role of the non-amyloid component Greek-key motif in the aggregation and cytotoxicity of human a-synuclein

5.1. Introduction

Parkinson's disease (PD) is the second most frequent human neurodegenerative disease, pathologically characterised by intracytoplasmic aggregates composed of α-synuclein (α-syn), called Lewy bodies and Lewy neurites (Maroteaux and Scheller, 1991; Spillantini *et al.,* 1997;. Spillantini *et al*., 1998; Goedert, 2001). α-syn is an abundant brain protein present in high concentration at presynaptic terminals and it's found to be located in both soluble and membrane-associated brain fractions. Furthermore, α-syn has been shown to physically interact with at least 30 proteins, underlying its role in cell signalling (Dev *et al*., 2003; Uversky, 2008; Payton *et al.*, 2008).

The biological role of this protein has not been completely elucidated yet, but is believed to be involved in the regulation of synaptic vesicle release and fatty acid-binding (Sharon *et al.* 2001). Due to the physiological role of α-syn as lipid binding protein, studies have been performed to investigate the *in vitro* behaviour of α-syn in the presence of lipids. Under physiological conditions, α-syn forms aggregates that structurally resemble those found in the brains of PD patients. It has been shown that α-syn is remarkably stable in the solution and the initiation of its aggregation usually requires the involvement of external factors and interfaces that promote its aggregation, such as air-water boundaries, the presence of nanoparticles or lipid vesicles (Grey *et al*., 2011; Butterfield and Lashuel, 2010; Rabe *et al*., 2013). In the presence of negatively charged lipid vesicles, the rate of primary nucleation can be increased by more than three orders of magnitude, at high protein:lipids ratios (Galvagnion C. *et al*., 2015).

In its monomeric state, this protein is an intrinsically disordered protein (IDP), having its primary structure composed of N-terminal domain (1-60), hydrophobic fibrillation-triggering non-amyloid-β component (NAC) (61-95), and C-terminal domain (96-140), as indicated in *Introduction*. During α-syn binding, the conformational change occurs manifested by the formation of two α-helices by both N-terminal and NAC domain, leaving C-terminal domain unstructured. The central hydrophobic NAC region has been implicated as the most important region for the aggregation properties of α-syn (Jethva *et al*., 2011), having a critical role in nucleation of the aggregation process (Hanz *et al*., 1995). Moreover, it has been shown that deletion of certain residues within the NAC region abolishes α-syn aggregation (Bodles *et al*., 2001; Giasson *et al*., 2001), peptides derived from the NAC region can form β-sheet fibrils (El-Agnaf *et al*., 1998; Han *et al*., 1995) and are toxic to cells (Bodles *et al*., 2001). Additionally to α-syn, there are other types of synucleins, β-syn and γ-syn, all sharing highly conserved N-terminal domain (George, 2002). The former has been reported to have a suppressing effect on α-syn aggregation *in vitro* and *in vivo* (Hashimoto *et al*., 2001; Fan *et al*., 2006), and later to be evolutionarily more closely related to to α-syn but has not been associated with the onset of PD (Nishioka *et al*., 2010). Interestingly, the most striking difference between these proteins is that α-syn, but not β-syn or γ-syn, possesses the NAC domain.

Recently, a structure of the pathogenic fibril of full-length human α-syn has been solved demonstrating that the NAC domain within the fibril forms the core structure that contains β-serpentine arrangement with a Greek key β-sheet topology (Tuttle *et al*., 2016). The solved structure indicates NAC region as a critical region involved in α-syn fibril nucleation, propagation and even strain formation.

Another important region of α-syn, which harbours all missense mutations identified to date that alter oligomerization and fibrillization properties of α-syn, is the N-terminal region (Ghosh *et al*., 2014). Furthermore, β1-β2 region, located at N-terminal region of α-syn and comprised of amino acids 35-59 (Waxman *et al*., 2009), has been implicated to have a critical role in the aggregation and pathogenesis of α-syn. The tertiary contacts between β1 and β2 region have been shown to be the most prevalent interactions among five identified β-strands of α-syn in the amyloid fibril state (Esteban-Martin *et al*., 2013). An importance of transient interactions between these two segments in the aggregation and pathogenesis of α-syn has been evaluated by engineering a novel variant of α-syn having the two segments covalently cross-linked through the engineered intramolecular disulfide bond (Shaykhalishahi *et al*., 2015). The generated variant of α-syn harbouring a novel disulfide bond was shown to be non-fibrillogenic *in vitro* and even having an inhibitory effect on the aggregation of wt α-syn and some other amyloidogenic peptides.

In our work, following a similar line of research, an artificial disulfide bond has been generated within the NAC region of full-length human α-syn, and the corresponding changes in the amyloid formation tendencies that the established disulfide bond promotes have been analysed. Since α-syn is a dynamic molecule whose secondary structure depends on its environment and it naturally lacks cysteine residues, the behaviour of wt α-syn and its engineered disulfide bridge-forming variant have been analysed in three different conditions to identify the differences in the strain formation. Moreover, to determine the step within the amyloidogenic pathway where the stapling of the Greek-key motif in the engineered disulfide bridge-forming variant has an effect, the aggregation kinetics have been performed and the differences in the aggregation pathways in the absence and presence of lipids were determined.

5.2. Results and discussion

5.2.1. Cross-linking of the Greek-key motif in the NAC region of α-syn

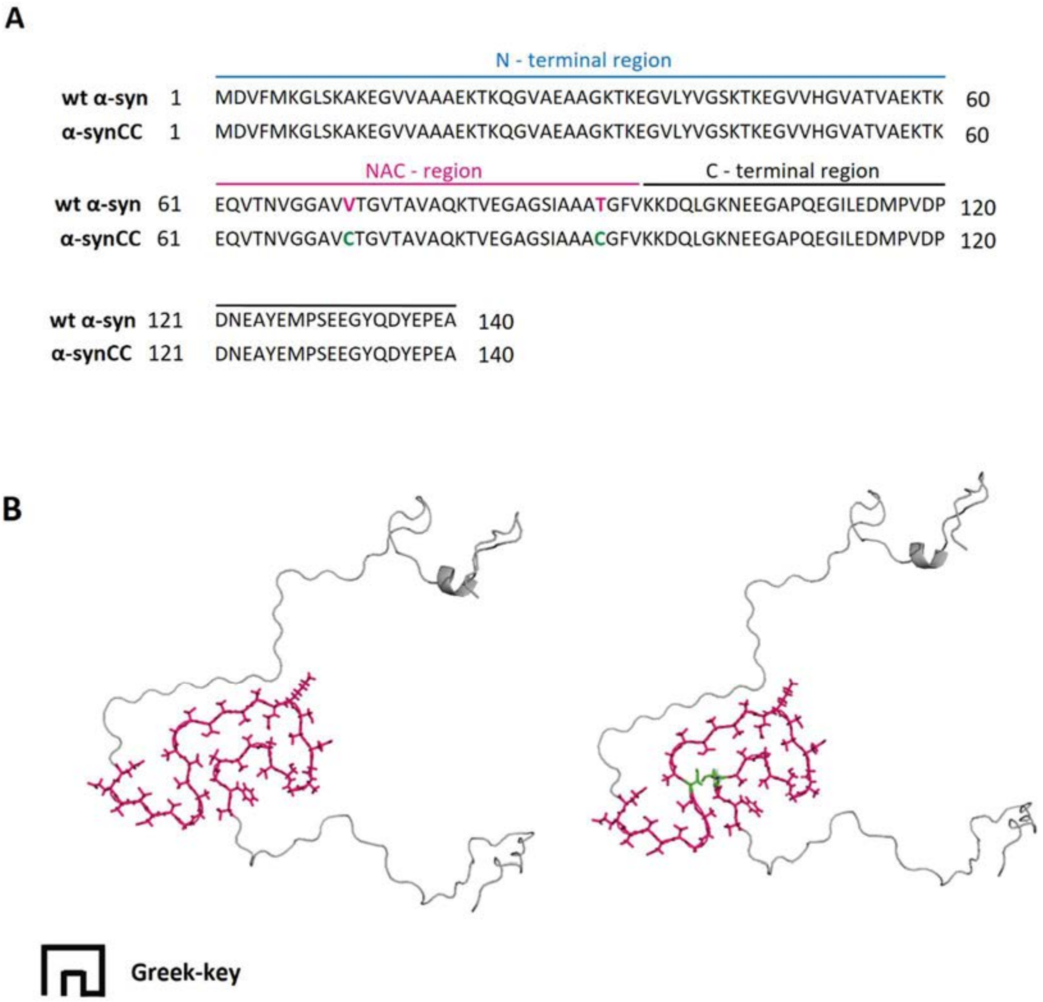
Disulfide by Design 2.0 (Craig and Dombkowski, 2013) was used to design α-syn variant with two cysteine residues introduced within NAC region, at the positions where disulfide bridge could be established between them. The central hydrophobic NAC region was selected due to its importance in the aggregation process of α-syn molecule (Fig. 5.1A). We searched for two residues close enough in the Greek-key motif to permit the formation of the desired covalent interaction after their mutation to Cys (Fig. 5.1B). Performed computational analysis indicated Val71 and Thr92 as only optimal candidates, due to their proximity (distance between the β-carbons = 3.1 A˚), face-to-face orientation, and proper dihedral angles. Therefore, we cloned, produced, and purified a α-syn variant with both residues replaced by Cys (α-synCC). The absence of detectable alkylation by vinylpyridine in the recombinant purified protein confirmed that the designed disulfide bond is spontaneously formed. The introduced covalent linkage closes off a loop of 20 residues.

5.2.2. NAC cross-linking does not alter the predicted aggregation propensity, disorder and hydrophobicity of α-syn

Sequential predictors of aggregation propensity, such as AGGRESCAN (Conchillo-Sole *et al.*, 2007), FoldAmyloid (Garbuzynskiy *et al.*, 2010) and Zyggregator (Tartaglia *et al.*, 2008), were used to estimate the effect of mutations on the intrinsic aggregation propensity of α-syn (Fig. 5.2A). Overall, the analysis demonstrates that the introduced mutations are not expected to promote by themselves significant changes in the aggregation propensity of α-syn, indicating that the mutations could have an effect on the structure instead.

In order to estimate if the mutations had an effect on the intrinsic order, IUPred (Dosztanyi *et al.*, 2005) and RONN (Yang *et al.*, 2005) predictors were used. Both wt α-syn and double cysteine mutant α-synCC are predicted to possess two disordered regions having probability scores ≥ 0,5, according to IUPred server: one common region containing residues 50-56, and a second region containing residues 97-140 for wt α-syn, and residues 99-140 for α-synCC, corresponding to the C-terminal domain of protein (Fig. 5.2B). According to RONN predictor, both proteins are predicted to be almost completely disordered, emphasizing that the sequences are typical of the intrinsically disordered proteins, with the exception in the region containing residues 61-93, assigning reduced intrinsic disorder in this region to α-synCC only (Fig. 5.2C). Interestingly, the mentioned region of protein comprises almost complete NAC region. This suggests that the corresponding server was able to detect the change in the intrinsic order promoted by the mutations introduced in α-synCC.

The hydrophobicity analysis according to the scale of Kyte and Doolittle (Kyte and Doolittle, 1982), demonstrates that for both wt α-syn and α-synCC, hydrophobicity plot profiles are almost the same, indicating the NAC region of α-syn as hydrophobic (Fig. 5.2D). The mutation of Val71 and Thr92, belonging to NAC region, to Cys in case of α-synCC, and disulfide cross-linking of this hydrophobic region could be entropically favourable, protecting the hydrophobic region from the exposure to the solvent. Moreover, α-synCC displaying polar free thiols in the reduced state would be entropically unfavourable.



wt α-syn 1	MDVFMKGLSKAKEGVVAAAEKTKQGVAEAAGKTKEGVLYVGSKTKEGVVHGVATVAEKTK	60
α-synCC 1	MDVFMKGLSKAKEGVVAAAEKTKQGVAEAAGKTKEGVLYVGSKTKEGVVHGVATVAEKTK	60
wt α-syn 61	EQVTNVGGAVVTGVTAVAQKTVEGAGSIAAATGFVKKDQLGKNEEGAPQEGILEDMPVDP	120
α-synCC 61	EQVTNVGGAVCTGVTAVAQKTVEGAGSIAAACGFVKKDQLGKNEEGAPQEGILEDMPVDP	120
wt α-syn 121	DNEAYEMPSEEGYQDYEPEA	140
α-synCC 121	DNEAYEMPSEEGYQDYEPEA	140

Fig. 5.1. Primary sequence and three-dimensional structure of wt α-syn and its double cysteine mutant α-synCC. (A) Primary sequence alignment of wt α-syn and α-synCC. The amino acid sequences of wt α-syn (UniProt ID: P37840) and α-synCC are shown, with the differences in α-synCC coloured in green. The three distinct regions of α-syn are indicated. (B) Structural models of wt α-syn (*left panel*) and α-synCC (*right panel*) monomers showing NAC domain in purple and the mutated residues Cys71 and Cys92 in α-synCC coloured in green. The Greek-key motif of the fibril core is shown. The models were generated in PyMol using the solved structure of pathogenic fibril of full-length human α-syn (PDB: 2N0A).

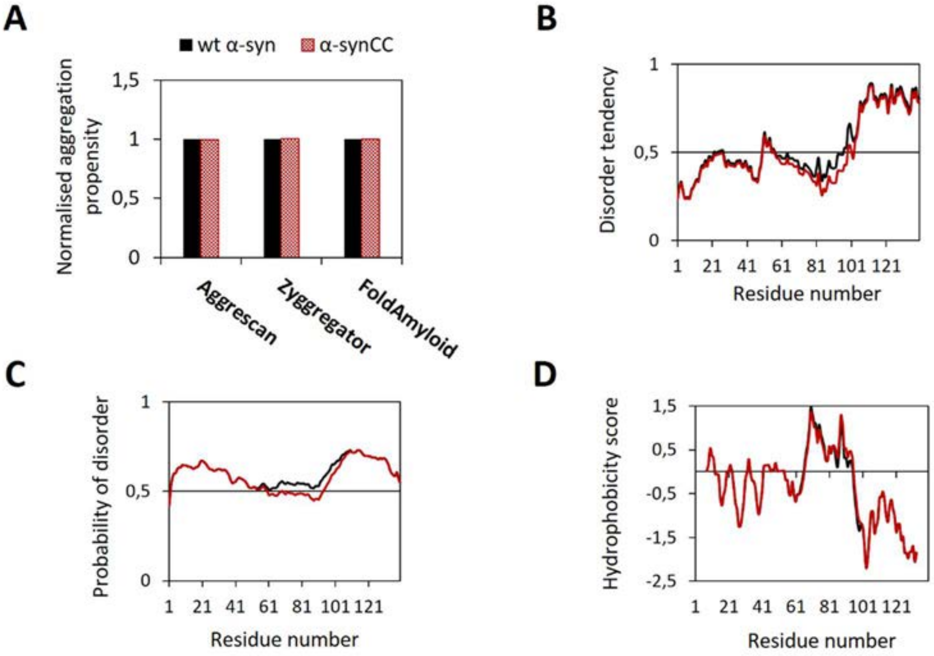


Figure 5.2. Sequence analysis of wt α-syn and α-synCC. (A) Aggregation propensity analysis of wt α-syn (black) and α-synCC (dark red) using different sequential predictors: AGGRESCAN (Conchillo-Sole *et al*., 2007), Zyggregator (Tartaglia *et al*., 2008) and FoldAmyloid (Garbuzynskiy *et al*., 2010). Intrinsic disorder predictions for wt α-syn and α-synCC using: (B) IUPred (Dosztanyi *et al.,* 2005) and (C) RONN (Yang *et al.,* 2005). Residues exhibiting values above 0.5 are considered to be disordered. (D) Kyte-Doolittle hydrophobicity plots (Kyte and Doolittle, 1982) of wt α-syn and α-synCC. Residues exhibiting values above 0 are considered to be hydrophobic. IUPred and RONN predictions were performed by implementing an 11-residue sliding window.

5.2.3. NAC cross-linking promotes α-syn compaction

Early experiments evaluating the contribution of disulfide bonds to the thermodynamic stability of proteins were involving methodically disrupting of native disulfide bonds, such as two disulfide bridges in ribonuclease T. It has been demonstrated that each cross-link contributed ~3.5 kcal/mol to the thermodynamic stability of the protein, consistent with a range of 2.3–5.2 kcal/mol reported in other experiments (Pace *et al*., 1988; Tidor and Karplus, 1993). It is generally assumed that the increase in stability promoted by the formation of disulfide bridges is determined to be due to the loss of conformational entropy (ΔS) of the unfolded state since the establishment of covalent link, such as cysteine disulfide bond, restricts motion of the unfolded random coil more efficiently than the restrictions imposed upon the folded protein. Direct comparison of the oxidized form of α-synCC with its dithiol form allows differentiating the effect of the mutations from the effect of the disulfide bond.

The contribution of a disulfide cross-linking to the change in conformational entropy of the protein was modelled by the expression:

$$\Delta S = 2.1 \left(\frac{3}{2}\right) R \ln \mathrm{N}$$

The ΔS is related to the size of the polypeptide loop enclosed by the disulfide bond (N).

Due to the establishment of disulfide bond in α-synCC, the thermodynamic stability of the mutant is increased for 78.46 J / mol. K. compared to wt α-syn.

To determine the existence of desired intracellular disulfide cross-linking between Cys71 and Cys92 in purified recombinant α-synCC, classical techniques were implemented. Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry (MALDI-TOF MS) analysis of soluble wt α-syn and α-synCC, in each case indicate the dominant peak due to the monomeric form of α-syn, with the molecular weight (MW) of 14460.6 Da for wt α-syn (Fig. 5.3A) and 14464.9 Da for α-synCC (Fig. 5.3B). For wt α-syn, this is very close to the theoretical MW of 14460.0 Da, however, there is an evident difference in case of α-synCC. This difference in the experimental (14464.9 Da) and theoretical average mass (14466.1 Da) of α-synCC accounts for two 1H present in the reduced form but absent in the oxidized from. To further analyse the formation of an intramolecular disulfide bridge between Cys71 and Cys92 in α-synCC, protein sample was incubated in 10 mM dithiothreitol (DTT) prior to the iodoacetamide (IAM) addition (Fig. 5.3B). The disulfide formation was confirmed by the fact that the respective peak indicating oxidized α-synCC (14464.9 Da) shiffted to 14578.5 Da (Fig. 5.3B). This observed mass shift of approximately 114 Da corresponds to alkylation of two Cys with two IAM molecules, each IAM accounting for 57.05 Da.

Furthermore, we used mass spectrometry (MS) under non-denaturing conditions (native MS), which permits monitoring compaction and oxidation distinctly and simultaneously (Fig. 5.3C; Fig. 5.3D and Fig. 5.3E). While accurate mass determination informs about the oxidation state, charge-state distributions allow detection of simultaneously coexisting conformational species. Electrospray Ionization Mass Spectrometry (ESI-MS) represents a key tool of structural biology, allowing the detection of non-covalent interactions (Marcoux and Robinson, 2013; Schmidt and Robinson, 2014). The measured molecular masses (14459.8 Da for wt α-syn, 14464.4 Da for oxidized α-synCC and 14466.1 Da for reduced α-synCC) are in good agreement with the theoretical ones (14460.0 Da, 14464.1 Da and 14466.1 Da, respectively). The main effects of the introduction of the disulfide bridge in α-synCC (evident in 4-components Gaussian fits) is the accumulation of the compact intermediate state (green curve; 68%) at the expenses of the less compact conformers (cyan and magenta curves; 13 and 11%) (Fig. 5.3D). Under reducing conditions, the conformational ensemble of the mutant is much more similar to the wt α-syn, indicating that the observed effect is due to the disulfide bridge itself and not merely to the mutation. The data demonstrates that the disulfide bridge leads to the accumulation of the green component, a decrease of the cyan component and the loss of the magenta component (fully extended conformation) (Fig. 5.3D).

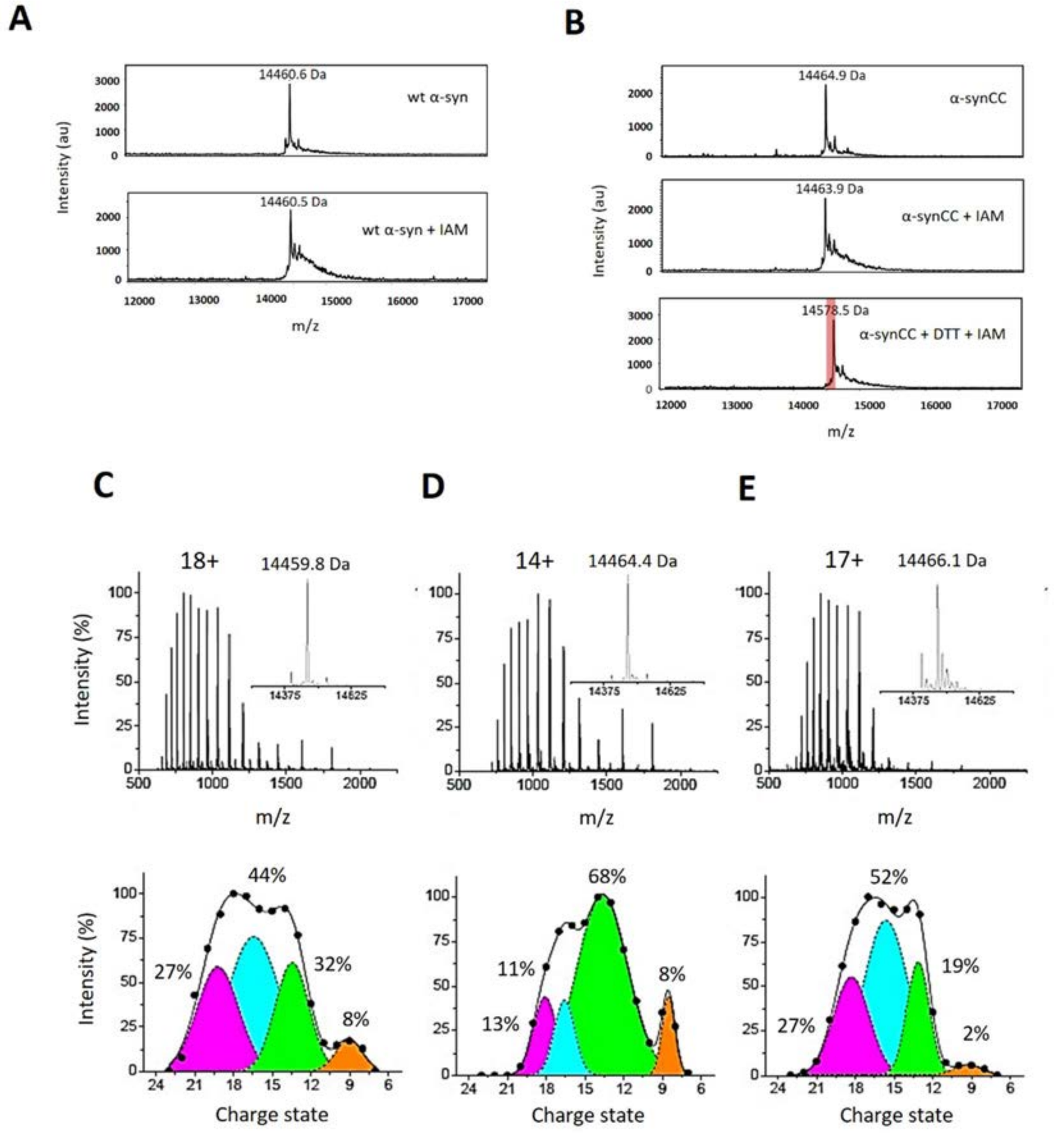


Figure 5.3. Comparison between compactness of wt α-syn and α-synCC. (A-B) MALDI-TOF MS analysis of soluble wt α-syn and α-synCC, together with the analysis of the intramolecular disulfide bridge between Cys71 and Cys92 of α-synCC. The peak at 14460.6 Da corresponds to wt α-syn (A) and the peak at 14464.9 Da corresponds to α-synCC (B). Addition of dithiothreitol (DTT) to α-synCC led to the reduction of the disulfide bond and the reaction with iodoacetamide (IAM) occurred. A mass shift of 114 kDa is shown (red square), indicating the alkylation of 2 Cys. (C-D) ESI-MS analysis showing the conformational ensembles of wt α-syn and α-synCC. ESI-MS spectra (*upper panels*), and Gaussian fits of the charge-state distributions obtained using 4 conformers (*lower panels*) for wt α-syn (C), oxidized α-synCC (D) and reduced α-synCC (E) are shown. The most intense peak of each ESI-MS spectra is labelled by the corresponding charge state. Inserts report the mass-deconvolution spectrum. Each Gaussian component is labelled by its relative amount in percentage.

5.2.4. Conformational properties of wt α-syn and α-synCC

To investigate if the covalent cross-linking in α-synCC has an effect on the exposure of hydrophobic patches, the Reverse-Phase High Performace Liquid Chromatography (RP-HPLC) has been performed with the oxidized α-synCC and α-synCC that has been reduced with DTT prior to HPLC analysis (Fig. 5.4A). RP-HPLC elution profiles clearly indicate that the oxidized α-synCC elutes faster than its fully reduced form, indicating that cross-linking in α-synCC renders the more compact conformation exhibiting less surface hydrophobicity. On the other hand, the profile of DTT-reduced α-synCC confirms that the Cys are exposed to solvent and in a flexible environment after reduction with DTT. The reduced form of α-synCC evidently has hydrophobic patches exposed to the solvent as well, which binds to the stationary phase during the separation and causes DTT-treated α-synCC to elute later. Since the disulfide cross-linking established in the oxidized α-synCC results in the closing of the NAC region, identified by hydrophobicity analysis as the most hydrophobic part of the protein, we assume that faster elution of oxidized α-synCC results from the inability of this same region to bind to the RP-HPLC column. Compared to oxidized α-synCC, the Cys residues in the reduced form are exposed to solvent and flexible, which influences the elution to be delayed.

The structural features of recombinant purified wt α-syn and α-synCC were evaluated by Nuclear Magnetic Resonance (NMR) spectroscopy. The temperature-induced unfolding of wt α-syn and its double cysteine mutant was analysed by monitoring the changes in their one-dimensional NMR (^{1}H-NMR) spectrum (Fig. 5.4B). ^{1}H-NMR spectrums of the wt α-syn and oxidized α-synCC under native conditions (298K and pH 7.4) display a narrow signal dispersion characteristic of natively unfolded proteins (Fig. 5.4B). α-synCC spectrum exhibits resonances with sharper peaks compared to the spectrum of wt α-syn. Both wt α-syn and α-synCC start to loose peak sharpness at 318K, wt α-syn to a greater extent than α-synCC, keeping their native signal dispersion. Above this temperature, the intensity of the signals reduces for wt α-syn, but it doesn't change significantly in case of α-synCC where the peak intensity is still preserved. Both proteins exhibit reversible thermal denaturation behaviour as indicated by the complete recuperation of the native spectrum at 298K after cooling down the sample (Fig. 5.4B).

Homonuclear or Heteronuclear Multidimensional (HetMulD) NMR spectroscopy assisted with a stable isotope labelling technology is a powerful technique for quantitatively investigating detailed structural features in IDPs (Lee *et al.*, 2000; Lee *et al.*, 2012). However, a fundamental problem in the structural characterization of IDPs is the definition of the conformational ensemble sampled by the polypeptide chain in solution which can often complicate sequential signal assignment in NMR spectroscopy. We performed the homonuclear two-dimensional NMR (2D NMR) spectroscopy, detecting signals of the same isotope (^{1}H), for the wt α-syn and its double cysteine mutant. Unfortunately, HetMulD NMR spectroscopy with a stable isotope labelling was not possible to perform due to the inability of Origami B (DE3) *E. coli* cells, which are the expression strain in case of α-synCC, to grow in the minimal medium during the isotope labelling process. A 2D ^{1}H-^{1}H spectrum of both α-syns indicates the narrow chemical shift dispersion in ^{1}H dimensions, characteristic of natively unfolded protein (Fig. 5.4C). In case of α-synCC, a slightly broadening of the signal dispersion can be observed, which suggests that the formation of a disulfide bridge in the oxidized α-synCC didn't influence significantly protein to adopt more folded conformation. Both wt α-syn and α-synCC remained mainly disordered, however, there is no signal overlapping. Due to the evidence that Cys71 and Cys92 are spontaneously cross-linked in α-synCC, the differences observed in ^{1}H-^{1}H spectrum between wt α-syn and α-synCC are indicating that the compaction that occurred in case of α-synCC, caused the different spatial distribution of the amino acid residues in the protein conformation. Overall, NMR data demonstrate that wt α-syn and α-synCC in the same conditions adopt different conformations and that the compaction of α-synCC is the cause of this.

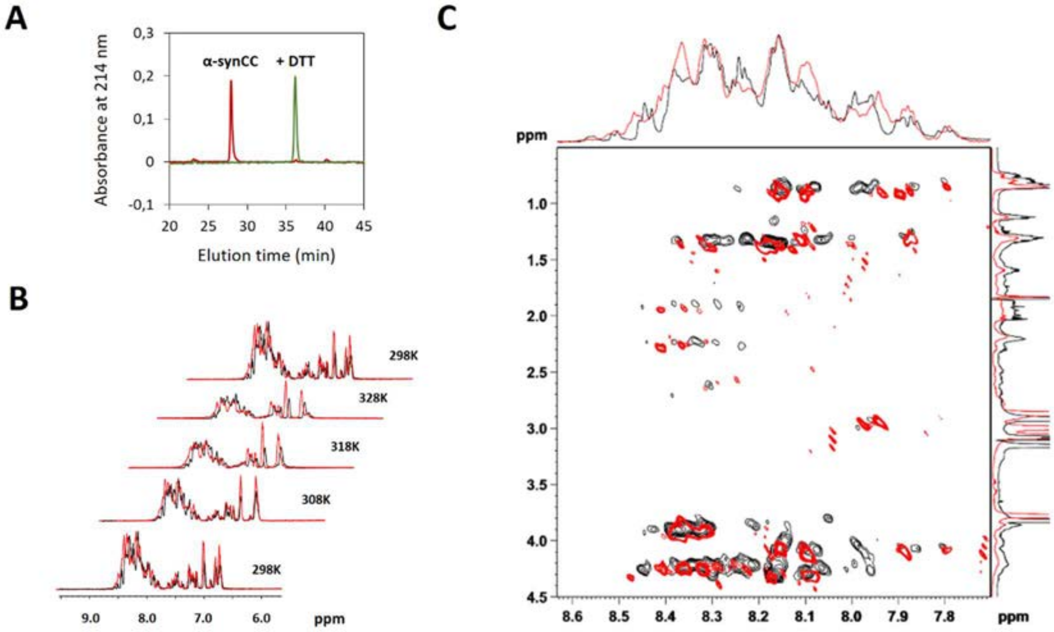


Figure 5.4. Conformational properties of wt α-syn and α-synCC. (A) RP-HPLC pattern of α-synCC before (red) and after reduction with DTT (green). (B) ^{1}H NMR variable-temperature spectra for wt α-syn (black) and α-synCC (reed) in 20 mM HEPES pH 7.4, using a 600 MHz NMR spectrometer showing the temperature range from 298 to 328 K. (C) 2D-TOCSY ^{1}H-^{1}H NMR-spectrum of wt α-syn and α-synCC dissolved in 20 mM HEPES, pH 7.4 at 25 °C.

It has been reported that natively unfolded nature of α-syn, associated with low hydrophobicity and high net charge, can be altered due to the increase in hydrophobicity and/or decrease in net charge, and partial folding can be induced by introducing changes in the environment of α-syn (Uversky *et al.*, 2001). The negative charge of α-syn molecule at neutral pH (pI = 4.7) is expected to be neutralized by acidification of the solution where protein resides, and the overall hydrophobicity of a protein is expected to increase with increasing temperature. Furthermore, it has been previously reported that α-syn is capable of partial compaction-driven by the sequestering of its N- and C-terminal via their interactions with each other and with the central NAC region. The existence of these long-range interactions was suggested to play an autoinhibitory role preventing the spontaneous α-syn oligomerization and aggregation (Bertoncini *et al*., 2005; Dedmon *et al*., 2005). This nascent highly dynamic tertiary structure of α-syn was shown to be abolished under conditions that induce α-syn aggregation *in vitro* (Bertoncini *et al*., 2005) or due to protein interaction with a polyamine. Furthermore, the presence of salt in the protein environment has been reported to modify the tendencies of α-syn to self-assemble due to the establishment of transient interactions with the terminal regions of synuclein aggregation properties of α-syn due to the modification of the repulsion between individual molecules. The pH of the solution also influences the net charge of the polypeptides through the protonation/deprotonation of the amino acid side chain.

Since α-syn naturally lacks cysteine residues in its sequence and studies revealed that it's largely unfolded and devoid of stable tertiary structure (Eliezer *et al*., 2001), we decided to investigate how does the covalent cross-linking of the highly hydrophobic NAC region in engineered α-synCC influences the ensemble of conformational species that this protein will adopt in a given environment compared to wt α-syn. We monitored the conformation of wt α-syn and α-synCC in different buffers at the beginning of the reaction, therefore, prior to aggregation experiments, using far-UV circular dichroism (CD) and staining with the fluorescent probe 4,4`-bis(1-anilinonaphthalene 8-sulfonate) bis-ANS. Proteins were prepared to 60 µM final concentration at room temperature in three different buffers: 50 mM sodium acetate pH 5.0 (buffer A); 10 mM sodium phosphate pH 7.0 (buffer B) and 10 mM sodium phosphate supplemented with 200 mM NaCl pH 7.0 (buffer C). Deconvolution of the CD spectra of freshly dissolved proteins in different buffers indicates that wt α-syn and α-synCC do not display significant differences in secondary structure content in the same buffer (Fig. 5.5A).

In all cases, both α-syn display CD spectra characteristic of random coil secondary structure. The detection of solvent-accessible hydrophobic clusters using bis-ANS demonstrated that in buffer A, both wt α-syn and α-synCC exhibit high affinity for the probe indicating that a protonation of α-syn in a low pH solution renders a molten globule (MG) conformation with hydrophobic patches exposed to the solvent (Fig. 5.5B). Despite the fact that buffer A-induced wt α-syn and α-synCC MG does not exhibit the existence of β-sheet conformations, this state is aggregation competent (Castillo *et al*., 2013). As expected, in buffer B and C, wt α-syn exhibits <2 times higher bis-ANS fluorescence emission than α-synCC, confirming that the cross-linking of NAC region in α-synCC, identified as the most hydrophobic region of α-syn, allows α-synCC to adopt conformation with a tightly packed core protecting NAC region from the exposure to the solvent.

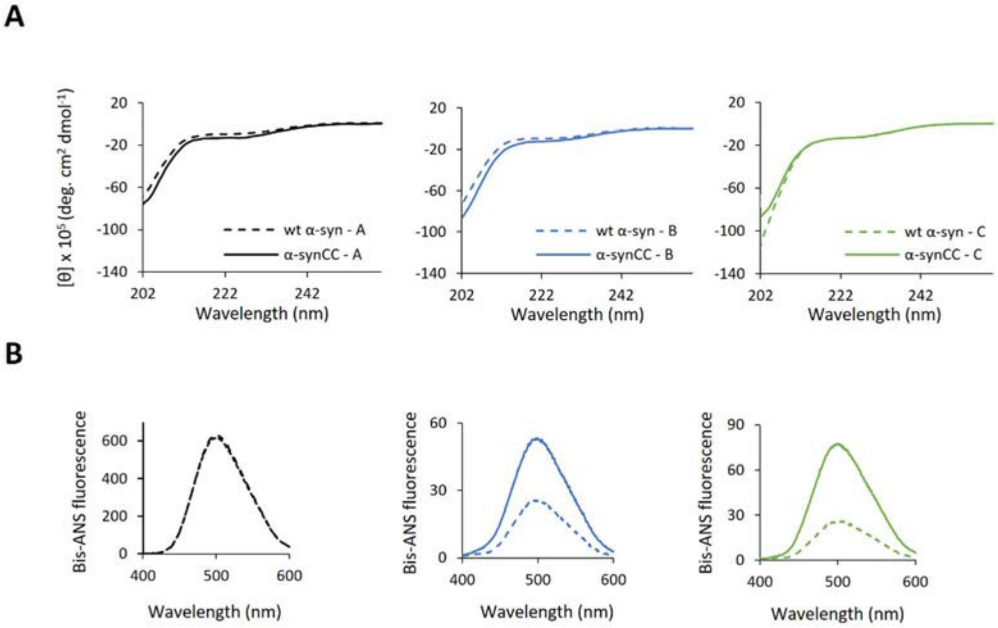


Figure 5.5. Characterisation of soluble wt α-syn and α-synCC. (A) Far-UV CD spectra of wt α-syn (dashed line) and its double cysteine mutant (full line). Soluble proteins at 15 µM were measured in buffer A, B and C. (B) Bis-ANS fluorescence spectra of 10 µM bis-ANS bound to 15 µM soluble wt α-syn and α-synCC in buffer A, B and C.

5.2.5. Structural and morphological characterisation of wt α-syn and α-synCC aggregation reactions

After determining the conformation of wt α-syn and α-synCC in three buffers at the beginning of the reaction, the samples were incubated at 700 rpm and 37°C in thermoshaker. The aggregation process was monitored during fifteen days by measuring the binding of the protein to the amyloid-specific dye Thioflavin-T (Th-T) (Fig. 5.6).

After 5 days, the wt α-syn and α-synCC aggregate formation in three different buffers was characterized by light scattering, Th-T fluorescence and Congo Red (CR) binding (Fig. 5.7A; Fig. 5.7B and Fig. 5.7C). In buffer A, the increase in the Th-T signal is observed only for wt α-syn. This confirms the assumption that in a low pH buffer aggregation reaction is promoted, based on the high bis-ANS binding properties of conformation that wt α-syn exhibited at the beginning of the reaction. Interestingly, contrary to this assumption, α-synCC after 5 days of incubation in buffer A exhibits insignificant Th-T increase, even though the bis-ANS binding at the initial point of the reaction was the same like in wt α-syn. In agreement with these data, the binding to CR is observed for wt α-syn fibrils, but not for the species formed by the α-synCC.

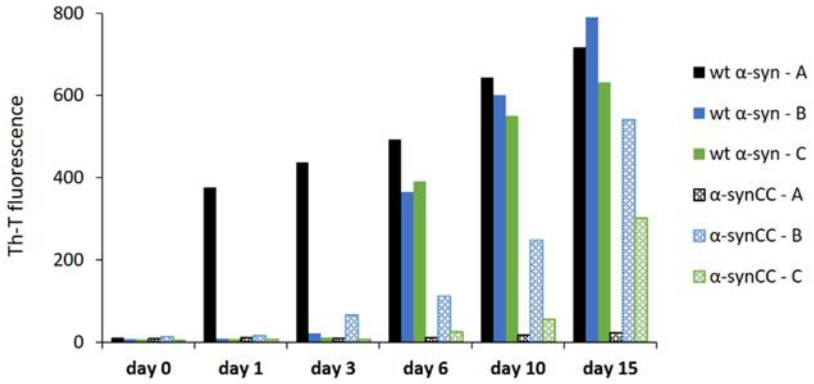


Figure 5.6. Aggregation kinetics of wt α-syn and α-synCC in three different buffers. The represented aggregation kinetics were monitored by following the change in relative Th-T fluorescence during 15 days-incubation of 60 µM protein in buffer A, B and C.

In neutral pH- buffer B, wt α-syn and α-synCC aggregates exhibit similar light scattering. Both bind Th-T and CR, however binding of Th-T in case of wt α-syn is around 2 times higher than in α-synCC. In the presence of salt, both wt α-syn and α-synCC form aggregates as confirmed by Th-T and CR binding. Interestingly, Th-T binding of wt α-syn is >10 times higher than the binding of its double cysteine mutant, even though their aggregates exhibit similar tendencies to bind CR. Moreover, we analysed the conformational properties of α-syn aggregates formed at this time point by bis-ANS binding (Fig. 5.7D). In agreement with the data, wt α-syn fibrils exhibit higher bis-ANS binding than the fibrils formed by the disulfide-containing α-synCC, in all cases except in buffer B. This data suggests the existence of a lower proportion of hydrophobic residues exposed to solvent in the fibrils formed by α-synCC in buffer A and C, but not in buffer B.

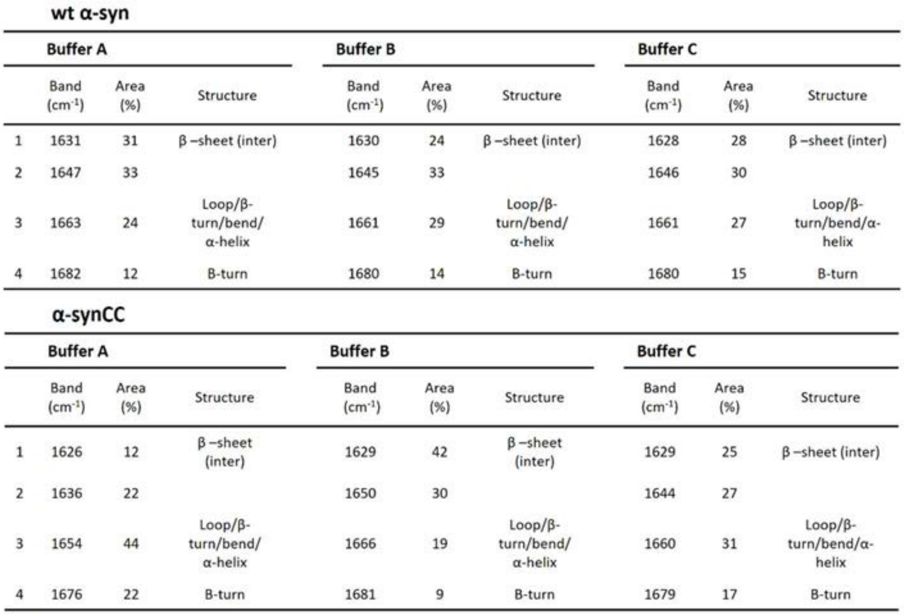
Analysis of the different aggregation reactions at 5 days by transmission electron microscopy (TEM) (Fig. 5.7E) demonstrates amyloid fibrils of wt α-syn formed in all three buffers, but clearly points out the inability of α-synCC to aggregate in the low pH buffer. However, differences in the morphology of the fibrils were observed. wt α-syn fibrils in buffer A exhibit uncommon massive shape. In buffer B, fibrils formed by wt α-syn and its cysteine mutant exhibit similar morphology, they are long and straight. In the presence of salts, fibrils formed by wt α-syn lack structural organisation and by shape they resemble protofibrils which does not justify high Th-T binding. α-synCC in the same condition forms shorter, discrete, and straight structures.

To further explore whether the observed differences in fibril morphology are related to differences in the interactions supporting their architecture, we analysed the conformational properties of α-syn aggregates formed at this time point by limited proteolysis and attenuated total reflectance-Fourier transformed infrared spectroscopy (ATR-FTIR). In agreement with previous results, the digestion patterns were different, underlying a different structure. Additionally, all the aggregates seemed to be resistant to proteinase K (PK) when compared to monomeric α-syn. Furthermore, we observed that, in our conditions, the wt α-syn fibrils in buffer A and C exhibit resistance to PK digestion, whereas the wild-type fibrils in buffer B were susceptible to proteolysis more than expected (Fig. 5.7F). In case of α-synCC, after 15 and 30 min of digestion with PK, assemblies formed in buffer A and C are digested which results in a specific digestion pattern. The fibrils formed in neutral pH-buffer B are exhibiting the highest resistance to PK digestion (Fig. 5.7F).

To assess the secondary structure content of the assemblies formed by wt α-syn and α-synCC, we analysed the amide I region of the ATR-FTIR spectrum (1700–1600 cm^{-1}). This region of the spectrum corresponds to the absorption of the carbonyl peptide bond of the main amino acid chain of the protein, and it represents a marker of the protein secondary structure. After deconvolution of the obtained ATR-FTIR spectra, the contribution of the individual secondary structure elements was assigned to the main absorbance signal for fibrils formed after 5 days of incubation (Fig. 5.7G and Table 5.1).
After deconvolution of the obtained ATR-FTIR spectra, the contribution of the individual secondary structure elements was assigned to the main absorbance signal for fibrils formed after 5 days of incubation (Fig. 5.7G and Table 5.1).

After 5 days of incubation, the condition that renders the biggest differences is low pH buffer. While the spectrum of α-syn is dominated by a peak at 1625 cm^{-1}, attributable to the presence of amyloid-like inter-molecular β-sheet structure (Fig. 5.7G), the spectrum of the α-synCC was dominated by a peak at 1654 cm^{-1} corresponding to disordered/random coil conformation (Fig. 5.7G and Table 5.1). The spectrum for α-synCC in buffer B is exhibiting significant content of intermolecular β-sheet structure compared to wt α-syn in the same buffer.

Table 5.1. Assignment of secondary structure components of wt α-syn and α-synCC after 5 days-incubation in described conditions.



wt α-syn

	Buffer A			Buffer B			Buffer C		
	Band (cm^{-1})	Area (%)	Structure	Band (cm^{-1})	Area (%)	Structure	Band (cm^{-1})	Area (%)	Structure
1	1631	31	β –sheet (inter)	1630	24	β –sheet (inter)	1628	28	β –sheet (inter)
2	1647	33		1645	33		1646	30	
3	1663	24	Loop/β-turn/bend/ α-helix	1661	29	Loop/β-turn/bend/ α-helix	1661	27	Loop/β-turn/bend/α-helix
4	1682	12	B-turn	1680	14	B-turn	1680	15	B-turn

α-synCC

	Buffer A			Buffer B			Buffer C		
	Band (cm^{-1})	Area (%)	Structure	Band (cm^{-1})	Area (%)	Structure	Band (cm^{-1})	Area (%)	Structure
1	1626	12	β –sheet (inter)	1629	42	β –sheet (inter)	1629	25	β –sheet (inter)
2	1636	22		1650	30		1644	27	
3	1654	44	Loop/β-turn/bend/ α-helix	1666	19	Loop/β-turn/bend/ α-helix	1660	31	Loop/β-turn/bend/α-helix
4	1676	22	B-turn	1681	9	B-turn	1679	17	B-turn

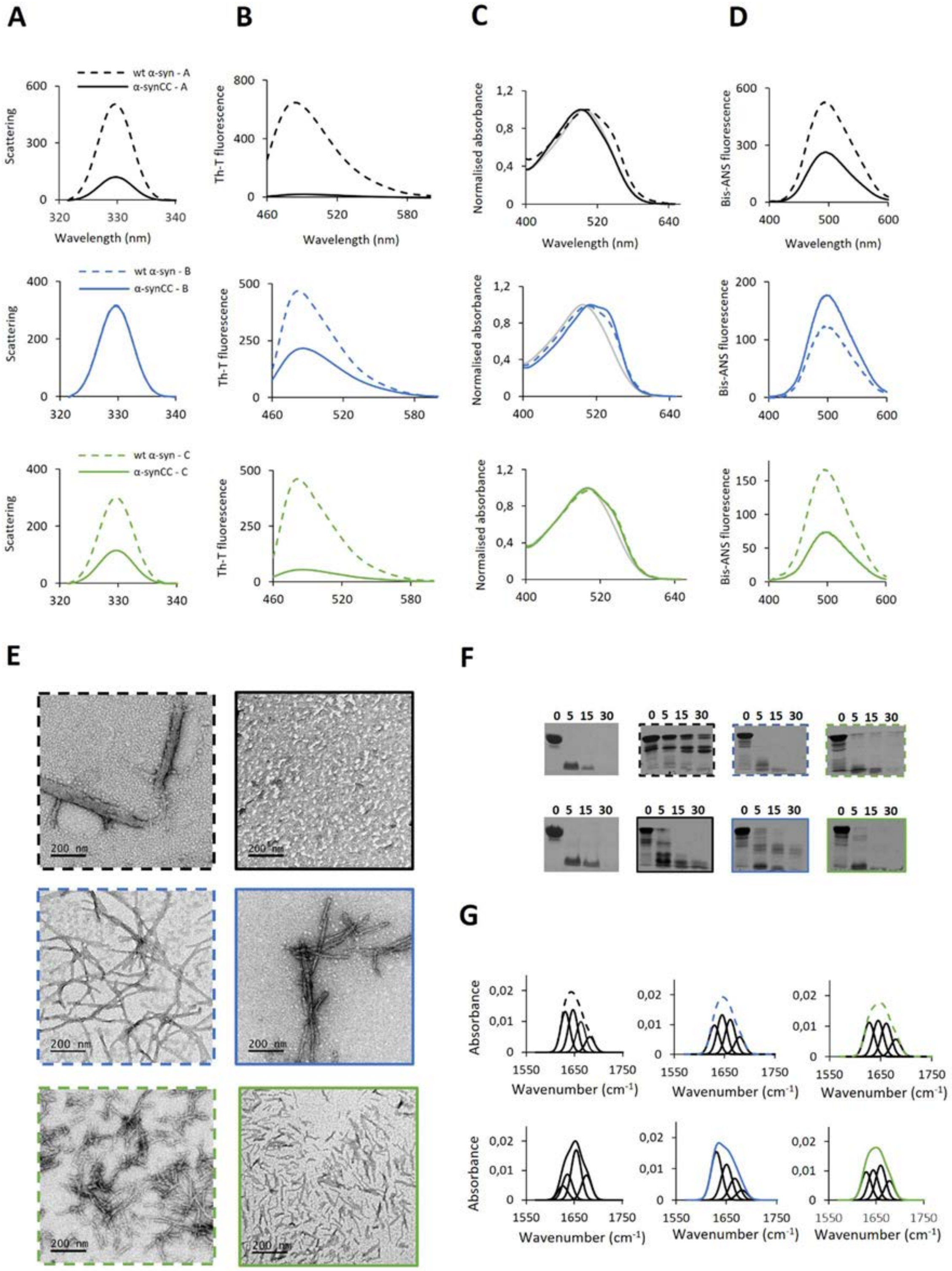
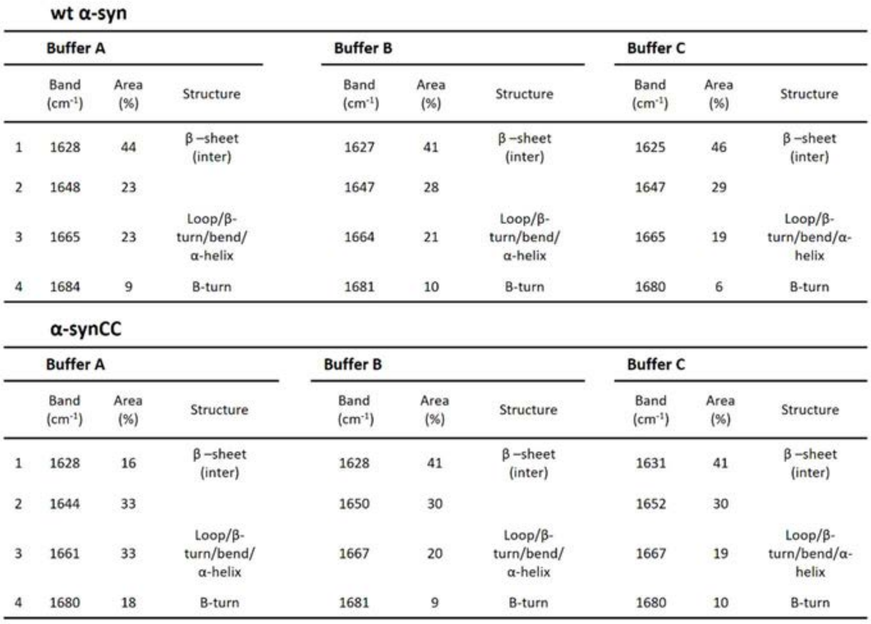

A
B
C
D
wt α-syn - A
α-synCC - A
wt α-syn - B
α-synCC - B
wt α-syn - C
α-synCC - C
Scattering
Wavelength (nm)
Th-T fluorescence
Normalised absorbance
Bis-ANS fluorescence
E
200 nm
F
0 5 15 30
G
Absorbance
Wavenumber (cm-1)

Figure 5.7. Characterisation of wt α-syn and α-synCC aggregates formed after 5 days-incubation in buffer A, B and C. (A) Static light scattering of 10 μM wt α-syn (dashed line) and α-synCC aggregates (full line) in different buffers. (B) Fluorescence emission spectra of Th-T upon incubation with 10 μM wt α-syn and α-synCC fibrils formed in different buffers. (C) CR normalised absorbance spectra in the presence of 10 μM α-syn aggregates. Free CR absorbance spectrum is represented in grey. (D) Bis-ANS fluorescence spectra of 10 μM bis-ANS bound to 15 μM wt α-syn and α-synCC aggregates in different buffers. (E) TEM micrographs of negatively stained aggregates formed by wt α-syn (left panel) and α-synCC (right panel) in three different conditions. Scale bars represent 200 nm. (F) PK degradation patterns of soluble wt α-syn and α-synCC (panels without frames), together with the digestion pattern of the aggregated forms (framed panels), monitored over time on Coomassie-stained SDS–PAGE. The time of digestion (in min) is indicated on the top of each panel. (G) The secondary structure of wt α-syn and α-synCC aggregates formed after 5 days. ATR-FTIR absorbance spectra in the amide I region was acquired (thick lines) and the fitted individual bands after Gaussian deconvolution are indicated (thin lines).

At the different time point, 14 days of incubation in the indicated conditions, the wt α-syn and α-synCC assemblies formed in three different buffers were characterised in the same manner (Fig. 5.8A; Fig. 5.8B; Fig. 5.8C and Fig. 5.8D). Overall, in buffer A, the amyloid-like signature is always higher for wt α-syn than for α-synCC. Even after 14 days, α-synCC exhibits low light scattering, does not bind Th-T or CR, and maintains lower bis-ANS binding than wt α-syn.

Table 5.2. Assignment of secondary structure components of wt α-syn and α-synCC after 14 days-incubation in described conditions.



wt α-syn

	Buffer A			Buffer B			Buffer C		
	Band (cm^{-1})	Area (%)	Structure	Band (cm^{-1})	Area (%)	Structure	Band (cm^{-1})	Area (%)	Structure
1	1628	44	β –sheet (inter)	1627	41	β –sheet (inter)	1625	46	β –sheet (inter)
2	1648	23		1647	28		1647	29	
3	1665	23	Loop/β-turn/bend/ α-helix	1664	21	Loop/β-turn/bend/ α-helix	1665	19	Loop/β-turn/bend/α-helix
4	1684	9	B-turn	1681	10	B-turn	1680	6	B-turn

α-synCC

	Buffer A			Buffer B			Buffer C		
	Band (cm^{-1})	Area (%)	Structure	Band (cm^{-1})	Area (%)	Structure	Band (cm^{-1})	Area (%)	Structure
1	1628	16	β –sheet (inter)	1628	41	β –sheet (inter)	1631	41	β –sheet (inter)
2	1644	33		1650	30		1652	30	
3	1661	33	Loop/β-turn/bend/ α-helix	1667	20	Loop/β-turn/bend/ α-helix	1667	19	Loop/β-turn/bend/α-helix
4	1680	18	B-turn	1681	9	B-turn	1680	10	B-turn

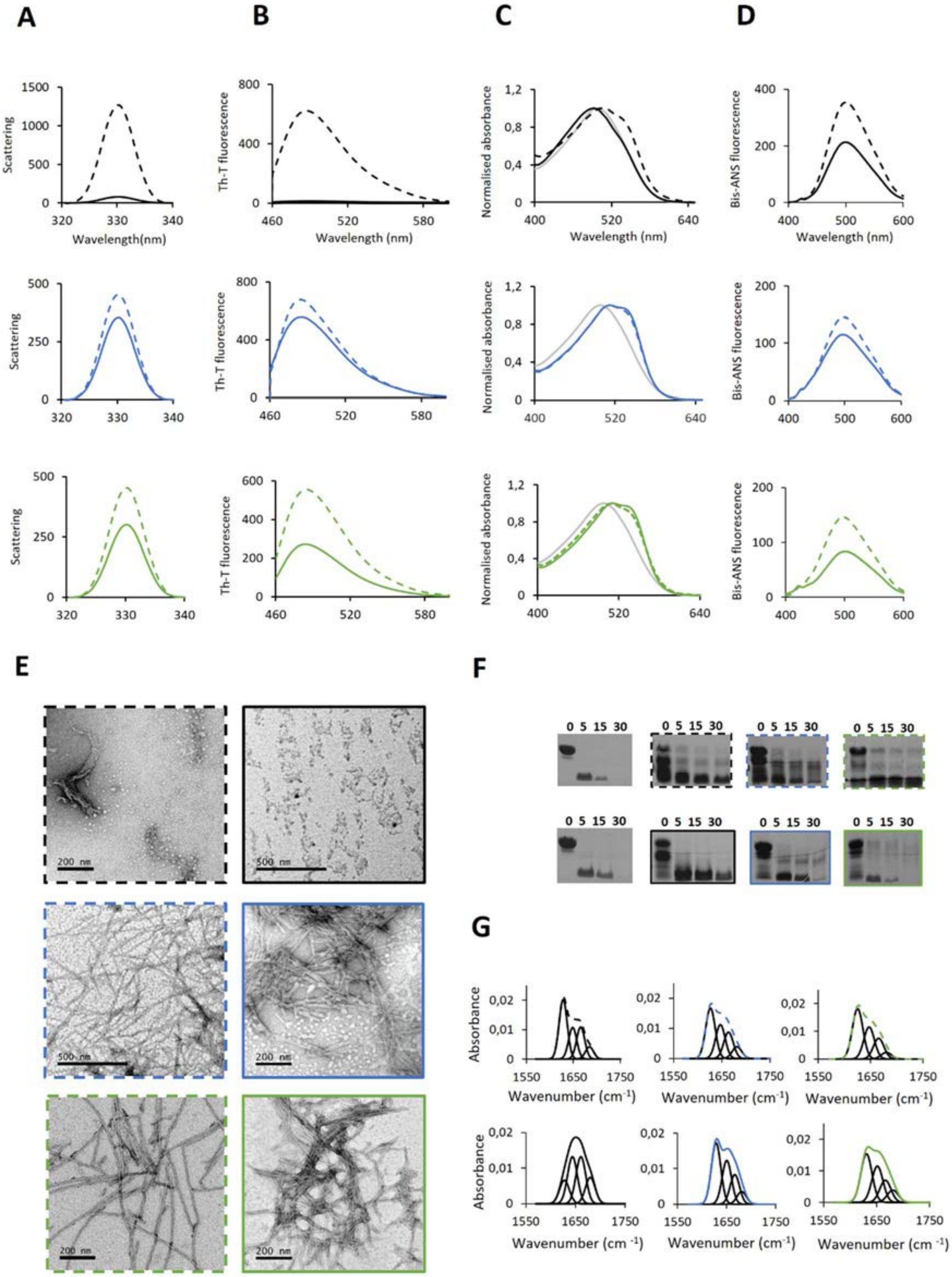

A
B
C
D
Scattering
Wavelength(nm)
Th-T fluorescence
Wavelength (nm)
Normalised absorbance
Bis-ANS fluorescence
E
200 nm
500 nm
F
0 5 15 30
G
Absorbance
Wavenumber (cm-1)
1550
1650
1750

Figure 5.8. Characterisation of wt α-syn and α-synCC aggregates formed after 14 days-incubation in buffer A, B and C. (A) Static light scattering of 10 µM wt α-syn (dashed line) and α-synCC aggregates (full line) in different buffers. (B) Fluorescence emission spectra of Th-T upon incubation with 10 µM wt α-syn and α-synCC fibrils formed in different buffers. (C) CR normalised absorbance spectra in the presence of 10 µM α-syn aggregates. Free CR absorbance spectrum is represented in grey. (D) Bis-ANS fluorescence spectra of 10 µM bis-ANS bound to 15 µM wt α-syn and α-synCC aggregates in different buffers. (E) TEM micrographs of negatively stained aggregates formed by wt α-syn (left panel) and α-synCC (right panel) in three different conditions. Scale bars represent 200 nm. (F) PK degradation patterns of soluble wt α-syn and α-synCC (panels without frames), together with the digestion pattern of the aggregated forms (framed panels), monitored over time on Coomassie-stained SDS–PAGE. The time of digestion (in min) is indicated on the top of each panel. (G) The secondary structure of wt α-syn and α-synCC aggregates formed after 14 days. ATR-FTIR absorbance spectra in the amide I region was acquired (thick lines) and the fitted individual bands after Gaussian deconvolution are indicated (thin lines).

In neutral pH- buffer B, longer time of incubation renders fibrils of wt α-syn and α-synCC that display similar light scattering, Th-T and CR binding. In the presence of salt, wt α-syn fibrils maintain higher amyloid-like signature than α-synCC, independently of the used spectral probe.

TEM analysis (Fig. 5.8E) demonstrates amyloid fibrils of wt α-syn formed in all three buffers, and fibrils formed by α-synCC in buffer B and C. Interestingly, the gels showing PK resistance of the aggregates formed in different conditions after 14 days, exhibit in all conditions the emergence of bands with a molecular weight (MW) lower than full-length wt α-syn (14460.0 Da) even before PK treatment (Fig. 5.8F). This observation was reported by other groups and analysis of these bands revealed that they arise from autoproteolysis in the N- and C-terminal, originating bands with approximately 12 and 10 kDa, respectively (Vlad *et al.*, 2011). This differential trait suggests that α-syn autoproteolytic activity depends on the aggregates morphology. Interestingly, in our experiment, the amyloid containing aggregates showed a most prominent band at 10 kDa. In all cases, the digestion resulted in a specific digestion patterns, indicating that aggregates formed by wt α-syn and its disulfide bond forming variant α-synCC contain different PK resistant cores. ATR-FTIR spectra in all cases, except in case of α-synCC in a low pH-buffer, are dominated by a peak at 1625 cm^{-1}, attributable to the presence of amyloid-like intermolecular β-sheet structure (Fig. 5.8G, Table 5.2.).

5.2.6. Conformational and toxic properties of α-syn and α-synCC oligomers

The formation of wt α-syn and α-synCC oligomers was determined by monitoring the Th-T fluorescence, light scattering, bis-ANS binding and secondary structure content by ATR-FTIR. In the last few years, soluble oligomers rather than insoluble amyloid fibrils were indicated as the main cytotoxic species involved in the pathogenesis of PD and other synucleinopathies

(Roberts and Brown, 2015; Lashuel *et al.*, 2013). The fact that it has been shown that the clinical symptoms of PD are manifested before the appearance of fibrillar deposits (Goldberg and Lansbury, 2000) and that *in vitro*, oligomers are more toxic than fibrils (Winner *et al.*, 2011), are contributing to this hypothesis.

We analysed the toxicity of oligomeric assemblies of wt α-syn and α-synCC generated after 21 h of incubation in 20 mM HEPES, pH 7.4, on cultured neuroblastoma cells from the SH-SY5Y cell line. In agreement with previous reports, the oligomers formed by wt α-syn were toxic (Avila *et al.* 2014). However, the oligomers formed by α-synCC did not have any effect on the viability of neuroblastoma cells (Fig. 5.9A). Furthermore, to explore if there are differences in morphology of oligomers, we analysed the morphological and conformational properties of both oligomeric assemblies formed at this time point. As confirmed by TEM (Fig. 5.9B), oligomers appeared as spheroidal polydisperse aggregates, with diameters ranging from 6 to 10 nm. In case of α-synCC, larger oligomeric forms could be observed. Moreover, α-syn oligomers did not significantly enhance Th-T fluorescence or exhibited differences in light scattering (Fig. 5.9C).

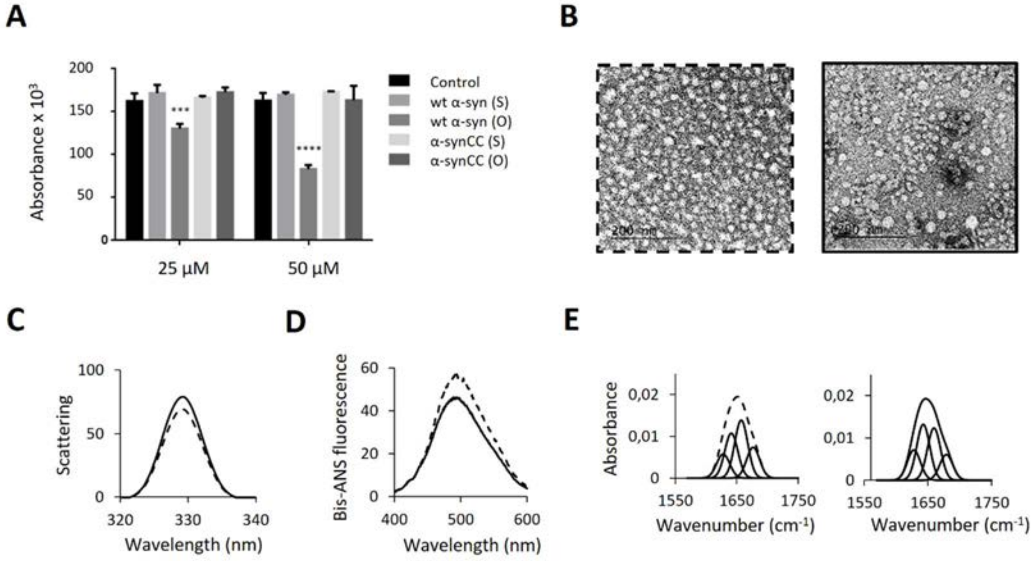
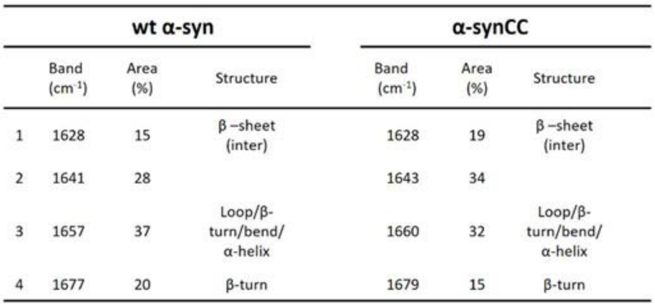


Figure 5.9. Evaluation of wt α-syn and α-synCC toxicity. (A) The cytotoxicity assay was used to estimate cell viability of SH-SY5Y neuroblastoma cells after 72 h-incubation in the presence of different concentrations of soluble (S) and oligomeric (O) forms of wt α-syn and α-synCC. Error bars indicate ± SE (n = 3). (B) Electron micrographs of oligomeric assemblies of wt α-syn (dashed line) and α-synCC (full line) generated after 21 h of incubation in 20 mM HEPES, pH 7.4. The scale bars in the electron micrographs correspond to 200 nm. (C) Static light scattering of 10 µM wt α-syn (dashed line) and α-synCC (full line) oligomeric assemblies. (D) Bis-ANS fluorescence spectra of 10 µM bis-ANS bound to 15 µM wt α-syn and α-synCC oligomers in 20 mM HEPES, pH 7.4. (E) The secondary structure of wt α-syn (*left panel*) and α-synCC oligomers (*right panel*) formed after 21 h incubation in indicated buffer. ATR-FTIR absorbance spectra in the amide I region was acquired and the fitted individual bands after Gaussian deconvolution are indicated (dashed or full line).

The differences in bis-ANS binding are not significant either, even though it has been reported for different and unrelated proteins that the binding to ANS-like dyes correlates with the toxicity of amyloid species, suggesting that the exposure of hydrophobic regions is a critical characteristic of these pathogenic assemblies (Fig. 5.9D).

The ATR-FTIR spectra of wt α-syn and α-synCC oligomers in the amide I region is showing a low content of β-sheet conformations characterised by the band appearing at 1615–1640 cm^{-1} (Fig. 5.9E). The proportion of amyloid-like intermolecular β-sheet secondary structure is higher in the oligomers formed by α-synCC (19%) than in those formed by wt α-syn (15%) (Table 5.3). Overall, these data suggest that the exposure of NAC region in α-syn molecule is not critical in the early stages of amyloidogenic pathway since at the same time point both wt α-syn and α-synCC form oligomeric assemblies that do not exhibit significant differences in their structural properties. However, the constriction of NAC region in α-synCC renders non-toxic oligomeric assemblies.

Table 5.3. Assignment of secondary structure components of wt α-syn and α-synCC oligomeric assemblies generated after 21 h of incubation in 20 mM HEPES, pH 7.4.



	wt α-syn			α-synCC		
	Band (cm^{-1})	Area (%)	Structure	Band (cm^{-1})	Area (%)	Structure
1	1628	15	β –sheet (inter)	1628	19	β –sheet (inter)
2	1641	28		1643	34	
3	1657	37	Loop/β-turn/bend/ α-helix	1660	32	Loop/β-turn/bend/ α-helix
4	1677	20	β-turn	1679	15	β-turn

5.2.7. Early-stage aggregation kinetics of wt α-syn and α-synCC in the presence of lipids

Next, we monitored how the disulfide bond formation in α-synCC impacted the aggregation kinetics in three different buffers by continuously monitoring the changes in Th-T binding over time for wt α-syn and α-synCC (Fig. 5.10). The kinetics of amyloid fibril formation for wt α-syn follow a sigmoidal curve that reflects a nucleation-dependent growth mechanism. However, the aggregation of α-synCC in all three buffers does not follow this pattern. The aggregation reactions of wt α-syn and in α-synCC were shown to diverge significantly in each buffer, with an exponential increase of Th-T fluorescence for wt α-syn in all conditions, and a steady and much slower increase for α-synCC, reaching more than 2-times lower fluorescence intensity in buffer A (Fig. 5.10A), more than 5-times lower intensity in buffer B (Fig. 5.10B), and more than 6-times lower intensity in buffer C (Fig. 5.10C) compared with the wt α-syn in the same conditions. Altogether, these data demonstrate that the formation of disulfide bond that constricts the NAC region in α-synCC reduces amyloid formation by α-synCC in all three conditions *in vitro.* This observation does not fit with the report from Shaykhalishahi *et a*l. in which the low hydrophobicity (GRAVY score = 0.411) of the loop closed by covalent disulfide bond in the engineered α-synCC(37-54), was indicated as a determinant that influenced this α-synCC variant to be unable to form either oligomers or fibrils, since in our case oxidized α-synCC variant has a loop closed having a similar hydrophobicity (GRAVY score = 0.745) but is capable to form oligomers and amyloid fibrils in different solutions.

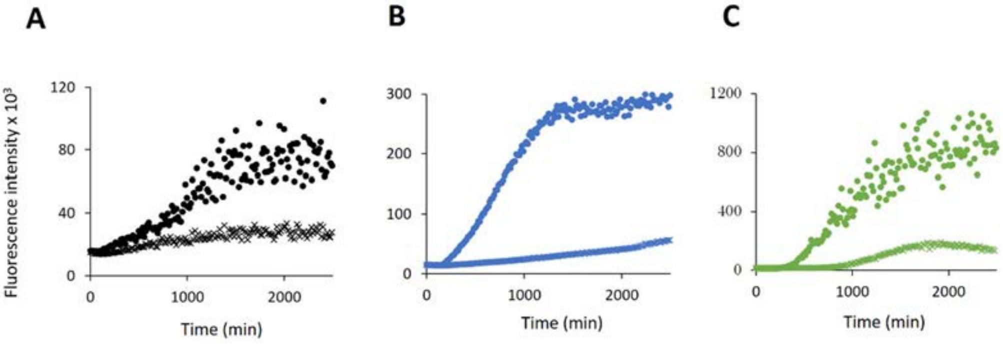


Figure 5.10. Aggregation kinetics of α-syn in buffer A, B and C. Change in Th-T fluorescence intensity when 70 µM monomeric wt α-syn and α-synCC were incubated under quiescent conditions at 37 °C, in (A) buffer A, (B) buffer B, and (C) buffer C.

α-syn exists in a dynamic equilibrium between a soluble state and a membrane-bound state and the interaction between α-syn and lipid surfaces is believed to be a key feature for mediating its cellular functions. It is known that the amphipathic N-terminus of α-syn plays a role in membrane interaction allowing lipid-induced coil→helix conformational transitions of the protein (Bartels *et al*., 2010), an unstructured C-terminal region is weakly associated with the membrane and the NAC region has been implicated as a sensor of the lipid properties that determines the affinity of α-syn membrane binding (Fusco *et al*., 2014). The differences in the binding and membrane interaction for wt α-syn and its double cysteine mutant α-synCC, in which NAC region has been constricted due to the disulfide bridge between mutated residues, were determined by following the aggregation kinetics in neutral pH-buffer in the presence and in the absence of lipids, together with the circular dichroism (CD) spectroscopy analysis of proteins in the presence of different types of lipids at the beginning of the reaction, prior to aggregation experiments (Fig. 5.11A and Fig. 5.11B). Unilamellar vesicles of varying net charge and lipid compositions served as model membranes. The results suggest that in case of wt α-syn, the incubation in the presence of neutral liposomes (DOPC) and neural liposomes (DMPS) with protein:lipid concentration ratio 1:1, resulted in the steady and slower aggregation reactions compared to the wt α-syn incubated alone, where at around 24 h plateau has been reached (Fig. 5.11B).

In case of double cysteine mutant α-synCC, the incubation of protein in the presence of unilamellar vesicles resulted in a significant retardation of the aggregation kinetics (Fig. 5.11B). Assemblies generated at the final point of the monitored aggregation kinetics were analysed by TEM (Fig. 5.11B; *lower panel*). wt α-syn without vesicles forms mature long unbranched fibrils while the fibrils formed by α-synCC resemble protofibrils. In the presence of DOPC the fibrils become thinner but well structured in case of wt α-syn, which is in contrast with disordered fibrils formed by α-synCC. In the presence of neural liposomes, both wt α-syn and its disulfide bond forming variant form fibrils that seem to be disordered.

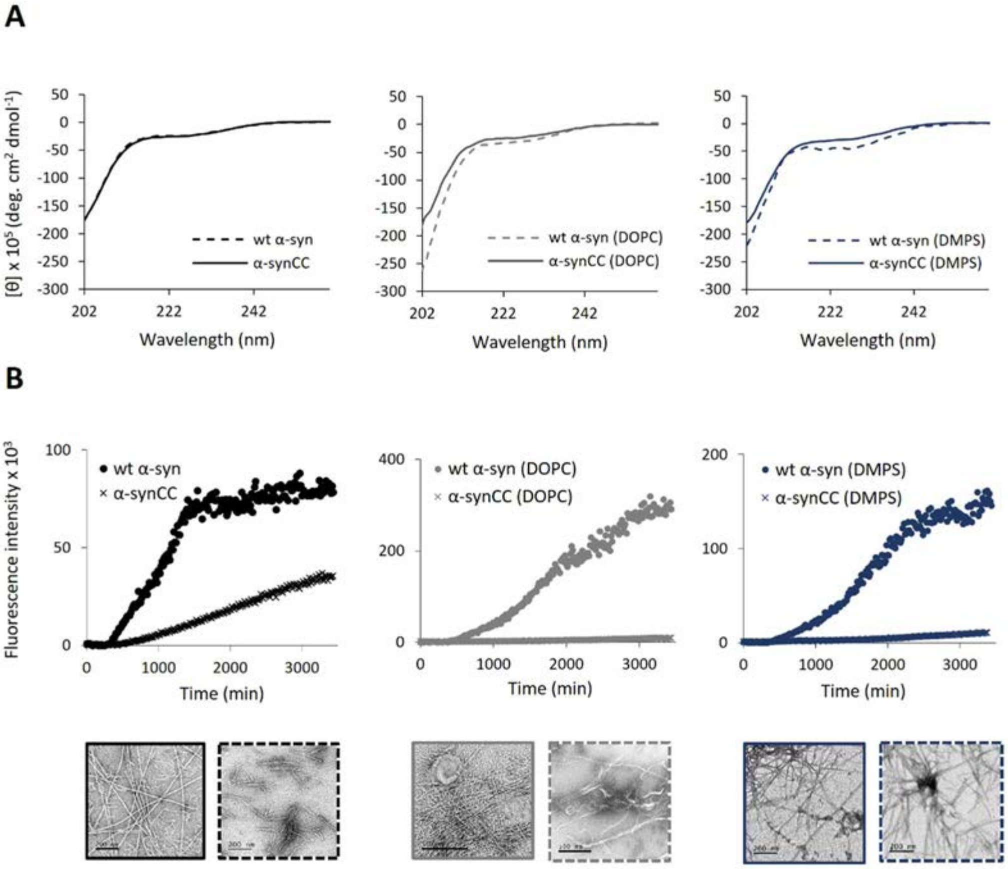


Figure 5.11. α-syn aggregation in the presence of lipids. (A) CD spectra of 15 µM wt α-syn and α-synCC in the presence of lipid vesicles are shown: α-syn alone (black, *left panel*); with neutral liposomes (DOPC, light grey, *middle panel*); with neural-like liposomes (DMPS, dark blue, *right panel*) with protein:lipid concentration ratio 1:1. (B) Change in Th-T fluorescence intensity when 70 µM wt α-syn and α-synCC were incubated in the absence (black; *left panel*) and in the presence of 70 µM DOPC (grey; *middle panel*) or DMPS (dark blue*; right panel*) under quiescent conditions in buffer B and 37 °C, together with the corresponding electron micrographs of fibrils generated at the final point of the monitored aggregation kinetics (*lower panel*). The scale bars in the electron micrographs correspond to 200 or 500 nm.

5.2.8. The inhibitory effect of α-synCC on the aggregation of wt α-syn

To test the effect of α-synCC on wt α-syn, fibrillization reactions of wt α-syn were performed in the absence and presence of increasing concentration of α-synCC and monitored in a fluorescence microplate reader (Fig. 5.12A and Fig. 5.12B).

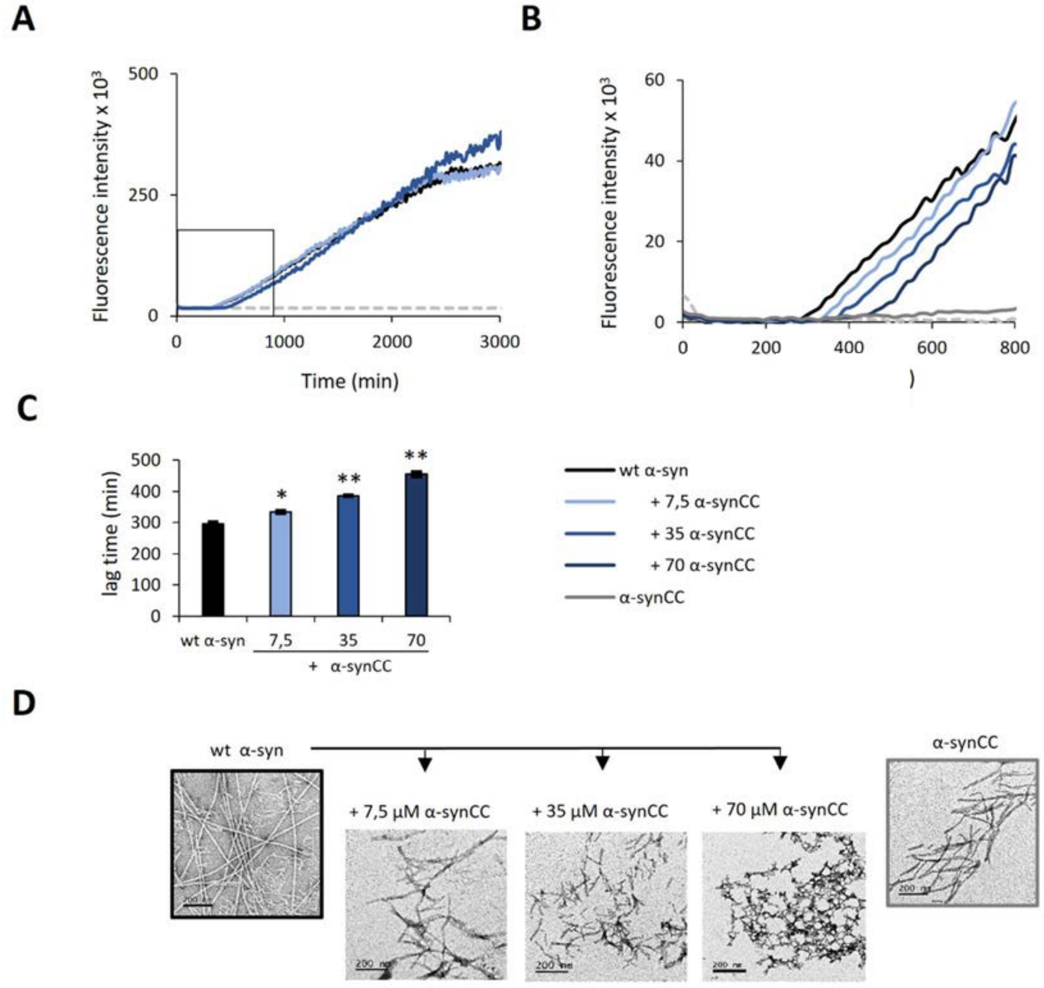


Figure 5.12. wt α-syn aggregation in the presence of α-synCC. (A) Th-T time course of fibrillization of wt α-syn in buffer B, in the absence (black line) and in the presence of increasing concentrations of soluble α-synCC (lines coloured in a gradient of blue). (B) The first 800 min showing the aggregation lag phase for the indicated conditions. (C) Bar diagram representation of lag time of the aggregation kinetics in the indicated conditions. Error bars indicate ± SE (n = 4). (D) The corresponding electron micrographs of amyloid fibrils generated at the final point of the monitored aggregation kinetics. The scale bars in the electron micrographs correspond to 200 or 500 nm.

α-synCC inhibited the aggregation of wt syn both at equivalent and at substoichiometric ratios only in the early stages of the aggregation (Fig. 5.12B), significantly increasing lag time of the aggregation reaction (Fig. 5.12C), indicating that α-synCC changes the aggregation pathway of wt α-syn and interferes with the elongation of wt α-syn fibrils. However, at the final stage of fibrillization reactions, observable both by aggregation kinetics and by TEM analysis of the assemblies formed at the end of the monitored aggregation, the formed wt α-syn fibrils become disordered with an increasing concentrations of α-synCC, rendering amorphous aggregates at the wt α-syn:α-synCC concentration ratio 1:1 (Fig. 5.12D).

5.3. Material and methods

Design of α-syn disulfide bond

The rational design of the disulfide bond in α-syn (α-synCC) was performed using Disulfide by Design 2.0 (Craig and Dombkowski, 2013). Two residues within NAC region of α-syn, Val71 and Thr92, were mutated to cysteines to form a disulfide bridge. The stability of the mutated variant α-synCC was assessed using FoldX (Schymkowitz *et al*., 2005). The atomic spatial coordinates of wt α-syn in the solved crystal structure of a pathogenic fibril of full-length human α-syn (PDB: 2N0A) were used as input.

Analysis of aggregation propensity, disorder, and hydrophobicity of α-syn sequences

The sequential analyses were performed using web-based algorithms for predicting aggregation tendencies such as AGGREGATOR (Conchillo-Sole *et al*., 2007), Zyggregator (Tartaglia *et al*., 2008) and FoldAmyloid (Garbuzynskiy et al., 2010). All predictions were performed using default settings. Intrinsic disorder for both wt α-syn and α-synCC was analysed using IUPred (Dosztanyi *et al*., 2005) and RONN (Yang *et al*., 2005), by implementing an 11-residue sliding window. The hydrophobicity values according to the scale of Kyte and Doolittle (Kyte and Doolittle, 1982) were calculated considering a 15-residue sliding window.

Expression and purification of wt α-syn and α-synCC

pET21a vector (Novagen) encoding for wt α-syn was transformed into *E.coli* BL21(DE3) cells and the same vector encoding for α-synCC was transformed into *E.coli* Origami(DE3) cells. For protein expression, 10 ml overnight culture of transformed cells was used to inoculate 1 L of LB medium with 100 µg/mL ampicillin and allowed to grow under agitation 250 rpm at 37 °C. At an OD_{600} 0.6, protein expression was induced with 1 mM of isopropyl-1-thio-β-D-galactopyranoside (IPTG) for 4 h at 37 °C. Then, the cultures were centrifuged and cell pellet frozen at -80 °C.

The protocol for α-syn purification was adapted from (Volles and Lansbury, 2007). Pellets were resuspended in 15 mL of lysis buffer (50 mM Tris·HCl, 150 mM NaCl, 1 µg/mL Pepstatin A, 20 µg/ml Aprotinin, 1 mM Benzamidine, 1 mM PMSF and 1 mM EDTA) and sonicated on ice. The lysate was boiled for 10 min at 95 °C and soluble and insoluble fraction were separated by centrifugation at 48,384 x g, for 15 min. 136 µL/mL of 10% streptomycin sulfate and 228 µL/mL of glacial acetic acid was added to the supernatant. The resulting solution was centrifuged for 5 min at 48,384 x g, and the supernatant was collected and precipitated with saturated ammonium sulfate at 4 °C in a ratio 1:1 (v/v). Further, the pellets were washed with a 1:1 (v/v) solution of ammonium sulfate and water. Next, pellets were resuspended in 900 µL of 100 mM ammonium acetate and the same volume of 100% ethanol was added to precipitated α-syn.

Ethanol precipitated α-syn was resuspended in starting buffer (25 mM Tris·HCl, pH 8.0) and filtered through a Millex-HP filter syringe-driven 0.45 µm filter unit (Millipore). Anion exchange high-performance liquid-chromatography was carried out on an AKTA-FPLC (GE Healthcare). The sample was loaded, bounded to Hi-Trap column (GE Healthcare) and eluted with a NaCl linear gradient of elution buffer (25 mM Tris·HCl at pH 8.0, 1 M NaCl). The fractions containing α-syn were collected and the buffer was exchanged for 20 mM ammonium acetate. The identity and purity of the recombinant proteins were assessed by MALDI-TOFF on an Autoflex Speed mass spectrometer (Bruker, Bremen) and SDS-PAGE electrophoresis.

Mass spectrometry

To prepare samples for MALDI mass spectrometry, lyophilised wt α-syn and α-synCC were resuspended in 50 mM ammonium bicarbonate. The molecular mass of samples, with or without a previous step of reduction with 500 mM dithiothreitol (DTT), and after derivatization with iodoacetic acid was determined with a MALDI-TOF UltrafleXtreme spectrometer (Bruker Daltonics) using 2',6'- dihydroxyacetophenone acid as a matrix. Samples and matrix were mixed to 1:1 ratio and 1 µL of the mixture was applied on a ground steel plate. The analyses were calibrated using external calibrants (Bruker Daltonics).

ESI-MS analysis

Mass spectra have been acquired on a hybrid triple quadruple-TOF instrument (Applied Biosystems QSTAR Elite) equipped with a nano-electrospray source. The samples (10 µL of 10 µM protein in 20 mM ammonium acetate pH 7.0) have been injected by borosilicate-coated capillaries with 1 µm internal diameter (Thermo Fisher Scientific). The α-synCC without the disulfide bridge has been obtained adding DTT to the sample solution to a final concentration of 20 mM and incubating 2 h at room temperature before the injection. The main instrumental parameters for spectra acquisition have been: ion spray 1.1 kV; declustering potential 60 V; curtain gas 20 PSI. Final spectra have been averaged over 1 minute acquisition time.

Circular dichroism (CD)

The stocks of wt α-syn and α-synCC were prepared by dissolving lyophilised protein in buffer A (50 mM sodium acetate pH 5.0), buffer B (10 mM sodium phosphate pH 7.0), and buffer C (10 mM sodium phosphate pH 7.0 supplemented with 200 mM NaCl). After filtering stocks through 0.2 µm filter units, protein was prepared to a final concentration of 15 µM in the corresponding buffers. Far-UV CD spectra were acquired in a thermostated spectropolarimeter Jasco-715. Spectra were recorded in a range from 260 to 200 nm with 1 nm bandwidth, and a scan speed of 100 nm/min. Average of ten accumulations were recorded for each spectrum.

Aggregation assay

From the prepared stocks of wt α-syn and α-synCC in buffer A, B and C, four aliquots of 300 µL of protein in each buffer were prepared to a final concentration of 60 µM. Samples were incubated in Eppendorf Thermomixer Comfort (Eppendorf) with 0.02% sodium azide at 600 rpm and 37 °C. The aggregation assays were performed in four replicates and up to two weeks.

Light scattering assay

The transition of α-syn from initial soluble form to aggregated state was determined by measuring light scattering at 328 nm in a Jasco FP-8200 spectrofluorometer (Jasco Inc.) with an excitation wavelength of 330 nm and emission range from 320 to 340 nm at 25 °C. Final protein concentration was 10 µM in the corresponding buffers. Buffers without protein were used as negative controls. The experiments were carried out in four replicates.

Thioflavin (Th-T) binding assay

The binding of Th-T to amyloid fibrils formed after 5 and 14 days of incubation, under described conditions, was measured using a JASCO FP-8200 spectrofluorometer (Jasco Inc.) with an excitation wavelength of 445 nm and emission range from 460 to 600 nm at 25 °C. The intensity of the spectra at the 488 nm maximum was recorded as an indication of the β-sheet structures in the aggregates. Protein samples were prepared to a final concentration of 10 µM in corresponding buffers, and supplemented with 25 µM Th-T. Buffers without protein were used as negative controls. All measurements were carried out in four replicates.

Congo Red (CR) binding assay

CR binding to amyloid fibrils was determined using a Cary-400 Varian spectrophotometer (Varian Inc.) by monitoring the absorbance spectra from 375 to 675 nm using a quartz cuvette of 1 cm optical length placed in a thermostated cell holder at 25 °C. Final CR and protein concentration were 10 µM in a native buffer. Same buffers without protein and solutions without CR were used for the baseline. All measurements were carried out in four replicates.

Bis-ANS binding assay

The fluorescence emission spectra of bis-ANS were determined by incubating soluble protein samples and those aggregating for 5 and 14 days at 15 µM in 10 µM bis-ANS. Using a JASCO FP-8200 spectrofluorometer (Jasco Inc.) the samples were excited at 370 nm and emission measured between 400 and 600 nm with slit widths of 5 nm.

Limited proteolysis

Protein samples were resuspended in corresponding buffers at a final concentration of 50 µM and incubated with 2.5 µg/mL proteinase K (PK) (Sigma-Aldrich) at 37 °C. After indicated time points during the digestion reaction, proteolytic aliquots were taken and the reactions were quenched by the addition of the same amount of 4 times concentrated denaturing sample buffer. Samples were heated at 99 °C for 10 min, and 10 µL of each sample was loaded on SDS-PAGE and analysed. Soluble protein in buffer B was digested in the same manner and used as a control.

Transmission Electron Microscopy (TEM)

For negative staining, wt α-syn and α-synCC incubated for 5 and 14 days in buffer A, B and C, were diluted in Mili-Q water to a final concentration of 20 µM. Prepared samples were adsorbed onto carbon-coated copper grids, washed with water and stained with 2% (w/v) uranyl acetate. The morphology of aggregates of wt α-syn and α-synCC was observed using a Joel 1400 transmission electron microscope at an accelerating voltage of 80 kV.

Secondary structure determination

ATR-FTIR spectroscopy analysis of amyloid fibrils was performed using a Bruker Tensor 27 FTIR Spectrometer (Bruker Optics Inc.) with a Golden Gate MKII ATR accessory. Incubated samples were centrifuged and the insoluble fractions were resuspended in Mili-Q water. Each spectrum consists of 16 independent scans, measured at a spectral resolution of 4 cm^{-1}, within the 1800-1500 cm^{-1} range. Second derivatives of the spectra were used to determine the frequencies at which the different spectral components were located. Fourier-deconvolution and determination of band position of the original amide I' band were performed using PeakFit software (Systat Software).

Preparation and analysis of oligomer-rich samples of wt α-syn and α-synCC

The oligomeric assemblies of wt α-syn and α-synCC were prepared following the reported method (Avila *et al.* 2014). The formation of wt α-syn and α-synCC oligomeric assemblies was confirmed by TEM analysis. The oligomers were structurally characterised by light scattering, bis-ANS binding and ATR-FTIR spectroscopy. Experiments were carried out in four replicates.

Cell viability assay

Human SH-SY5Y neuroblastoma cells cultured in F-12 medium supplemented with 10% FBS, were seeded in 96-well plates at a density of 4000 cells/well and maintained at 37 °C. Cell cultures were incubated for 72 h in the presence and in the absence (control) of 25 µM and 50 µM soluble oligomeric assemblies of wt α-syn and α-synCC. Then, 20 µL of PrestoBlue® Cell Viability Reagent (Invitrogen) was added to treated cells and fluorescence was determined after incubation for 45 min at 37 °C, recording fluorescence at 615 nm, with an excitation wavelength of 531 nm in a Victor fluorescent plate reader (Perkin Elmer).

Aggregation kinetics

Aggregation of 70 µM wt α-syn and α-synCC in buffer A, B and C was monitored by measuring the transition from non-aggregated to aggregated state, according to the Th-T fluorescence at 486 nm measured on a 96-wells microplate reader (Victor Microplate reader, Perkin Elmer) for 72 h at 37 °C. Experiments were carried out in four replicates.

Preparation of lipid vesicles

Dioleoylphosphatidylcholine (DOPC) vesicles and dimyristoylphosphatidylserine (DMPS) vesicles (50:30:20 DOPE/DOPS/DOPC) (Avanti Polar Lipids), a lipid composition suggested to mimic the composition of synaptic vesicles, were prepared by the lipid hydration and extrusion method (MacDonald *et al*., 1991). Briefly, stock lipids in chloroform were thoroughly mixed in a round glass tube. Organic solvents were removed to yield a lipid film using a dry nitrogen stream. The dried lipids were then hydrated in buffer B. Multilamellar liposomes (MLV) were formed by three cycles of vigorously mixing and sonication. MLV were extruded 20 times through 200 nm polycarbonate membranes (Avanti Polar Lipids) in an extruder device to obtain small unilamellar vesicles (LUV). The size of the lipid vesicles was determined by dynamic light scattering in order to have a homogenous sample.

Furthermore, far-UV CD spectra were acquired in a spectropolarimeter Jasco-715 using 15 µM protein samples resuspended in buffer B either in the presence or absence of 50 µM DOPC or DMPS vesicles. Spectra were recorded in a range from 260 to 200 nm with 1 nm bandwidth, and a scan speed of 100 nm/min. Average of ten accumulations were recorded for each spectrum.

Aggregation kinetics in the presence of lipid vesicles

The aggregation kinetics of wt α-syn and α-synCC in the presence of LUVs was monitored with time by measuring the change in the Th-T fluorescence signal at 486 nm, using a 96-wells microplate reader (Victor Microplate reader, Perkin Elmer). Lyophilized protein was prepared in buffer B to a final concentration of 70 µM, supplemented with 10 µM Th-T, and 70 µM DOPC or DMPS vesicles. The prepared 96-well plate was incubated under agitation at 37 °C for 72 h. Experiments were carried out in four replicates.

RP-HPLC analysis

α-synCC was resuspended in 50mM Tris pH 8.4 supplemented with 100 mM NaCl buffer to a final concentration of 0.5 mg/mL, with and without 100 mM DTT, and analysed by RP-HPLC in a Waters 2690 HPLC system. Protein resuspended in DTT was incubated for 2 h and both samples were mixed with TFA 2% (v/v) to freeze the reaction before the injection. A linear gradient from 30% to 50% of buffer B was applied to a standard C4 column (Phenomenex) 250 x 4.6 5 mm at a flow rate 0.75 mL/min. Composition of the chromatography buffers: pure water + 0.1% TFA (Buffer A) and Acetonitrile (Teknokroma) + 0.1% TFA (Buffer B).

NMR spectroscopy

Samples of α-syn wt and α-synCC were prepared at a final concentration of 160 µM in 20 mM HEPES pH 7.4, and further diluted using a 9:1 H2O/D2O ratio. One-dimensional NMR spectra were collected at increasing temperature with 10 min interval time on a Bruker 600 MHz - Avancell spectrometer using solvent suppression WATERGATE. Accumulation of 32 scans was recorded at a spectral resolution of 0.58 Hz. Spectra were processed and analyzed with Topspin2.0 software packages (Bruker GmbH).

The 2D-TOCSY pulse sequence used was solvent suppression dipsi2etgpsi19. DIPSI-2 TOCSY mixing sequence applied was 80 ms at a spectral resolution of 6 x 96 Hz (F2 x F1) (8.3 kHz). Accumulation of 32 scans was recorded in a total acquisition time 80 min. DOSY NMR experiments were performed with a pulsed-field gradient stimulated echo (ledbpgppr2s) using bipolar gradients. The spectral resolution was 0.46 Hz. The evolution of the pulsed-field gradient during the NMR diffusion experiments was established in 16 steps, applied linearly between 2 and 98%. Each NMR diffusion experiment was obtained from the accumulation of 128 scans.

6. Discussion

There are multiple approaches that have been implemented during the last decades with the aim to study aggregation reactions of different protein models. Accordingly, it this thesis, a combination of biophysical and computational methods, together with the use of *S. cerevisiae* as *in vivo* experimental model, has assisted us to reach the understanding about molecular determinants of protein aggregation.

In *Chapter 1.*, we have characterised the aggregation properties of a significant fraction of *E. coli* protein structures. The data we have obtained indicate that structural and sequential aggregation propensities of bacterial proteins are somehow connected but do not necessarily experiment the same pressure to avoid aggregation, and that selective pressure against protein aggregation plays an important role in shaping proteins conformational space. There is an observable trade-off between protein functionality and structural aggregation propensities (SAP) in such a way that the proteins SAP is linked to their abundance, their size, essentially, cellular location, molecular function, secondary structure, synthesis rate, thermodynamic stability and the monomeric/multimeric form in which they are active. We usually assume that protein structures and stabilities have evolved to perform their functions efficiently in the cell; this premise continues being valid, but in order to attain data evolution has also to sculpt their SAP. An implication of this link between SAP and function is that the lessons learned from our analysis might turn to be usually to rationally manipulate and control the production of soluble and functional recombinant proteins in bacteria.

In *Chapter 2.*, we used yeast as a model to systematically analyse the intracellular effect of expressing 21 variants of the amyloid-β-peptide, engineered to cover a continuous range of intrinsic aggregation propensities when expressed in the yeast cytosol. We were able to demonstrate the existence of a striking negative correlation between the aggregation propensity of a given variant and the oxidative stress it elicits. Recent studies have demonstrated that the aggregation process can result in a high diversity of protein conformational states, however the identity of the specific species responsible for the cellular damage is still unclear. Interestingly, each amyloid-β-peptide variant generated a specific distribution of protein assemblies in the cell. This allowed us to identify the aggregated species that remain diffusely distributed in the cytosol and are unable to coalesce into large protein inclusions as those causing the highest levels of oxidative damage. It is likely that, according to their aggregated and protease-resistant nature, these species correspond to oligomeric and/or protofibrillar assemblies (Deas *et al.*, 2016). Our data are consistent with *in vitro* assays on Aβ40 showing that maximum production of free radicals occurs early in the aggregation reaction, where these small assemblies are maximally populated (Tabner *et al.*, 2005).

Moreover, our data provide evidence for a causal link between the accumulation of diffuse aggregates and cellular oxidative stress and suggest that protein inclusion (PI) formation acts as a second line protective strategy against the intracellular oxidative stress caused by these species. These findings are consistent with previous data for other proteins related with conformational disorders, such as poly(Q)-expanded huntingtin, ataxin-1 and α-synuclein, demonstrating a dissociation between PI formation and cell death. Our results have important implications for the therapeutics of the neurodegenerative disorders linked to protein aggregation. Approaches intended to prevent the formation of PI in the brain can convert a locally recruited inert protein into a toxic aggregated molecule disseminated around the cell, prolonging oxidative stress and the associated neurotoxicity. This is consistent with the observation that pharmacological or genetic disruption of the physiologic processes required for PI formation results in increased proteotoxicity (Tanaka *et al*., 2004). On the other way around, reducing the lifetime of protein aggregates and/or promoting their sequestration into PI appear as worth to explore therapeutic strategies for the treatment of protein aggregation associated diseases.

In *Chapter 3*., we used protein design to cross-link the Greek-key motif in human α-synuclein (α-syn) through a novel disulfide bond. Substantial evidence indicates that the aggregation of α-syn is a critical factor in the pathogenesis of Parkinson's disease (PD). In spite of a lot of research on aggregation and fibrillation of α-syn, as well as experiments aimed to estimate how it induces cytotoxicity, there is still no clear picture of the molecular determinants regulating these processes. This issue is even more exacerbated by the recent discovery that α-syn aggregation can result in different fibrillar structures displaying different toxicity and, more importantly, a differential ability to propagate within the brain. The recent solid-state NMR structure of a pathogenic fibril of full-length human α-syn (Tuttle et al., 2016) have provided important clues to understand these phenomena. An outstanding feature of this structure is the presence of steric zippers stabilizing a new orthogonal Greek-key topology that connects distant regions in the NAC α-syn domain. It has been argued that this characteristic contributes to the robust propagation of this fibril pathogenic fibril form, while providing a drugable interface. We expected that a novel engineered disulfide bond cross-linking the Greek-key motif should facilitate amyloid formation in case if this structural element plays a role at the initial stages of aggregation. Our results argue against this view and indicate that the NAC folding is an important structural element in mature fibrils. Its formation at an early stage promotes compaction but precludes the progress of the reaction and impedes the formation of toxic oligomeric species. Our data also illustrate how the solution conditions can dramatically modulate the properties α-syn protein aggregates, in which the Greek-key motif cannot exist.

7. Concluding remarks

1. A structure-based protein aggregation profile of the *Escherichia coli* proteome

    - The work confirms previously reported evolutionary selection that shaped bacterial proteome in order to reduce the probability of promiscuous interactions between highly abundant proteins, implicating the aggregation propensity of protein structures, instead of the aggregation propensity of sequences, as the feature of protein that is being constrained by evolution. This indicates that misassembly, rather than misfolding, accounts for an observed negative correlation between structural aggregation propensity and abundance.
    - The structural aggregation propensity increases with the protein length. However, this correlation is not uniform but it exists for protein below 500 Da in size, indicating that there is an aggregation limit that cannot be overpassed for very long proteins.
    - The biological functions of proteins are clearly related to their structural aggregation propensities. Moreover, the set of genes in an operon share a common gene expression regulation and are generally regulating proteins that perform similar biological functions and exhibit related structural aggregation properties. This suggests a link between protein aggregation propensities and the rates of transcriptional initiation.
    - The importance of protein length is observable for the structures of essential bacterial proteins, which suffer a stronger selection against misassembly and aggregation than those of non-essential ones.
    - Analysis of aggregation propensity of structures belonging to bacterial membrane proteins allows the visualization of high aggregation-prone fringe in the outside flanked by low aggregation-prone regions that protrude out of the membrane, giving this structures overall high solubility. In contrast, linear predictors cannot give context to the scattered aggregation-prone regions in the sequence. This duality between the aggregation properties of the sequence and the structure might be important for the biogenesis of membrane proteins.
    - The soluble structures were found to be significantly enriched in charged residues and α-helices, indicating the importance of charge−charge interactions in promoting the protein solubility by inducing long-range repulsion. On the other hand, the aggregation-prone structures were enriched in aromatic residues and β-sheets, features that facilitate the probability of competing non-functional interactions.
    - Aggregation-prone structures are significantly more stable and are synthesized at higher rates in contrast to the soluble structures that are less stable and synthesized

at lower rates. These differences in stability might reflect a safeguarding strategy to prevent that the partial unfolding might further increase the possibility of the establishment of non-functional interactions.

- There is a relationship between structural aggregation propensity and oligomeric assembly in bacteria, proteins active as oligomers exhibiting high aggregation propensities due to the protein interfaces necessary for the formation of the functional complexes. This comes at the risk of a misassembly especially in conditions that favour multimers dissociation.

2. Protein aggregation into insoluble deposits protects from oxidative stress

- The different aggregation properties of the 20 Aβ42-GFP variants expressed in *S. cerevisiae* translate into a binary ability to form protein inclusions, half of the variants forming visible aggregates and the rest remaining diffusely distributed in the cytosol. This illustrates the exquisite control that the sequence exerts on *in vivo* aggregation propensities.
- There is an observable almost perfect negative correlation between the protein aggregation potential of Aβ42-GFP variants and the oxidative stress they elicit in yeast cells. The proteins in the Aβ42-GFP set can be classified into two classes: I) variants with low aggregation propensity and high oxidative stress that activates the antioxidant defences and II) mutants with high aggregation tendency and low oxidative potential. The sequestering the most aggregation-prone variants in protein inclusions appears to be as effective to abrogate oxidative stress as their active degradation by the proteostasis machinery.
- The class II Aβ42-GFP variants, that do not form protein inclusions after 16 h of expression in yeast, are mainly distributed in the insoluble cell fractions. The protein found in the insoluble fraction that is homogeneously distributed through the cytosol corresponds to diffuse aggregates. Moreover, from this observation it becomes evident that inclusions and aggregates are not interchangeable terms.
- Engineering and expressing a highly soluble 19Asp/34Leu Aβ42-GFP variant, the soluble, likely monomeric protein, is discarded as a significant source of oxidative stress and this property can be assigned to the diffuse aggregates. It is likely that, according to their aggregated and protease-resistant nature, these species correspond to oligomeric and/or protofibrillar assemblies.
- Comparative analyses of the distribution, solubility and oxidative stress for different Aβ42-GFP variants at 16 and 48 h, suggest that when the protein aggregates are immobilized in intracellular protein inclusion, they are less dangerous. Therefore, oxidative stress can be predicted just by monitoring the levels of diffuse, but aggregated species at a certain time point.

- The placement of polar or charged residue in the position 19 of the central hydrophobic cluster of Aβ42 might disrupt this stretch and abrogate the misfolding recreation, which could explain why all the variants displaying hydrophobic mutations were recruited into PI whereas the charged and polar ones were not.

3. Role of the non-amyloid component Greek-key motif in the aggregation and cytotoxicity of human α-synuclein

- Cross-linking of the Greek-key motif in the NAC region of α-syn renders more compact and more thermally stable form. Furthermore, under reducing conditions the conformational ensemble of the generated α-synCC is similar to wt α-syn, indicating that the observed effect is due to the disulfide bridge itself and not merely to the mutation. This is further supported by the analysis of aggregation propensity and disorder of wt α-syn and α-synCC using sequential predictors, which indicate that there are no observable differences between wt α-syn and its double cysteine mutant. The data demonstrates that the disulfide bridge in α-syn leads to the loss of the fully extended conformation.
- In different buffers, wt α-syn and α-synCC exhibit differences in their aggregative potential, forming different assemblies in the same conditions. This confirms previously reported data that NAC region is implicated in the formation of α-syn strains.
- α-synCC cannot form toxic oligomers compared to wt α-syn. These data suggest that the exposure of NAC region in α-syn molecule is not critical in the early stages of amyloidogenic pathway since at the same time point both wt α-syn and α-synCC form oligomeric assemblies that do not exhibit significant differences in their structural properties. However, the constriction of NAC region in α-synCC renders non-toxic oligomeric assemblies.
- Even in the conditions which generally promote wt α-syn aggregation, such as presence of lipids, neutral (DOPC) or neural-like (DMPS) liposomes, α-synCC aggregates at the lower rate compared to the wt α-syn.
- α-synCC can interfere with the aggregation pathway of wt α-syn, most likely having an effect on the elongation/polymerization stage of wt α-syn aggregation pathway.

8. References

Aguzzi, A. and O'Connor, T. (2010). Protein aggregation diseases: pathogenicity and therapeutic perspectives. *Nat Rev Drug Discov*, 9, 237–248.

Albu, R. F. et al. (2015). A feature analysis of lower solubility proteins in three eukaryotic systems*. J Proteomics*, 118, 21-38.

Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science*, 181(4096), 223-230.

Ansari, M. A. and Scheff, S. W. (2010). Oxidative stress in the progression of Alzheimer disease in the frontal cortex. *J Neuropathol Exp Neurol*, 69, 155–167.

Apetri, M. M. et al. (2006). Secondary structure of α-synuclein oligomers: Characterization by Raman and atomic force microscopy. *J Mol Biol*, 355*,* 63-71.

Arrasate, M. et al. (2004). Inclusion body formation reduces levels of mutant huntingtin and the risk of neuronal death. *Nature*, 431, 805–810.

Ávila, C. L. et al. (2014). Structural characterization of heparin-induced glyceraldehyde-3-phosphate dehydrogenase protofibrils preventing α-synuclein oligomeric species toxicity. *J Biol Chem*, 289(20), 13838-50.

Baba, T. et al. (2006). Construction of Escherichia Coli K-12 In-Frame, Single-Gene Knock-Out Mutants: The Keio Collection. *Mol syst Biol*, 2, 2006. 008.

Baldwin, R. (1986). Temperature dependence of the hydrophobic interaction in protein folding. *Proc Natl Acad Sci*, 83, 8069–8072.

Baldwin, R. L. (1989). How does protein folding get started? *Trends Biochem Sci,* 14, 291.

Ballard, C. et al. (2011). Alzheimer's disease. *Lancet*, 377(9770), 1019-1031.

Bartels, T. et al. (2010). The N-Terminus of the Intrinsically Disordered Protein α-Synuclein Triggers Membrane Binding and Helix Folding*. Biophys J*, 99(7), 2116–2124.

Baynes, B. M. and Trout, B. L. (2004). Rational design of solution additives for the prevention of protein aggregation. *Biophys J*, 87(3), 1631-1639.

Berman, H.M. et al. (2002). The Protein Data Bank*. Acta Crystallogr D Biol,* 58 (Pt 6 No 1), 899-907.

Bertoncini, C. W. et al. (2005). Release of long-range tertiary interactions potentiates aggregation of natively unstructured α-synuclein. *Proc Natl Acad Sci USA*, 102(5), 1430-1435.

Bharadwaj, P. R. et al. (2012). Latrepirdine (dimebon) enhances autophagy and reduces intracellular GFP-

Aβ42 levels in yeast. *J Alzheimers Dis*, 32, 949–67.

Bitan, G., Vollers, S. S. and Teplow, D. B. (2003). Elucidation of primary structure elements controlling early amyloid beta-protein oligomerization. *J Biol Chem*, 278, 34882–9.

Bodles, A. M. et al. (2001). Identification of the region of non-Aβ component (NAC) of Alzheimer's disease amyloid responsible for its aggregation and toxicity, *J Neurochem*, 78, 384–395.

Boeckmann, B. et al. (2003). The SWISS-PROT Protein Knowledgebase and Its Supplement TrEMBL In 2003. *Nucleis Acid Res,* 31, 365-370.

Braak, H. et al. (2003). Idiopathic Parkinson's disease: possible routes by which vulnerable neuronal types may be subject to neuroinvasion by an unknown pathogen. *J Neural Transm (Vienna)*, 110(5), 517-536.

Branden C., T. J. (1998). *Introduction to protein structure.* Garland Publishing, New York.

Breydo, L., Wu, J. and Uversky, V. (2012). α-synuclein misfolding and Parkinson's disease. *Biochim Biophys Acta*. 1822(2), 261–285.

Bryan, A. W. et al. (2009). BETASCAN: probable beta-amyloids identified by pairwise probabilistic analysis. *PLoS Comput Biol*, 5(3), e1000333.

Buchberger, A., Bukau, B. and Sommer, T. (2010). Protein quality control in the cytosol and the endoplasmic reticulum: brothers in arms. *Mol Cell*, 40(2), 238-252.

Buck, P. M., Kumar, S. and Singh, S. K. (2013). On the Role of Aggregation Prone Regions in Protein Evolution, Stability, and Enzymatic Catalysis: Insights from Diverse Analyses. *PLoS Comput Biol,* 9(10), e1003291.

Butterfield, S. M. and Lashuel, H. A. (2010). Amyloidogenic protein-membrane interactions: mechanistic insight from model systems. *Angew Chem Int Ed Engl*, 49(33), 5628-5654.

Calero, M. and Gasset, M. (2004). Fourier Transform Infrared and Circular Dichroism Spectroscopies for Amyloid Studies. *Methods Mol Biol*, 299, 129-151.

Carrio, M. et al. (2005). Amyloid-like properties of bacterial inclusion bodies. *J Mol Biol,* 347(5), 1025-1037.

Castillo, V. and Ventura, S. (2009). Amyloidogenic regions and interaction surfaces overlap in globular proteins related to conformational diseases. *PLoS Comput Biol*, 5: e1000476.

Castillo, V., Chiti, F. and Ventura, S. (2013) The N-terminal Helix Controls the Transition between the Soluble and Amyloid States of an FF Domain. *PLoS ONE*. 8(3), e58297.

Chiti, F. and Dobson, C. M. (2009). Amyloid formation by globular proteins under native conditions. *Nat Chem Biol*, 5(1), 15-22.

Chiti, F. and Dobson, C.M. (2006). Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem*, 75, 333-366.

Chiti, M. et al. (2003). Rationalization of the effects of mutations on peptide and protein aggregation rates. *Nature*, 424(6950), 805-808.

Ciechanover, A. (2005). Proteolysis: from the lysosome to ubiquitin and the proteasome. *Nat Rev Mol Cell Biol*, 6(1), 79-87.

Ciryam P et al. (2013). Widespread aggregation and neurodegenerative diseases are associated with supersaturated proteins. *Cell Rep*, 5(3), 781-790.

Cohen, E. et al. (2006). Opposing activities protect against age-onset proteotoxicity. *Science*, 313, 1604–1610.

Conchillo-Sole, O. et al. (2007). AGGRESCAN: A server for the prediction and evaluation of “hot spots” of aggregation in polypeptides. *BMC Bioinformatics*, 8, 65.

Cooper, A. A. et al. (2006). Alpha-synuclein blocks ER-Golgi traffic and Rab1 rescues neuron loss in Parkinson’s models. *Science*, 313, 324–328.

Cowan, S. W. et al. (1992). Crystal Structures Explain Functional Properties Of Two E. Coli Porins. *Nature*, 358, 727–733.

Craig, D. B. and Dombkowski, A. A. (2013). Disulfide by Design 2.0: a web-based tool for disulfide engineering in proteins. *BMC Bioinformatics*,14, 346.

Creighton, T. E. (1993). Proteins: structures and molecular properties. W. H. Freeman, New York, c1993.

Crowther, R. et al. (1998). Synthetic filaments assembled from C-terminally truncated α-synuclein*. FEBS Letters*, 436, 309-312.

de Groot, N. S. and Burgas, M. T. (2015). Is membrane homeostasis the missing link between inflammation and neurodegenerative diseases? *Cell Mol Life Sci*, 72, 4795–4805.

de Groot, N. S. and Ventura, S. (2010). Protein Aggregation Profile of the Bacterial Cytosol. *PLoS One*, 5(2): e9383.

de Groot, N. S. et al. (2006). Mutagenesis of the central hydrophobic cluster in Aβ42 Alzheimer's peptide: Side-chain properties correlate with aggregation propensities. *FEBS J*, 273, 658-668.

Deas, E. et al. (2016). Alpha-Synuclein Oligomers Interact with Metal Ions to Induce Oxidative Stress and Neuronal Death in Parkinson’s Disease. *Antioxid Redox Signal*, 24, 376–391.

Dedmon, M. M. (2005). Mapping long-range interactions in alpha-synuclein using spin-label NMR and ensemble molecular dynamics simulations, *J Am Chem Soc*, 127, 476–477.

Dennis, G. Jr. et al. (2003). DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol,* 4 (5), P3.

Dev et al. (2003). Part II: alpha-synuclein and its molecular pathophysiological role in neurodegenerative disease. *Neuropharmacology*, 45(1), 14-44.

Dill, K. A. and Chan, H. S. (1997). From Levinthal to pathways and funnels. *Nat Struct Biol,* 4(1), 10-19.

Dill, K. A. and MacCallum, J. L. (2012). The protein-folding problem, 50 years on. *Science*, 338(6110), 1042-1046.

Dill, K. A. et al. (2007). The protein folding problem: when will it be solved? *Curr Opin Struct Biol*, 17(3), 342-346.

Dobson, C. M. (2002). Getting out of shape. *Nature*, 418, 729–730.

Dobson, C. M. (2003). Protein Folding and Misfolding. *Nature*, 426, 884-890.

Doig, A. J. and Williams, D. H. (1992). Why water-soluble,compact, globular proteins have similar specific enthalpies of unfolding at 110°C. *Biochemistry*, 31, 9371–9375.

Dosztanyi, Z. et al. (2005). IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics*, 21, 3433-3434.

Dougan, D. A., Mogk, A. and Bukau, B. (2002) Protein Folding And Degradation In Bacteria: To Degrade Or Not To Degrade? That Is The Question. *Cell Mol Life Sci*, 59, 1607–1616.

Douglas, P. M. et al. (2008). Chaperone-dependent amyloid assembly protects cells from prion toxicity. *Proc Natl Acad Sci USA*, 105, 7206–7211.

Drummond, D. A. and Wilke, C. O. (2009). The evolutionary consequences of erroneous protein synthesis. *Nat Rev Genet*, 10(10), 715-724.

DuBay, K. F. et al. (2004). Prediction of the absolute aggregation rates of amyloidogenic polypeptide chains. *J Mol Biol*, 341(5),1317-1326F.

Dunker, A. K. et al. (2005). Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS J,* 272(20), 5129-5148.

El-Agnaf, O. M. A. et al. (1998). The N-terminal region of non-Aβ component of Alzheimer's disease amyloid is responsible for its tendency to assume β-sheet and aggregate to form fibrils. *Eur J Biochem*, 258, 157–163.

Eliezer, E. et al. (2001). Conformational properties of alpha-synuclein in its free and lipid-associated states. *J Mol Biol*, 307, 1061-1073.

Ellis, R. J. (1993). The general concept of molecular chaperones. *Philos Trans R Soc Lond B Biol Sci*, 339(1289), 257-261.

Ellis, R. J. and Pinheiro, T. J. (2002). Medicine: danger-misfolding proteins. *Nature*, 416(6880), 483-484.

Esteban-Martin, S. et al. (2013). Identification of Fibril-Like Tertiary Contacts in Soluble Monomeric α-Synuclein. *Biophys J*, 105, 1192 – 1198.

Evangelisti, E. et al. (2016). Binding affinity of amyloid oligomers to cellular membranes is a generic indicator of cellular dysfunction in protein misfolding diseases. *Sci Rep*, 6, 32721.

Facchinetti, F., Dawson, V. L. and Dawson, T. M. (1998). Free radicals as mediators of neuronal injury. *Cell Mol Neurobiol*, 18, 667–682.

Fan, Y. et al. (2006). Beta-synuclein modulates alpha-synuclein neurotoxicity by reducing alpha-synuclein protein expression, *Hum Mol Genet,* 15(20), 3002-3011.

Fernandez-Busquets, X. et al. (2008). Recent structural and computational insights into conformational diseases. *Curr Med Chem*, 15, 1336.

Fernandez-Escamilla, A. M. et al. (2004). Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat Biotechnol*, 22(10), 1302-1306.

Ferron, F. et al. (2006). A practical overview of protein disorder prediction methods. *Proteins*, 65(1), 1-14.

Fersht, A. R. (1995). Optimization of rates of protein folding: the nucleationcondensation mechanism and its implications. *Proc Natl Acad Sci USA*, 92, 10869–10873.

Finn, R. D. et al. (2014). The Pfam protein families database. *Nuclei Acids Res*, 42, D222-D230.

Firdaus, W. J. et al. (2006). Huntingtin inclusion bodies are iron-dependent centers of oxidative events. *FEBS J*, 273, 5428–41.

Fitzpatrick, A. W. et al. (2011). Inversion of the balance between hydrophobic and hydrogen bonding interactions in protein folding and aggregation. *PloS Comput Biol*, 7(10), e1002169.

Forloni, G. et al. (2016). Oligomeropathies and pathogenesis of Alzheimer and Parkinson's diseases. *Mov Disord*, 31, 771–781.

Frousios, K. K. et al. (2009). Amyloidogenic determinants are usually not buried*. BMC Struct Biol,* 9:44.

Frydman, J. (2001). Folding of newly translated proteins in vivo: the role of molecular chaperones. *Annu Rev Biochem*, 70, 603–647.

Fusco, G. et al. (2014). Direct observation of the three regions in α-synuclein that determine its membrane-bound behaviour. *Nat Commun*, 5, 3827.

Galvagnion C. et al. (2015). Lipid vesicles trigger *α*-synuclein aggregation by stimulating primary nucleation. *Nat Chem Biol*, 11, 229–234.

Garbuzynskiy, S. O., Lobanov, M. Y. and Galzitskaya, O. V. (2010). FoldAmyloid: a method of prediction of amyloidogenic regions from protein sequence. *Bioinformatics*, 26(3):326-332.

George, J. M. (2003). The synucleins [published online ahead of print December 20, 2001]. *Genome Biol.* 3(1), REVIEWS3002.

Gerdes, S. Y. et al. (2003). Experimental determination and system level analysis of essential genes in Escherichia Coli MG1655. *J Bacteriol,* 185, 5673-5684.

Gerdes, S. Y. et al. (2003). Experimental determination and system level analysis of essential genes in Escherichia Coli MG1655 (2003*). J Bacteriol,* 185, 5673-5684.

Giasson, B. I. et al. (2001). A hydrophobic stretch of 12 amino acid residues in the middle of α-synuclein is essential for filament assembly, *J Biol Chem*., 276, 2380–2386.

Goedert, M. (2001). *Alpha-synuclein and neurodegenerative diseases*. *Nat. Rev Neurosci.* 2, 492–501.

Goldberg, A. L. (2003). Protein degradation and protection against misfolded or damaged proteins. *Nature*, 426(6968), 895-899.

Goldberg, M. S. and Lansbury, P. T. (2000). Is there a cause-and-effect relationship between alpha-synuclein fibrillization and Parkinson's disease? *Nat Cell Biol*, 2(7), E115-E119.

Golding, I. and Cox, E. C. (2006). Physical nature of bacterial cytoplasm. *Phys Rev Lett*, 96(9), 098102.

Grantcharova, V. P. and Baker, D. (2001). Circularization changes the folding transition state of the src SH3 domain. *J Mol Biol*, 306(3), 555–563.

Graña-Montes, R. et al. (2012). Contribution of disulfide bonds to stability, folding, and amyloid fibril formation: the PI3-SH3 domain case. *Antioxid Redox Signal*, 16(1), 1-15.

Gregersen, N. (2006). Protein misfolding disorders: pathogenesis and intervention. *J Inherit Metab Dis*, 29(2-3), 456-70.

Grey, M. et al. (2011). Membrane interaction of α-synuclein in different aggregation states. *J Parkinsons Dis*, 1(4), 359-371.

Guo, J. L. et al. (2013). Distinct α-synuclein strains differentially promote tau inclusions in neurons. *Cell*, 154(1), 103-117.

Hajipour, M. J. et al. (2017). Advances in Alzheimer’s Diagnosis and Therapy: The Implications of Nanotechnology. *Trends Biotechnol*, pii: S0167-7799(17), 30130-0.

Hamada, D. et al. (2009). Competition between folding, native-state dimerisation and amyloid aggregation in beta-lactoglobulin. *J Mol Biol*, 386, 878-890.

Han, H., Weinreb, P.H. and Lansbury, P.T., Jr. (1995). The core Alzheimer's peptide NAC forms amyloid fibrils which seed and are seeded by β-amyloid: is NAC a common trigger or target in neurodegenerative disease? *Chem Biol*, 2, 163–169.

Hartl, F. U. and Hayer-Hartl, M. (2009). Converging concepts of protein folding in vitro and in vivo. *Nat Struct Mol Biol*, 16(6), 574-581.

Hartl, F. U., Bracher, A. and Hayer-Hartl, M. (2011). Molecular chaperones in protein folding and proteostasis. *Nature,* 475, 324–332.

Hashimoto, M. et al. (2001). Beta-Synuclein inhibits alpha-synuclein aggregation: a possible role as an anti-parkinsonian factor. *Neuron*, 32(2), 213-223.

Herczenik, E. and Gebbink, M. F. (2008). Molecular and cellular aspects of protein misfolding and disease. *FASEB J*, 22(7), 2115-2133.

Horwich, A. L. et al. (2007). Two families of chaperonin: physiology and mechanism. *Annu Rev Cell Dev Biol,* 23, 115-145.

Hoyer, W. et al. (2002). Dependence of α-Synuclein Aggregate Morphology on Solution Conditions. *J Mol Biol* , 322, 383-393.

Huerta, A. M. et al. (1998). Regulondb: A Database On Transcriptional Regulation In Escherichia Coli. *Nuclei Acids Res,* 26, 55-59.

Invernizzi, G. et al. (2012). Protein aggregation: mechanisms and functional consequences. *Int J Biochem Cell Bio*, 44 (9), 1541-1554.

Ishihama, Y. et al. (2008). Protein Abundance Profiling Of The Escherichia Coli Cytosol. *BMC Genomics*, 9, 102.

Itzhaki, L. and Wolynes, P. (2008). The quest to understand protein folding. *Curr Opin Struct Biol*, 18(1), 1-3.

Itzhaki, L. S. et al. (1995). Search for nucleation sites in smaller fragments of chymotrypsin inhibitor 2, *J Mol Biol*, 254, 289-304.

Ivankov, D. N. et al. (2003). Contact order revisited: influence of protein size on the folding rate. *Protein Sci*, 12(9), 2057-2062.

Iwatsubo, T. et al. (1994). Visualization of A beta 42(43) and A beta 40 in senile plaques with end-specific A beta monoclonals: evidence that an initially deposited species is A beta 42(43), *Neuron*, 13(1), 45-53.

Jackson, S. E. (1998). How do small single-domain proteins fold? *Folding and design*, 3, R81-R91.

Jahn, T. R. and Radford, S. E. (2008). Folding versus aggregation: polypeptide conformations on competing pathways. *Arch Biochem Biophys,* 469, 100-117.

Jakob-Roetne, R. and Jacobsen, H. (2009). Alzheimer's disease: from pathology to therapeutic approaches. *ADD Angew Chem Int Ed Engl,* 48(17), 3030-3059.

Jarrett, J. T., and Lansbury, P. T., J. (1993). Seeding „one-dimensional crystallization“ of amyloid: a pathogenic mechanism in alzheimer's disease and scarpie? *Cell*, 73(6), 1055-1058.

Jeppesen, M. D., Westh, P. and Otzen, D.E. (2010). The role of protonation in protein fibrillation. *FEBS Lett*, 584(4), 780-784.

Jethva, P. N., Kardani, J. R. and Roy, I. (2011). Modulation of α-synuclein aggregation by dopamine in the presence of MPTP and its metabolite. *FEBS J*, 278, 1688-1698.

Jiang, P., Xu, W. and Mu, Y. (2009). Amyloidogenesis Abolished by Proline Substitutions but Enhanced by Lipid Binding. *PLoS Comput Biol*, 5(4), e1000357.

Jin, F. and Liu, Z. (2013). Inherent relationships among different biophysical prediction methods for intrinsically disordered proteins. *Biophys J*, 104(2), 488-495.

Kaganovich, D., Kopito, R. and Frydman, J. (2008). Misfolded proteins partition between two distinct quality control compartments, *Nature,* 454(7208), 1088-1095.

Kampinga, H. H. and Craig, E. A. (2010). The HSP70 chaperone machinery: J proteins as drivers of functional specificity. *Nat Rev Mol Cell Biol*, 11(8), 579-592.

Kawaguchi, Y. *et al.* (2003). The deacetylase HDAC6 regulates aggresome formation and cell viability in response to misfolded protein stress. *Cell*, 115, 727–738.

Kern, D. and Zuiderweg, E. R. (2003). The role of dynamics in allosteric regulation. *Curr Opin Struct Biol.* 13(6), 748-757.

Khurana, V. and Lindquist, S. (2010). Modelling neurodegeneration in Saccharomyces cerevisiae: why cook with baker's yeast? *Nat Rev Neurosci*, 11, 436–49.

Kim, Y. E. et al. (2013). Molecular chaperone functions in protein folding and proteostasis. *Annu Rev Biochem,* 82, 323-355.

King, O. D., Gitler, A. D. and Shorter, J. (2012). The tip of the iceberg: RNA-binding proteins with prion-like domains in neurodegenerative disease. *Brain Res*, 1462, 61–80.

Kitada, T. et al*.* (1998). Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. *Nature,* 392, 605–608.

Knowles, T. J. et al. (2009). Membrane Protein Architects: The Role Of The BAM Complex In Outer Membrane Protein Assembly. *Nat Rev Microbiol*, 7, 206–214.

Kopito, R. R. (2000). Aggresomes, inclusion bodies and protein aggregation. *Trends Cell Biol*, 10(12), 524-530.

Krebs, M. R. et al. (2004). Observation of sequence specificity in the seeding of protein amyloid fibrils. *Protein Sci,* 13(7), 1933-1938.

Kyte, J. and Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *J Mol Biol*, 157(1), 105-132.

Lashuel H.A. et al. (2013) The many faces of alpha-synuclein: from structure and toxicity to therapeutic target. *Nat Rev Neurosci*, 14, 38–48.

Laskowska, E. et al. (2004). Aggregation of heat-shock-denatured, endogenous proteins and distribution of the IbpA/B and Fda marker-proteins in *Escherichia coli* WT and grpE280 cells. *Microbiology*, 150, 247–259.

Lawrence, M. S., Phillips, K.J. and Liu, D. R. (2007). Supercharging proteins can impart unusual resilience. *J Am Chem Soc*, 129(33), 10110-1012.

Lee, H. et al. (2000). Local structural elements in the mostly unstructured transcriptional activation domain of human p53. *J Biol Chem*, 275, 29426-29432.

Lee, S.H. et al. (2012). Understanding pre-structured motifs (PreSMos) in intrinsically unfolded proteins. *Curr Protein Pept Sci*, 13, 34-54.

Leuenberger, P. et al. (2017). Cell-wide analysis of protein thermal unfolding reveals determinants of thermostability. *Science*, 355(6327). pii: eaai7825.

Levy, E. D, Michnick, S. W, and Landry, C. R. (2012). Protein abundance is key to distinguish promiscuous from functional phosphorylation based on evolutionary information. *Philos Trans R Soc Lond B Biol Sci*, 367(1602), 2594–2606.

Levy, E. D. and Teichmann, S. A. (2012). Cellular crowding imposes global constraints on the chemistry and evolution of proteomes. *Proc Natl Acad Sci U S A*, 109(50), 20461-20466.

Li, G. W. et al. (2014). Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell,* 157, 624–635.

Linding, R. et al. (2004). A comparative study of the relationship between protein structure and beta-aggregation in globular and intrinsically disordered proteins*. J Mol Biol,* 342(1), 345-353.

Lührs, T. et al. (2005). 3D structure of Alzheimer's amyloid-beta(1-42) fibrils. *Proc Natl Acad Sci U S A*, 102, 17342–7.

Ma, M. R. et al. (2016). Phosphorylation induces distinct alpha-synuclein strain formation. *Sci Rep*, 6, 37130.

MacDonald, R. C. et al. (1991). Small volume extrusion apparatus for preparation of large unilamellar vesicles. *Biochim Biophys Acta*, 1061, 297–303.

Mankar, S. et al. (2011). Nanomaterials: amyloids reflect their brighter side. *Nano Rev*, 2, 6032-6043.

Marcoux, J. and Robinson, C. V. (2013). Twenty years of gas phase structural biology. *Structure,* 21, 1541–1550.

Maroteaux, L. and Scheller, R. H. (1991). The rat brain synucleins: Family of proteins transiently associated with neuronal membrane, *Brain Res Mol Brain Res*, 11, 335-343.

Maurer-Stroh, S. et al. (2010). Exploring the sequence determinants of amyloid structure using position-specific scoring matrices. *Nat Methods*, 7(3), 237-242.

McCann, H. et al. (2014). α-Synucleinopathy phenotypes. *Parkinsonism Relat Disord*, 20 Suppl 1, S62–S67.

Melki, R. (2015). Role of Different Alpha-Synuclein Strains in Synucleinopathies, Similarities with other Neurodegenerative Diseases. *J Parkinsons Dis*, 5(2), 217-227.

Merlini, G. and Bellotti, V. (2003). Molecular mechanisms of amyloidosis. *N Engl J Med*, 349(6), 583-596.

Minor, D. L. and Kim, P. S. (1996). Context-dependent secondary structure formation of a designed protein sequence. *Nature*, 380(6576), 730-734.

Mitrea, D. M. and Kriwacki, R. W. (2016). Phase separation in biology; functional organization of a higher order. *Cell Commun Signal*, 14, 1.

Modler, A. J. et al. (2003). Assembly of amyloid protofibrils via critical oligomers-A novel pathway of amyloid formation. *J Mol*, *325*, 135-148.

Monsellier, E. and Chiti, F. (2007). Prevention of amyloid-like aggregation as a driving force of protein evolution. *EMBO Rep*, 8(8), 737-742.

Monsellier, E. et al. (2008). Aggregation Propensity of the Human Proteome. *PLoS Comput Biol*, 4(10), e1000199.

Morel, B. et al. (2010). Environmental conditionsaffect the kinetics of nucleation of amyloid fibrils and determine their morphology. *Biophys J*, 99(11), 3801-3810.

Morell, M. et al. (2011). Linking amyloid protein aggregation and yeast survival. *Mol Biosyst*, 7, 1121–8.

Morimoto, A. et al. (2004). Analysis of the secondary structure of beta-amyloid (Abeta42) fibrils by systematic proline replacement*. J Biol Chem*, 279, 52781–8.

Mossuto, M. F. et al. (2011). Disulfide bonds reduce the toxicity of the amyloid fibrils formed by an extracelular protein. *Angew Chem Int Ed Engl*, 50(31), 7048-7051.

Natalello, A. and Doglia, S. M. (2015). Insoluble protein assemblies characterized by fourier transform infrared spectroscopy. *Methods Mol Biol*, 1258, 347-369.

Navarro, S., Villar-Piqué, A. and Ventura, S. (2014) Selection against toxic aggregation-prone protein sequences in bacteria. *Biochim Biophys Acta -Mol Cell Res*, 1843, 866–874.

Nilsson, M. R. (2004). Techniques to study amyloid fibril formation in vitro. *Methods*, 34, 151-160.

Nishioka, K. et al. (2010). Association of α-, β-, and γ-Synuclein With Diffuse Lewy Body Disease. *Arch Neurol*, 67(8), 970–975.

Niwaa, T. et al. (2009). Bimodal protein solubility distribution revealed by an aggregation analysis of the entire ensemble of Escherichia coli proteins. *Proc Natl Acad Sci U S A*, 106(11), 4201-4206.

Nixon, R. A. (2007). Autophagy, amyloidogenesis and Alzheimer disease. *J Cell Sci*, 120(Pt 23), 4081-4091.

Ohgushi, M. and Wada, A. (1983). Molten-globule state: a compact form of globular proteins with mobile side-chains. *FEBS Lett*, 164 (1), 21-24.

Orchard, S. et al. (2013) The MIntAct project – IntAct as a common curation platform for 11 molecular interaction databases. *Nuclei Acids Res*, 42, D358–D363.

Pace, C. N. (1988). Conformational stability and activity of ribonuclease T1 with zero, one, and two intact disulfide bonds. *J Biol Chem*, 263 (24), 11820–11825.

Payton, J. E. et al. (2001). Protein-protein interactions of alpha-synuclein in brain homogenates and transfected cells. *Brain Res Mol Brain Res*, 95(1-2), 138-145.

Pechmann, S. et al. (2009). Physicochemical principles that regulate the competition between functional and dysfunctional association of proteins. *Proc Natl Acad Sci USA*, 106(25), 10159-10164.

Petty, S. A. and Decatur, S. M. (2005). Intersheet rearrangement of polypeptides during nucleation of ß-sheet aggregates. *Proc Natl Acad Sci USA*, *102*, 14272-14277.

Ptitsyn, O. B. (1995). Structures of folding intermediates. *Curr Opin Struct Biol*, 5(1), 74-78.

Quintas, A. (2013). What drives an amyloid protein precursor from an amyloidogenic to a native-like aggregation pathway? *OA Biochemistry*, 1(1), 6.

Rabe, M. et al. (2013). On-surface aggregation of α-synuclein at nanomolar concentrations results in two distinct growth mechanisms. *ACS Chem Neurosci,* 4(3), 408-417.

Radford, S. E. (2000). Protein folding: progress made and promises ahead. *Trends Biochem Sci*, 25(12), 611-618.

Radivojac, P. et al. (2007). Intrinsic disorder and functional proteomics. *Biophys J,* 92(5), 1439-1456.

Rauscher, S. et al. (2006). Proline and glycine control protein self-organization into elastomeric or amyloid fibrils. *Structure*, 14(11), 1667-1676.

Requiao, R. D. et al. (2017). Protein charge distribution in proteomes and its impact on translation. *PLoS Comput Biol*, 13(5), e1005549.

Reumers, J. et al. (2009). Protein sequences encode safeguards against aggregation. *Hum Mutat*, 30(3), 431-437.

Roberts, H.L. and Brown, D.R. (2015) Seeking a mechanism for the toxicity of oligomeric alpha-synuclein. *Biomolecules*, 5, 282–305.

Rodriguez, F. et al. (2008). Molecular basis for regulation of the heat shock transcription factor sigma32 by the DnaK and DnaJ chaperones. *Mol Cell*, 32(3), 347-358.

Rokney, A. *et al.* (2009). *E. coli* transports aggregated protein*s* to the poles by a specific and energy-dependent process*. J Mol Biol,* 392, 589–601.

Ross, C. A. and Pickart, C. M. (2004). The ubiquitin-proteasome pathway in Parkinson's disease and other neurodegenerative diseases. *Trends Cell Biol,* 14, 703–711.

Ross, C. A. and Poirier, M. A. (2005). Opinion: What is the role of protein aggregation in neurodegeneration? *Nat Rev Mol Cell Biol*, 6, 891–898.

Rousseau, F., Serrano, L. and Schymkowitz, J. W. (2006). How evolutionary pressure against protein aggregation shaped chaperone specificity. *J Mol Biol,* 355(5), 1037-1047.

Rudiger, S. et al. (2010). Modulation of substrate specificity of the DnaK chaperone by alteration of a hydrophobic arch. *J Mol Biol*, 304(3), 245-251.

Sabate, R. et al. (2010). The role of protein sequence and amino acid composition in amyloid formation: scrambling and backward reading of IAPP amyloid fibrils. *J Mol Biol*, 404(2), 337-52.

Sabate, R. et al. (2015). What Makes a Protein Sequence a Prion? *PLoS Comput Biol* 11(1), e1004013.

Sabate, R. and Ventura, S. (2013). Cross-beta-sheet supersecondary structure in amyloid folds. Tecniques for detection and characterization*. Methods Mol Biol*, 932, 237-257.

Sanchez de Groot, N. et al. (2015). Proteome response at the edge of protein aggregation. *Open Biol,* 5, 140221.

Santoni, V., Molloy, M. and Rabilloud, T. (2000). Membrane Proteins And Proteomics: Un Amour Impossible? *Electrophoresis*, 21, 1054–1070.

Schirmer, T. (1998). General And Specific Porins From Bacterial Outer Membranes. *J Struct Biol*, 121, 101–109.

Schmidt, C. and Robinson, C. V. (2014). Dynamic protein ligand interactions insights from MS. *FEBS J,* 281, 1950–1964.

Schubert, U. et al. (2000). Rapid degradation of a large fraction of newly synthesized proteins by proteasomes, *Nature*, 404, 770–774.

Schymkowitz, J. et al. (2005). The FoldX web server: an online force field. *Nucleic Acids Res*, 33 (Web Server issue): W382–W388.

Selinger, D. W. et al. (2000). RNA Expression Analysis Using A 30 Base Pair Resolution Escherichia Coli Genome Array. *Nat Biotechnol*, 18, 1262–1268.

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

Sharon, R.et al. (2001). Alpha-Synuclein occurs in lipid-rich high molecular weight complexes, binds fatty acids, and shows homology to the fatty acid-binding proteins. *Proc Natl Acad Sci U S A*, 98, 9110–9115.

Shaykhalishahi, H. et al. (2015). Contact between the β1 and β2 Segments of α-Synuclein that Inhibits Amyloid Formation. *Angew Chem Int Ed Engl*, 54(30), 8837-3840.

Sievers, F and Higgins, D. G. (2014). Clustal Omega, accurate alignment of very large numbers of sequences. *Methods Mol Biol,* 1079, 105-116.

Silva, B. A., Breydo, L. and Uversky, V. (2013). Targeting the chameleon: a focused look at α-synuclein and its roles in neurodegeneration. *Molecular neurobiology* 47, 446-459.

Sorolla, M. A. et al. (2011). Sir2 is induced by oxidative stress in a yeast model of Huntington disease and its activation reduces protein aggregation*. Arch Biochem Biophys*, 510, 27–34.

Spencer, H. L. et al. (1985). Influence of an extrinsic crosslink on the folding pathway of ribonuclease A. Kinetics of folding-unfolding. *Biochemistry*, 24 (11), 2680–2686.

Spillantini, M. G. et al. (1997). R-Synuclein in Lewy bodies. *Nature*, 388, 839-840.

Spillantini, M. G. et al. (1998). Alpha-Synuclein in filamentous inclusions of Lewy bodies from Parkinson’s disease and dementia with lewy bodies. *Proc Natl Acad Sci USA,* 95, 6469–6473.

Stromer, T. and Serpell, L. C. (2005). Structure and morphology of the Alzheimer's amyloid fibril. *Microsc Res Tech*, 67(3-4), 210-217.

Sydow, A. et al. (2011). Tau-induced defects in synaptic plasticity, learning, and memory are reversible in transgenic mice after switching off the toxic Tau mutant. *J Neurosci*, 31, 2511–2525.

Tabner, B. J. et al. (2005). Hydrogen peroxide is generated during the very early stages of aggregation of the amyloid peptides implicated in Alzheimer disease and familial British dementia. *J Biol Chem*, 280 (2005) 35789–92.

Tabner, B. J. et al. (2005). Protein aggregation, metals and oxidative stress in neurodegenerative diseases. *Biochem Soc Trans*, 33, 1082–1806.

Taipale, M., Jarosz, D. F. and Lindquist, S. (2010). HSP90 at the hub of protein homeostasis: emerging mechanistic insights. *Nat Rev Mol Cell Biol,* 11(7), 515-528.

Tanaka, M. et al. (2004). Aggresomes formed by α-synuclein and synphilin-1 are cytoprotective. *J Biol Chem,* 279, 4625–463.

Tartaglia, G. G. et al. (2005). Organism complexity anti-correlates with proteomic beta-aggregation propensity. *Protein Sci*, 14(10), 2735-2740.

Tartaglia, G. G. et al. (2008). Prediction of aggregation-prone regions in structured proteins*. J Mol Biol,* 380(2), 425-436.

Tartaglia, G. G. et al. (2008). The Zyggregator Method for Predicting Protein Aggregation Propensities. *Chem Soc Rev*, 37 (7), 1395-1401.

Tartaglia, G.G. et al. (2007). Life On The Edge: A Link Between Gene Expression Levels And Aggregation Rates Of Human Proteins. *Trends Biochem Sci*, 32, 204–206.

Tenreiro, S. and Outeiro, T. F. (2010). Simple is good: yeast models of neurodegeneration. *FEMS Yeast Res*, 10, 970–979.

Tidor, B. and Karplus, M. (1993). The contribution of cross-links to protein stability: a normal mode analysis of the configurational entropy of the native state. *Proteins*, 15 (1), 71–79.

Tolosa, E., Wenning, G. and Poewe, W. (2006). The diagnosis of Parkinson's disease. *Lancet Neurol*, 5(1), 75-86.

Tompa, P. (2002). Intrinsically unstructured proteins. *Trends Biochem Sci*, 27(10), 527-533.

Tompa, P. (2012). Intrinsically disordered proteins: a 10-year recap. *Trends Biochem Sci*, 37(12), 509-516.

Treusch, S. et al. (2011). Functional links between Aβ toxicity, endocytic trafficking, and Alzheimer’s disease risk factors in yeast. *Science,* 334, 1241–5.

Trivedi, M. V., Laurence, J. S., and Siahaan, T. J. (2009). The role of thiols and disulfides on protein stability. *Curr Protein Pept Sci.*10(6), 614-625.

Trovato, A., Seno, F. and Tosatto, S. C. (2007). The PASTA server for protein aggregation prediction. *Protein Eng Des Sel*, 20(10), 521-523.

Tsolis A. C. et al. (2013). A consensus method for the prediction of 'aggregation-prone' peptides in globular proteins. *PLoS One*, 8(1), e54175.

Tuttle, M. D. et al. (2016). Solid-state NMR structure of a pathogenic fibril of full-length human α-synuclein. *Nat Struct Mol Biol*, 23(5), 409-415.

Tyedmers, J., Mogk, A. and Bukau, B. (2010). Cellular strategies for controlling protein aggregation. *Nat Rev Mol Cell Biol*, 11(11), 777-788.

Uversky, V. and Eliezer, D. (2009). Biophysics of Parkinson’s Disease: Structure and Aggregation of α-Synuclein. *Curr* Protein *Pept Sci*, 10, 483-499.

Uversky, V. N. (2002). Natively unfolded proteins: a point where biology waits for physics. *Protein Sci,* 11(4), 739-756.

Uversky, V. N. (2008). Alpha-synuclein misfolding and neurodegenerative diseases.*Curr Protein Pept Sci*, 9(5), 507-540.

Uversky, V. N., Li, J. and Fink, A. L. (2001). Evidence for a partially folded intermediate in alpha-synuclein fibril formation. *J Biol Chem*, 276, 10737-10744.

Uversky, V. N., Gillespie, J. R. and Fink, A. L. (2000). Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins*, 41(3), 415-427.

van den Berg, B., Ellis, R. J. and Dobson, C. M. (1999). Effects of macromolecular crowding on protein folding and aggregation. *EMBO J*, 18(24), 6927-6933.

Van Durme, J. et al. (2016). Solubis: a webserver to reduce protein aggregation through mutation. *Protein Eng Des Sel*, 29(8), 285-289.

Vavouri, T. et al. (2009).Intrinsic protein disorder and interaction promiscuity are widely associated with dosage sensitivity. *Cell*, 138(1), 198-208.

Vendruscolo, M. and Dobson, C. M. (2013). Structural biology: Protein self-assembly intermediates. *Nat Chem Biol*, 9(4), 216-217.

Vendruscolo, M. et al. (2003). Structures and relative free energies of partially folded states of proteins. *Proc Natl Acad Sci USA*, 100(25), 14817-14821.

Villar-Pique, A. and Ventura, S. (2012). Protein Aggregation Acts as Strong Constraint During Evolution. Evolutionary Biology: Mechanisms and Trends. Berlin: Springer, pp. 103–120.

Villar-Piqué, A. and Ventura, S. (2013). Protein aggregation propensity is a crucial determinant of intracellular inclusion formation and quality control degradation. B*iochim Biophys Acta*, 1833, 2714–24.

Vlad, C. et al. (2011). Autoproteolytic fragments are intermediates in the oligomerization/aggregation of the Parkinson’s disease protein alpha-synuclein as revealed by ion mobility mass spectrometry. *Chembiochem*, 12, 2740– 2744.

Volle,s M. J. and Lansbury, P. T. Jr. (2007). Relationships between the sequence of alpha-synuclein and its membrane affinity, fibrillization propensity, and yeast toxicity. *J Mol Biol*, 366(5), 1510-22.

Wales, P. et al. (2013). Limelight on alpha-synuclein: pathological and mechanistic implications in neurodegeneration. *J Parkinsons Dis*, 3(4), 415-459.

Walter, S. and Buchner, J. (2002). Molecular chaperones--cellular machines for protein folding. *Angew Chem Int Ed Engl,* 41(7), 1098-1113.

Wang, X. et al. (2014). Oxidative stress and mitochondrial dysfunction in Alzheimer’s disease. *Biochim Biophys Acta*, 1842, 1240–1247.

Wang, M. et al. (2012). PaxDb, a database of protein abundance averages across all three domains of life. *Mol Cell Proteomics*, 11(8), 492–500.

Watters, A. L. et al. (2007). The highly cooperative folding of small naturally occurring proteins is likely the result of natural selection. *Cell*, 128(3), 613-624.

Waxman, E. A., Mazzulli, J. R. and Giasson. B. I. (2009). Characterization of hydrophobic residue requirements for alpha-synuclein fibrillization. *Biochemistry*, 48(40), 9427-9436.

Weiss, W. F., Young, T. M. and Roberts, C. J. (2009). Principles, approaches, and challenges for predicting protein aggregation rates and shelf life*. J Pharm Sci*, 98(4), 1246-1277.

Wen-Hui, X. and Guang-Hong, W. (2016). Amyloid-*β* peptide aggregation and the influence of carbon nanoparticles. *Chinese Physics B*, 25(1), 018704.

Willingham, S. et al. (2003). Yeast genes that enhance the toxicity of a mutant huntingtin fragment or alpha-synuclein*. Science*, 302, 1769–1772.

Winner, B. et al. (2011). In vivo demonstration that alpha-synuclein oligomers are toxic. *Proc Natl Acad Sci USA*, 108, 4194–4199.

Wong, E. et al. (2012). Molecular determinants of selective clearance of protein inclusions by autophagy. *Nat Commun*, 3, 1240.

Wright, C. F. et al. (2005). The importance of sequence diversity in the aggregation and evolution of proteins. *Nature,* 438 (7069), 878-881.

Wright, P. E. and Dyson, H. J. (1999). Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol,* 293, 321–331.

Xu, S. and Chan, P. (2015). Interaction between Neuromelanin and Alpha-Synuclein in Parkinson's Disease. *Biomolecules*, *5*(2), 1122-1142.

Yang, Z. R. et al. (2005). RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics*, 21, 3369-3376.

Yoon, S. and Welsh, W. J. (2004). Detecting hidden sequence propensity for amyloid fibril formation. *Protein Sci,* 13(8), 2149-2160.

Yoshimura, Y. et al. (2017). MOAG-4 promotes the aggregation of alpha-synuclein by competing with self-protective electrostatic interactions. *J Biol Chem*, 292(20), 8269-8278.

Zambrano, R. et al. (2015a). PrionW: a server to identify proteins containing glutamine/asparagine rich prion-like domains and their amyloid cores. *Nucleic Acids Res*, 43(W1), W331-W337.

Zambrano, R. et al. (2015b). AGGRESCAN3D (A3D): server for prediction of aggregation properties of protein structures. *Nucleic Acids Res*, 43(W1), W306-W313.

Zhou, J. and Rudd, K. E. (2013). *EcoGene 3.0. Nuclei Acids Res*, 41 (D1), D613-D624.

Zhou, T., Weems, M. and Wilke, C. O. (2009). Translationally optimal codons associate with structurally sensitive sites in proteins. *Mol Biol Evol*, 26(7), 1571-1580.

Zibaee, S. et al. (2007). A simple algorithm locates beta-strands in the amyloid fibril core of alpha-synuclein, Abeta, and tau using the amino acid sequence alone. *Protein Sci*, 16(5), 906-918.

9. Appendices

Publications:

1. *Environmental and genetic factors support the dissociation between α-synuclein aggregation and toxicity*
2. *The effects of the novel A53E alpha-synuclein mutation on its oligomerization and aggregation*
3. *Data on correlation between Aβ42 structural aggregation propensity and toxicity in bacteria*

Environmental and genetic factors support the dissociation between α-synuclein aggregation and toxicity

Anna Villar-Piqué[a,1], Tomás Lopes da Fonseca[a,1], Ricardo Sant'Anna[b,c,d], Éva Mónika Szegö[a], Luis Fonseca-Ornelas[e], Raquel Pinho[a], Anita Carija[c], Ellen Gerhardt[a], Caterina Masaracchia[a], Enrique Abad Gonzalez[f,g], Giulia Rossetti[f,h], Paolo Carloni[f], Claudio O. Fernández[i,j], Debora Foguel[b], Ira Milosevic[k], Markus Zweckstetter[e,l,m,n], Salvador Ventura[c,d], and Tiago Fleming Outeiro[a,o,2]

[a]Department of Neurodegeneration and Restorative Research, University Medical Centre Göttingen, 37073 Goettingen, Germany; [b]Instituto de Bioquímica Médica Leopoldo de Meis, Universidade Federal do Rio de Janeiro, 21941-901 Rio de Janeiro, Brazil; [c]Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain; [d]Departament de Bioquímica i Biologia Molecular, Facultat de Biociències, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain; [e]Max-Planck-Institut für Biophysikalische Chemie, 37077 Goettingen, Germany; [f]Computational Biomedicine, Institute for Advanced Simulation IAS-5 and Institute of Neuroscience and Medicine INM-9, Forschungszentrum Jülich, 52425 Jülich, Germany; [g]German Research School for Simulation Sciences, Forschungszentrum Jülich, 52425 Jülich, Germany; [h]Department of Oncology, Hematology, and Stem Cell Transplantation, Medical School, Rheinisch-Westfälische Technische Hochschule Aachen University, 52074 Aachen, Germany; [i]Max Planck Laboratory for Structural Biology, Chemistry, and Molecular Biophysics of Rosario, Universidad Nacional de Rosario, Ocampo y Esmeralda, S2002LRK Rosario, Argentina; [j]Instituto de Investigaciones para el Descubrimiento de Fármacos de Rosario UNR-Consejo Nacional de Investigaciones Científicas y Técnicas de Argentina, Universidad Nacional de Rosario, Ocampo y Esmeralda, S2002LRK Rosario, Argentina; [k]European Neuroscience Institute, 37077 Goettingen, Germany; [l]Deutsches Zentrum für Neurodegenerative Erkrankungen, 37075 Goettingen, Germany; [m]Department of Neurology, University Medicine Göttingen, 37073 Goettingen, Germany; [n]Center for Nanoscale Microscopy and Molecular Physiology of the Brain, University Medicine Göttingen, 37075 Goettingen, Germany; and [o]Max Planck Institute for Experimental Medicine, Goettingen, Germany

Edited by Gregory A. Petsko, Weill Cornell Medical College, New York, NY, and approved September 2, 2016 (received for review April 28, 2016)

Synucleinopathies are a group of progressive disorders characterized by the abnormal aggregation and accumulation of α-synuclein (aSyn), an abundant neuronal protein that can adopt different conformations and biological properties. Recently, aSyn pathology was shown to spread between neurons in a prion-like manner. Proteins like aSyn that exhibit self-propagating capacity appear to be able to adopt different stable conformational states, known as protein strains, which can be modulated both by environmental and by protein-intrinsic factors. Here, we analyzed these factors and found that the unique combination of the neurodegeneration-related metal copper and the pathological H50Q aSyn mutation induces a significant alteration in the aggregation properties of aSyn. We compared the aggregation of WT and H50Q aSyn with and without copper, and assessed the effects of the resultant protein species when applied to primary neuronal cultures. The presence of copper induces the formation of structurally different and less-damaging aSyn aggregates. Interestingly, these aggregates exhibit a stronger capacity to induce aSyn inclusion formation in recipient cells, which demonstrates that the structural features of aSyn species determine their effect in neuronal cells and supports a lack of correlation between toxicity and inclusion formation. In total, our study provides strong support in favor of the hypothesis that protein aggregation is not a primary cause of cytotoxicity.

α-synuclein | copper | H50Q mutation | inclusions | protein aggregation

Synucleinopathies are a group of neurodegenerative disorders that include Parkinson's disease (PD), dementia with Lewy bodies, and multiple system atrophy. Although the clinical manifestations in these disorders are heterogeneous, ranging from movement impairment to dementia, they all share the abnormal accumulation of α-synuclein (aSyn) in proteinaceous inclusions (1). The vast majority of synucleinopathy cases are sporadic. Nevertheless, mutations in the gene encoding for human aSyn have been reported in both familial and sporadic forms of synucleinopathies. In addition to gene multiplications, six point-mutations have been described to cause familial forms of PD: A30P, E46K, H50Q, G51D, A53T, and A53E (2, 3).

aSyn is a soluble 140-residue intrinsically disordered protein that is abundant in neuronal cells, in particular in presynaptic terminals (4). The primary sequence of aSyn can be divided into three regions: (*i*) the N-terminal domain, with strong propensity to adopt α-helical structure, responsible for membrane binding; (*ii*) the central region, known as nonamyloid component, displays high aggregation propensity and is crucial for the amyloidogenicity of the entire protein; and (*iii*) the C-terminal region, highly acidic and disordered, described as a chaperone-like domain (5). Although the biological function of aSyn is still not well understood, increasing evidence suggests it plays a major role in the synapse, where it regulates synaptic vesicle release upon functional multimerization (6). Remarkably, all pathological point-mutations are located in the N-terminal region, required for membrane interactions (3).

Misfolding and aggregation of aSyn is a central event in synucleinopathies. In the case of PD and dementia with Lewy bodies,

Significance

Many neurodegenerative diseases are characterized by the abnormal accumulation of aggregated proteins in the brain. In Parkinson's disease and related disorders, this process involves the accumulation of α-synuclein (aSyn). Thus, understanding the relationship between aSyn aggregation and pathological conditions is essential for the development of novel and efficient therapies against these disorders. Here, we studied the effects that different aSyn species have on neurons using a combination of neurodegeneration-associated factors: the H50Q aSyn mutant and the presence of copper. Importantly, we demonstrate that exogenous aSyn promotes toxicity and inclusion formation, and that these effects are inversely correlated. Our data shed light onto the pathological mechanisms associated with aSyn aggregation, forming the foundation for future therapeutic strategies.

Author contributions: A.V.-P., T.L.d.F., S.V., and T.F.O. designed research; A.V.-P., T.L.d.F., R.S., É.M.S., L.F.-O., R.P., A.C., E.G., E.A.G., G.R., and I.M. performed research; C.M., P.C., C.O.F., D.F., I.M., M.Z., S.V., and T.F.O. contributed new reagents/analytic tools; A.V.-P., T.L.d.F., R.S., É.M.S., L.F.-O., G.R., P.C., C.O.F., D.F., I.M., M.Z., S.V., and T.F.O. analyzed data; and A.V.-P., S.V., and T.F.O. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]A.V.-P. and T.L.d.F. contributed equally to this work.

[2]To whom correspondence should be addressed. Email: touteir@gwdg.de.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1606791113/-/DCSupplemental.

aSyn accumulates mainly in intraneuronal inclusions called Lewy bodies (LBs) and Lewy neurites (LNs) (7). The major component of these inclusions is aggregated and phosphorylated aSyn on serine 129 (S129) (8), which adopts fibrillar morphology compatible with that of amyloids (7). LBs and LNs are abundant in various regions of diseased brains, such as the midbrain, hippocampus, and cortex, which exhibit neuronal cell loss and degeneration (9, 10). The mechanisms triggering aSyn aggregation are not clear, nor is the relationship of aSyn aggregation with pathology. Several theories attempt to explain the toxicity of misfolded/aggregated aSyn, including the impairment in the endoplasmic reticulum trafficking pathway, dysfunction of mitochondria, blocking of protein clearance mechanisms, alteration of axonal transport, and disruption of the plasma membrane (11–13). The precise nature of the toxic aSyn species is still unclear, although it is believed that specific oligomeric species might be cytotoxic, rather than the mature aggregates (14, 15).

Based on the Braak staging hypothesis (16) and on the observation of LB pathology in young dopaminergic neurons grafted into the brains of PD patients (17), it was proposed that aSyn pathology might spread in a prion-like manner. Subsequently, in vitro and in vivo studies confirmed the ability of aSyn to be released/secreted and taken up by neighboring cells (18–20), further supporting the prion-like spreading of aSyn.

Environmental factors, such as exposure to pesticides or metals, are thought to increase the risk for PD and, possibly, other synucleinopathies (21, 22). Indeed, deregulation of metal homeostasis, including copper, manganese, zinc, iron, and lead, is characteristic of these disorders (23, 24). However, there is still no unequivocal evidence for a causative role of metals in synucleinopathies. In particular, the precise relationship between those metals and aSyn in the brain is not clear, although they bind to and promote the aggregation of the protein in vitro (25, 26).

In the present study, we systematically screened for the effect of transition metals on the intracellular aggregation of disease-associated aSyn mutants and found a significant effect of copper(II) ions (Cu^{2+}) on the H50Q mutant. We produced recombinant aggregates from WT and H50Q aSyn variants in the presence and absence of Cu^{2+}, performed detailed biophysical and structural analyses, and assessed their impact on neuronal cells. Importantly, our studies of the interplay between environmental and genetic factors affecting aSyn aggregation support dissociation between cytotoxicity and intracellular aggregation, forming the basis for future studies aimed at targeting specific species of aSyn as therapeutic strategies.

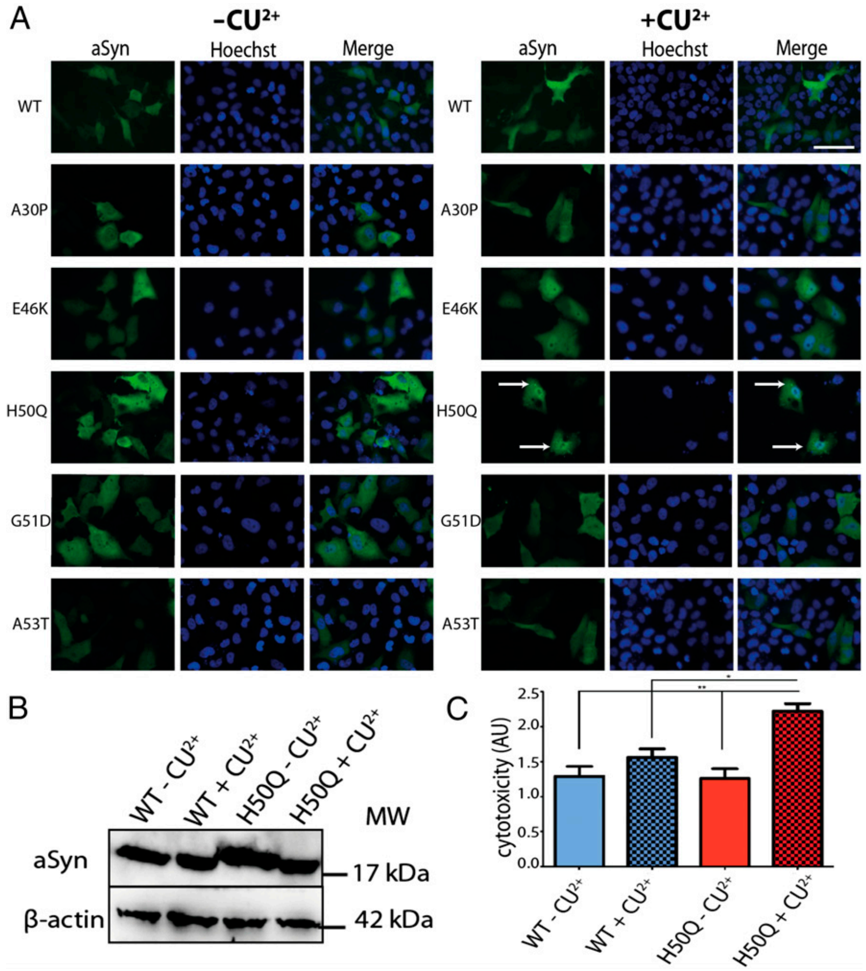


Fig. 1. Copper ions promote the intracellular accumulation of H50Q aSyn inclusions. (*A*) The addition of Cu^{2+} to the medium of cells transfected with different aSyn mutations revealed the formation of intracellular inclusions only in the presence of the H50Q aSyn mutant (white arrows). (Scale bar, 5 μm.) (*B*) The effect observed was not associated with an increase in intracellular levels of expression of the H50Q mutant, as determined by immunoblot analysis. (*C*) The interaction between Cu^{2+} and H50Q promotes an increase in cytotoxicity compared with WT $-Cu^{2+}$, WT $+Cu^{2+}$, and H50Q $-Cu^{2+}$. MW, molecular weight/molecular mass. $^{*}P < 0.05$; $^{**}P < 0.01$.

Results

Copper(II) Induces H50Q aSyn Aggregation in Human Cells. Transition metals, such as copper, interact with aSyn and modify its aggregation properties (26). To assess the interplay between Cu^{2+} and aSyn mutations, we took advantage of an established cell model of aSyn aggregation in human H4 neuroglioma cells and used it as a system to screen the effect on five disease-associated aSyn mutants. We supplemented the cell culture medium with CuCl$_2$ 24 h after transfection with plasmids encoding for the different aSyn mutants and incubated cells for 24 h. Upon immunohistochemistry and microscopy analysis we observed that Cu^{2+} promoted the formation of intracellular inclusions with the H50Q aSyn (Fig. 1*A*). This effect was not because of a change in the levels of expression of the H50Q mutant, which remained identical to those of WT aSyn (Fig. 1*B*). Interestingly, we observed an increase in cytotoxicity in cells expressing H50Q in the presence of Cu^{2+} (Fig. 1*C*). Importantly, the effect of Cu^{2+} on the H50Q aSyn mutant appears to be specific to this cation, because other relevant transition metals (24) did not induce the formation of aSyn inclusions (*SI Appendix*, Fig. S1). Therefore, we decided to focus our studies on the characterization of the interplay between H50Q aSyn and Cu^{2+}.

The Effect of Cu^{2+} on aSyn Aggregation in Vitro. First, we evaluated the effect of Cu^{2+} on the aggregation of WT and H50Q aSyn (Fig. 2*A*). The kinetics of amyloid fibril formation usually follows a sigmoidal curve that reflects a nucleation-dependent growth mechanism (27). Both in the absence and presence of Cu^{2+}, WT aSyn aggregation followed this kinetics. In the absence of Cu^{2+}, the WT aSyn reaction (WT –Cu^{2+}) exhibited a long lag phase of ∼100 h followed by a fast increase in thioflavin-t (Th-T) signal that reached maximum fluorescence intensity after ∼250 h of incubation in the conditions tested. The presence of Cu^{2+} (WT +Cu^{2+}) accelerated the aggregation reaction, shortening the lag phase to ∼50 h and, consistently, the equilibrium was reached at ∼200 h. We found a distinct scenario with the H50Q mutant. Although the aggregation of H50Q in the absence of Cu^{2+} (H50Q –Cu^{2+}) also followed a typical sigmoidal curve, the lag phase was shorter than for WT aSyn (∼25 h), followed by an exponential phase that plateaued at ∼90 h. In contrast, in the presence of Cu^{2+} (H50Q +Cu^{2+}), aggregation was strikingly faster, lacking a detectable lag phase, and Th-T signal remained very weak throughout the reaction, suggesting either the formation of nonfibrillar amyloid species or reduced aggregation. For WT aSyn, the presence of Cu^{2+} led to an increase in Th-T binding at the end of the reaction, whereas for the H50Q mutant it resulted in the opposite effect (Fig. 2*B*). To further confirm that the four samples aggregated and, especially, that the weak Th-T signal in H50Q + Cu^{2+} did not correspond to a marked reduction in the amount of deposited protein, we performed a sedimentation assay, measuring the protein remaining in the supernatant after centrifugation. In this assay, all samples showed similar amounts of insoluble material (Fig. 2*C*), despite the differential binding to Th-T. Finally, we assessed the potential dependence of aSyn aggregation on Cu^{2+} concentration, from sub- to suprastoichiometric concentrations, using the Th-T-binding assay. With both aSyn variants, an increase in Cu^{2+} concentration caused a shortening of the lag phase, although this effect was limited to the equimolar ratio. Exposing aSyn to suprastoichiometric Cu^{2+} concentration did not significantly accelerate the aggregation kinetics of WT or H50Q (*SI Appendix*, Fig. S2) relative to an equimolar ratio. Another neurodegeneration-related metal, iron, was introduced in the aggregation assay, but no effect was observed, indicating that the effect of Cu^{2+} on aSyn aggregation is probably specific for this cation (*SI Appendix*, Fig. S2).

Next, we performed biophysical characterization of the resulting aggregates using circular dichroism (CD), attenuated total reflectance–Fourier transform infrared spectroscopy (ATR–FTIR), 1-anilinonaphthalene-8-sulfonic acid (ANS) binding, and transmission electron microscopy (TEM). We analyzed the secondary structure of the samples by CD before and after aggregation. As expected, freshly prepared aSyn samples, in the presence or absence of Cu^{2+}, exhibited a characteristic negative CD peak at 200 nm, consistent with the lack of significant secondary structure in aSyn, because it is known to be an intrinsically disordered protein (*SI Appendix*, Fig. S3*A*). After aggregation, the CD spectra strongly changed, showing a minimum at around 216 nm, consistent with the presence of β-sheet structure. Deconvolution of the ATR–FTIR raw spectra of aggregated states into the corresponding contributing components ($r^2 < 0.998$ in all cases) indicated that both WT –Cu^{2+} and WT +Cu^{2+} shared similar secondary structure elements (*SI Appendix*, Fig. S3*B*). The peak at 1,623–1,641 cm^{-1}, assignable to intermolecular β-sheet structure, dominated both spectra (Table 1). However, the ratio between the β-sheet and the disordered spectral component was higher in the absence than in the presence of Cu^{2+} (2.0 and 1.3). The intermolecular β-sheet component was also the main contributor to the H50Q –Cu^{2+} spectra with a ratio between this signal and that of the disordered state similar to that observed with WT –Cu^{2+} (1.9). In contrast, in the case of H50Q +Cu^{2+}, the intermolecular β-sheet signal did not dominate the spectra but contributed the same amount as the disordered regions (ratio of 1.0) with an additional increase in β-turn content by 27%. Thus, the presence of Cu^{2+} resulted in fewer ordered aggregates for both the WT and the H50Q mutant, but the decrease in β-structure was more pronounced for H50Q aSyn. Structural differences of the

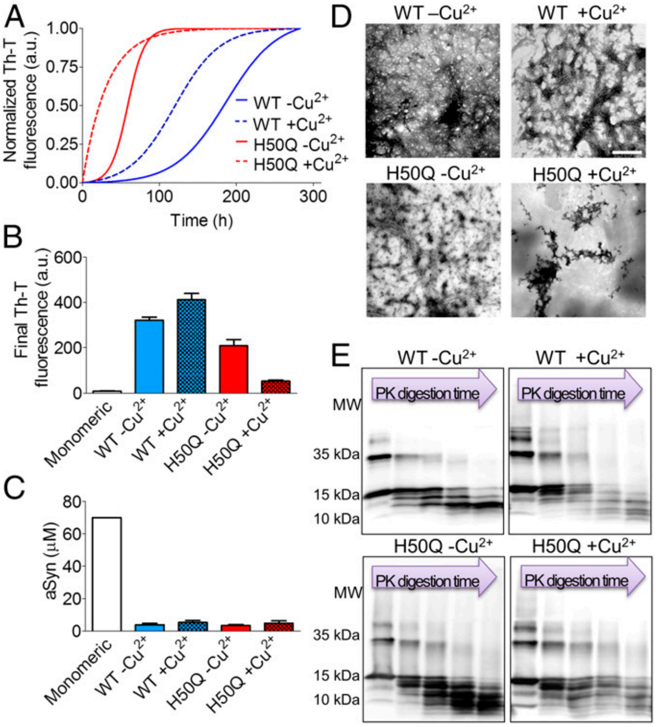


Fig. 2. Biophysical characterization of the effect of copper ions on the aggregation of WT and H50Q aSyn. (*A*) Aggregation kinetics of WT and H50Q aSyn with and without Cu^{2+} evaluated by normalized Th-T binding at the indicated time points. Fluorescence intensity peak at 482 nm was used to probe for amyloid formation. (*B*) Fluorescence intensity at the end-time point of the Th-T binding assay. (*C*) Remaining amount of soluble aSyn determined by sedimentation assay at the end time point of aggregation. (*D*) TEM images of the resulting aggregates stained with uranium acetate. (Scale bar, 500 nm.) (*E*) Partial PK digestion of the aggregates visualized by Coomasie staining of SDS/PAGE gels. Each lane represents different digestion times: 0, 1, 2.5, 5, and 10 min.

Table 1. Secondary structure composition of the fibrils formed by WT $-Cu^{2+}$, WT $+Cu^{2+}$, H50Q $-Cu^{2+}$ and H50Q $+Cu^{2+}$

Sample	β-Sheet intermolecular	Disordered	Turns
WT $-Cu^{2+}$	56.73	27.63	15.63
WT $+Cu^{2+}$	46.10	36.33	17.56
H50Q $-Cu^{2+}$	54.61	29.06	16.32
H50Q $+Cu^{2+}$	38.97	39.17	21.84

The assignments are represented in percentage and were attributed to each peak observed according to Susi and Byler (75).

aggregates were also determined using the ANS binding assay (*SI Appendix*, Fig. S3*C*). With both aSyn variants, the presence of Cu^{2+} increased the ANS fluorescence of the aggregates (more than 1.5-fold and 3-fold increase in WT and H50Q, respectively), suggesting that Cu^{2+} increased the exposure of hydrophobic surfaces in the resulting aggregates. The initial soluble species did not show differences in ANS binding among them, suggesting that the impact of Cu^{2+} on the exposure of hydrophobic clusters in aSyn assemblies occurs during the aggregation process. The morphology of the aggregated samples was further investigated by TEM (Fig. 2*D*). Both WT $-Cu^{2+}$ and WT $+Cu^{2+}$ formed well-defined and classic amyloid filaments. This scenario was different for the H50Q mutant, where the presence of Cu^{2+} induced the formation of nonfibrillar amorphous aggregates. Overall, the secondary structure found in the aggregates and their morphology correlated well with their relative ability to bind Th-T and is indicative of a different conformational effect of Cu^{2+} in WT and in H50Q aSyn. In addition, we examined the resistance of each aSyn aggregate to proteolysis by performing a time-dependent partial proteinase K (PK) digestion. We found that the species assembled in presence of Cu^{2+} displayed decreased resistance to proteolysis, whereas those formed without Cu^{2+} exhibited a strongly resistant core (Fig. 2*E*). Notably, in all cases the most susceptible region was the C terminus, as evidenced by the use of an antibody recognizing this region of the protein (*SI Appendix*, Fig. S4).

Because spreading of disease pathology might depend on the ability of the transmitted protein to seed aggregation in the recipient cell, we investigated the capacity of the aggregates of WT and H50Q aSyn, formed in the absence and presence of Cu^{2+}, to influence the aggregation of monomeric aSyn. To this end, the time-dependent decrease in the concentration of monomeric aSyn was monitored using NMR spectroscopy, which allows the highly sensitive detection of the initial steps of aggregation. Species formed without Cu^{2+} displayed a strong seeding activity, rapidly decreasing the amount of free monomers in solution. However, this activity was reduced when the aggregates formed in the presence of Cu^{2+} were used as seeds. Indeed, H50Q $+Cu^{2+}$ did not induce significant monomer consumption, indicating the absence of seeding activity at the time points analyzed (Fig. 3).

aSyn–Cu^{2+} Complexes at the Histidine 50 Site. Because the histidine 50 (H50) residue is key for anchoring Cu^{2+} binding to aSyn, the results suggested that the critical step for the effects observed in Cu^{2+}-mediated aSyn aggregation might be the formation aSyn–Cu^{2+} complexes at this site. In vitro NMR (28) and electron paramagnetic resonance (26, 29) experiments have suggested that the ion binds in a square planar or tetrahedral distorted geometry, involving 2 or 3 N ligands and 2 or 1 O ligands, respectively (29–31). Models A and B (*SI Appendix*, *SI Molecular Simulations*) have been proposed in the literature (28, 29, 32). The Cu^{2+} ion binds to the H50 side chain in both of them. The three additional ligands are the H50 amide group, a water molecule, and either V48 carbonyl O (model A) or V49 deprotonated amide (model B). Quantum mechanics/molecular mechanics simulations were carried out to test their structural stability (*SI Appendix*, *SI Molecular Simulations*). All of the coordination bonds exhibited relatively small fluctuations around their average values, except for the Cu^{2+}–V48 carbonyl and Cu^{2+}–V49 amide. This finding suggests an interconversion between the two forms. Added to the models in which H50 participation in binding to aSyn would involve the formation of an intramolecular macrochelate or would lead to intermolecular metal-bridged protein molecules (33), our calculations suggested that: (*i*) both models could be significantly populated binding modes; (*ii*) aSyn conformation is preorganized to accommodate the square planar conformation of Cu^{2+}, suggesting the latter may bind using a conformational selection mechanism (34); and (*iii*) Cu^{2+} is able to bind independently to the imidazole nitrogen of H50 in a N2O2 (A) or N3O1 (B) interconverting ligand donor sets, with the additional ligands mentioned above. In contrast, the H50Q variant lacks all these modes of Cu^{2+} binding, indicating differences with the WT protein, possibly underlying different conformations. Other models, initially considered, were discarded (*SI Appendix*, *SI Molecular Simulations*).

aSyn Aggregates Are Cytotoxic When Applied to Primary Neuronal Cultures. The aSyn aggregates we generated in the various conditions had distinct conformational and seeding properties. Thus, we next asked whether they might have differential effects when applied to cultured cells. We exposed rat primary cortical cultures for 13 d to the species resulting from aSyn aggregation. We found that all aSyn species were toxic on neuronal cultures as assessed by the release of adenylate kinase to the medium because of cell damage (Fig. 4*A*). Quantification of the remaining cells and of the percentage of condensed nuclei also revealed cell loss after treatment with aSyn aggregates (Fig. 4 *B* and *C*). Immunostaining against microtubule-associated protein 2 (MAP2) confirmed that the reduced number of cells was a result of neuronal loss. We also found a reduction in the average neurite length, particularly pronounced in the cultures treated with WT $-Cu^{2+}$ and H50Q $-Cu^{2+}$ (Fig. 4*D*). Consistently, these cultures displayed reduced levels of MAP2, as determined by immunoblot analysis (Fig. 4*E*). Because MAP2 is specifically present in dendrites, we also used Tau as an axonal marker of mature neurons (35) to assess whether these projections could also be affected, but found no significant differences (Fig. 4*E*). To measure possible synaptic damage induced by the aSyn aggregates, we quantified the levels of presynaptic (synapsin and synaptophysin) and postsynaptic (postsynaptic density protein 95) markers (*SI Appendix*, Fig. S5). We observed a significant reduction of these markers in cells that were treated with WT $-Cu^{2+}$ and H50Q $-Cu^{2+}$. Because the application of

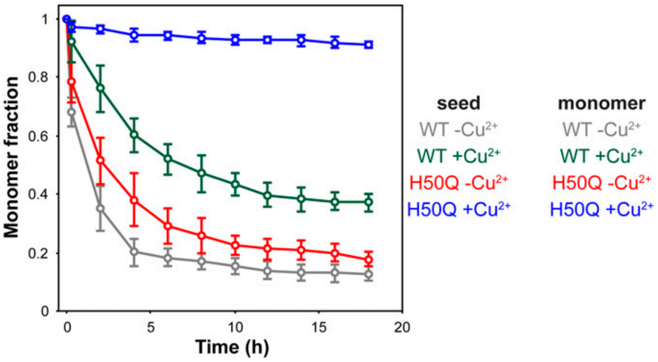


Fig. 3. Copper ions reduce the seeding capacity of aSyn aggregates. Time-dependence of 1H NMR spectra of monomeric aSyn upon addition of aggregation seeds [4% (vol/vol)]. The integrated intensity of ^1H signals in two regions (0.50–1.05 ppm, 6.50–8.00 ppm) was calculated and plotted as a function of time. The solution with 70-μM monomeric protein (WT or H50Q) contained either no Cu^{2+} or an equimolar concentration of Cu^{2+}, as indicated.

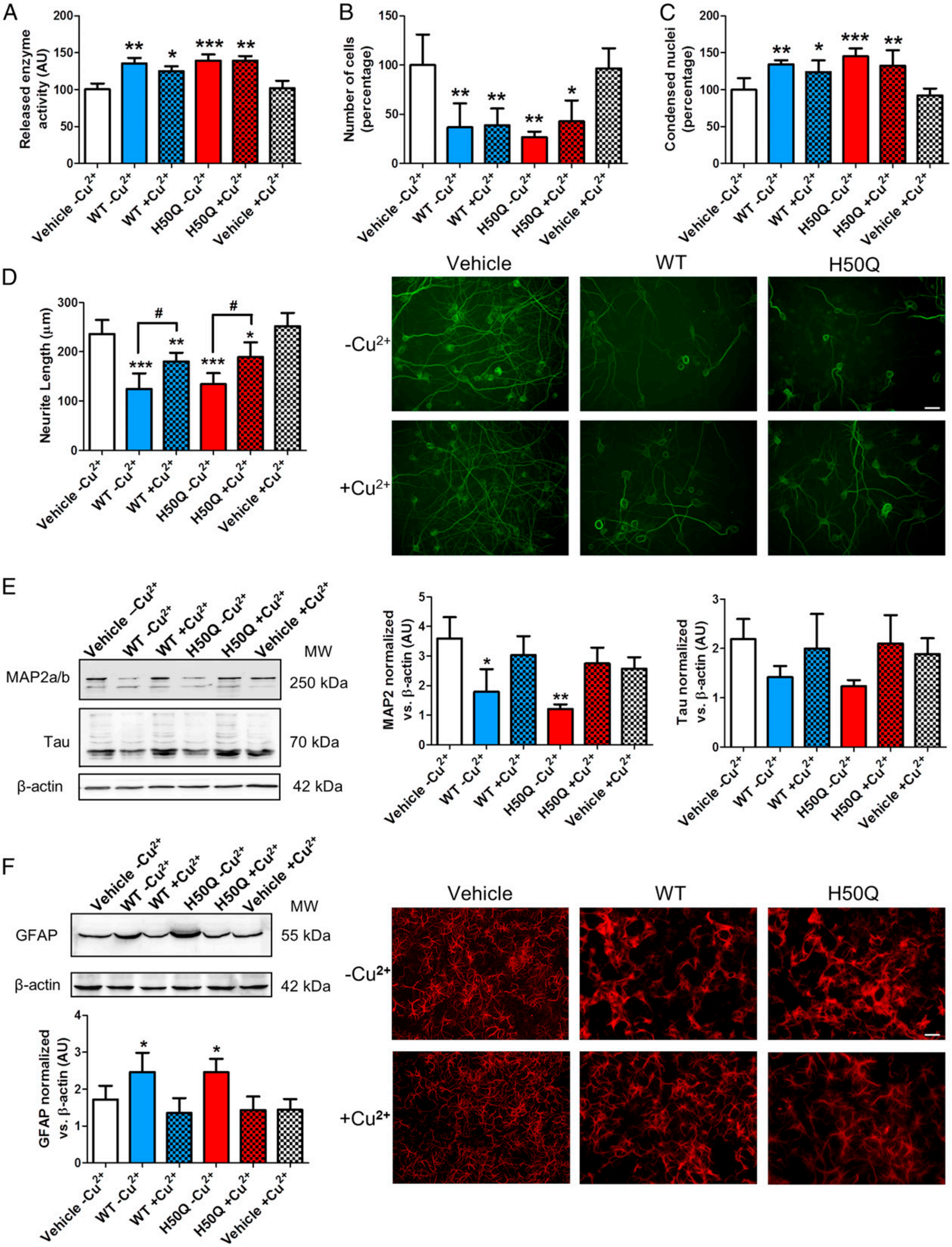


Fig. 4. Cellular effects as a result of the exogenous application of aSyn aggregates. (*A*) Cytotoxicity quantification in cultures exposed to aggregates compared with the vehicle control, measured as the presence of adenylate kinase in the medium. (*B*) Cell death quantification, measured as the reduction in the number of cells. (*C*) Increase in the percentage of condensed nuclei in the same experiment represented in *B*. (*D*) Quantification of neurite length of cells exposed to aggregates compared with the vehicle control (*Left*) and representative images of MAP2 stained cultures, used to obtain the former parameter (*Right*). (Scale bars, 20 μm.) (*E*) Quantification of MAP2 and Tau levels by Western blot. (*F*) Astrogliosis caused by the application of aSyn aggregates, measured by Western blot quantification of GFAP (*Left*), and representative images evidencing the presence of reactive astrocytes (*Right*). (Scale bars, 20 μm.) *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$ for comparison between samples and their appropriate vehicle control. $^{\#}P < 0.05$; $^{\#}P < 0.01$ for comparison between different samples.

pathological aSyn species has been reported to induce a glial response in vivo (20, 36), we investigated the state of the astrocytes by probing for glial fibrillary acidic protein (GFAP). An increase in GFAP levels could only be detected in cells exposed to WT –Cu^{2+} and H50Q –Cu^{2+}, likely because of the development of reactive astrocytes, as GFAP staining revealed a phenotype compatible with reactive astrocytes (37) (Fig. 4*F*).

Overall, we demonstrated that aggregated recombinant aSyn species promoted neuronal death and damage. Importantly, the effects observed depended on the type of aggregate that was used, being more acute in the case of aggregates formed without Cu^{2+}.

aSyn Aggregates Induce the Formation of Phosphorylated aSyn Inclusions. Next, we analyzed the capacity of the different aggregated species of aSyn to induce the conversion of endogenous aSyn into misfolded/aggregated forms in cultured cells. The experimental set-up was the same as described above. We performed immunostaining with an antibody against phosphorylated aSyn (p-aSyn) on S129, because aggregated aSyn is commonly phosphorylated in this residue, as in LBs (8). This approach also enabled us to specifically detect intracellular aSyn, and not external protein, as this was not phosphorylated and, therefore, could not react with the antibody against p-aSyn (38). Although we observed a wide distribution of p-aSyn in the cell body of vehicle-treated cells (Fig. 5*A*), those exposed to aSyn aggregates displayed more intense and concentrated signal, corresponding to inclusions positive for p-aSyn (Fig. 5*A*). Cells treated with H50Q –Cu^{2+}, and particularly with WT –Cu^{2+} species, contained fewer inclusions. In contrast, those treated with WT +Cu^{2+} and H50Q +Cu^{2+} presented several p-aSyn^{+} inclusions, which we classified as neuritic, perykarial, and nuclear (Fig. 5*B*). Because of technical limitations in the quantification of inclusions in neurites, we performed a semiquantitative assessment of the frequency of all types of inclusions in Table 2. We also performed quantification of total p-aSyn on S129 by immunoblot (Fig. 5*C*). Strikingly, the occurrence of intracellular inclusions was highly specific to the type of aggregates applied to the cells. We observed neuritic inclusions in all treatments, although very few were observed in cells treated with WT –Cu^{2+} and H50Q –Cu^{2+}. With the latter treatment (H50Q –Cu^{2+}), we also observed dispersed perykarial inclusions. Cultures treated with aggregates formed in the presence of Cu^{2+} presented more inclusions in neurites and perykaria. Specifically, cells treated with H50Q +Cu^{2+} displayed the largest number of all types of inclusions (Table 2). To rule out issues of unspecificity of the p-aSyn antibody (39), we validated the observed staining with an antibody against total aSyn (*SI Appendix*, Fig. S6). The insoluble nature of the induced p-aSyn inclusions was confirmed using an established fixation protocol during the immunostaining that included the presence of Triton X-100 to remove soluble proteins and structures (38). Furthermore, detection of cellular aSyn in the insoluble fraction in the cultures displaying more inclusions (those treated with WT +Cu^{2+} and H50Q +Cu^{2+}) corroborated the presence of aggregated aSyn (Fig. 5*D*).

We next aimed at determining which particular aSyn protein composed the inclusions observed. The use of an antibody specific for rat aSyn revealed that part of the inclusions present in neurites and perykaria are mainly formed by human recombinant protein that was internalized, whereas all nuclear inclusions were made up of endogenous aSyn (Fig. 6*A*). To further investigate the presence of exogenous aSyn in the nuclear inclusions, we labeled the aggregates formed by H50Q +Cu^{2+} in vitro, as those were the most active in promoting these types of inclusions, and applied them to the cells using the same procedure as above. Immunostaining against p-aSyn confirmed the absence of exogenous aSyn in the nuclear inclusions (Fig. 6*B*). To unequivocally validate this finding, labeled aSyn species were applied to neurons expressing GFP and live cell imaging was performed. Recombinant material was internalized into the cells, but was unable to cross the nuclear membrane (Fig. 6*C*). Therefore, we confirmed that nuclear p-aSyn inclusions were exclusively formed by endogenous aSyn.

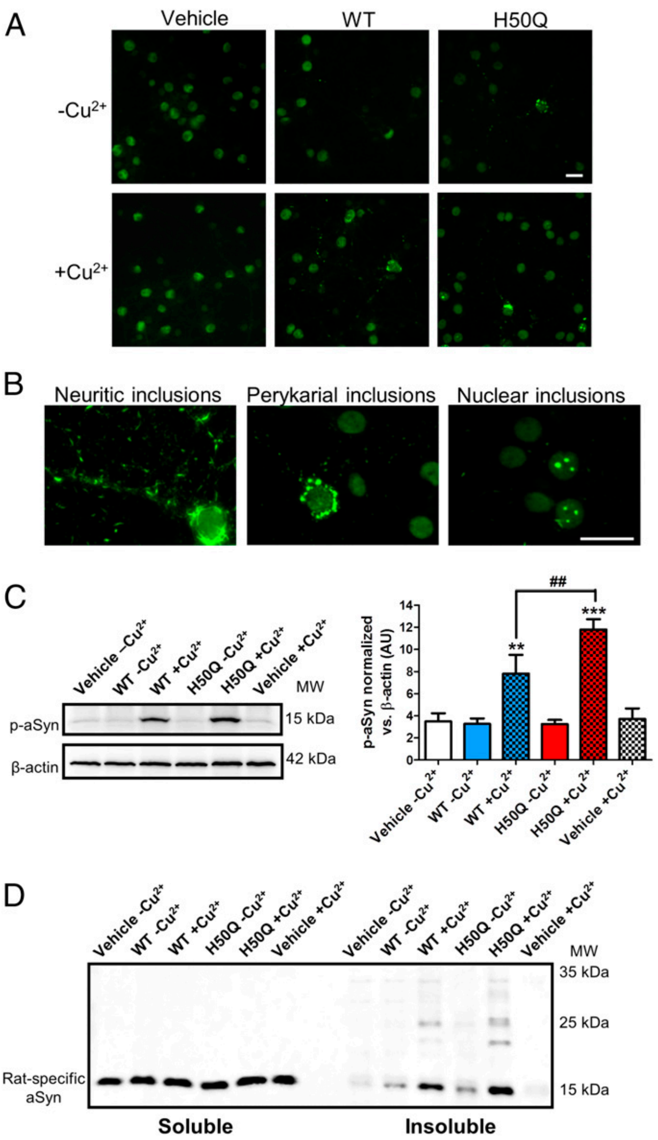


Fig. 5. Formation of S129 p-aSyn^{+} inclusions upon exposure to aSyn aggregates. (*A*) Representative images of each culture showing the presence of p-aSyn (S129) inclusions in cells treated with exogenous aggregates. (Scale bars, 20 μm.) (*B*) Distinct types of p-aSyn inclusions classified as neuritic, perykarial and nuclear. (Scale bars, 20 μm.) (*C*) Western blot measurements of p-aSyn (S129) in cell lysates. (*D*) Soluble/insoluble distribution of the cellular aSyn detected with a rat-specific aSyn antibody. $^{**}P < 0.01$; $^{***}P < 0.001$ for comparison between samples and their appropriate vehicle control. $^{\#\#}P < 0.01$ for comparison between different samples.

We performed further characterization of the p-aSyn^{+} inclusions. On the one hand, inclusions present in neurites were visible in Tau^{+} projections, confirming their axonal localization. They appeared as puncta or short filaments and, part of them, showed a discontinuous linearity that could be followed up to the neuronal cell body (Fig. 5*B* and *SI Appendix*, Fig. S7*A*). On the other hand, perykarial inclusions were examined by costaining with Lamin B1, which stains the nuclear envelope. We confirmed their presence in the cell body and around the cell nucleus (*SI Appendix*, Fig. S7*B*). Both neuritic and perykarial inclusions appeared positive for heat-shock protein 90 (Hsp90), a molecular chaperone commonly present in LBs and other protein

Table 2. Table presenting semiquantification of the frequency of p-aSyn (S129) inclusions in each culture measured from immunostaining pictures

aSyn species	Neuritic	Perykarial	Nuclear
WT $-Cu^{2+}$	+	—	—
WT $+Cu^{2+}$	++	+++	+
H50Q $-Cu^{2+}$	+	+	—
H50Q $+Cu^{2+}$	+++	++	+++

aggregates (40) (*SI Appendix*, Fig. S7 *A* and *B*). Finally, the localization of nuclear inclusions was confirmed by costaining with Hoechst, a widely used nuclear stain. On average, one to six p-aSyn$^+$ round structures appeared in the nuclei of cells containing inclusions. These were usually present in areas that were Hoechst$^-$. The lack of costaining with nucleolar markers (nucleolin and fibrillarin) ruled out their presence in the nucleoli (*SI Appendix*, Fig. S7*C*). However, we could detect costaining with ubiquitin, one of the main hallmarks of disease-associated protein deposits (41). Despite the presence of p-aSyn inclusions, we did not detect significant alteration in markers of the heat shock response or of the endoplasmic reticulum stress in our experimental conditions (*SI Appendix*, Fig. S8).

Discussion

PD is characterized by the accumulation of aSyn$^+$ fibrillar inclusions, and is associated with an imbalance in metal ion homeostasis (e.g., Cu, Zn, and Fe) (24). In the present study we first screened the effect of Cu^{2+} over different disease-associated aSyn mutants. After the exogenous addition of this cation, H50Q aSyn was the only mutant exhibiting intracellular accumulation. The interplay between this variant and Cu^{2+} was highly specific, because upon testing other relevant metals (24), no H50Q aSyn inclusions were observed. This unique synergy is not surprising, considering that the N-terminal region of aSyn in which H50 acts as a key metal anchoring residue constitutes the preferential binding interface for Cu^{2+} (28, 26, 33, 42). Previous studies, using other cell models, revealed that exposure to Cu^{2+} promotes inclusion formation and toxicity in aSyn-expressing cells (43) and that the H50Q aSyn mutant exhibits a higher aggregation propensity when expressed intracellularly (44). Consistently, in our screen we detected a specific increase in aSyn aggregation only with the combination of H50Q aSyn and Cu^{2+}, an effect that exacerbates cytotoxicity.

In cell-free conditions and in agreement with previous evidence (26), we show here that Cu^{2+} specifically increases the aggregation propensity of WT aSyn, reducing the lag phase and accelerating self-assembly. However, the effect of Cu^{2+} in the aggregation of H50Q aSyn is clearly different from that in the WT protein, resulting in a very rapid formation of amorphous aggregates that—despite containing intermolecular β-sheet structure—bind to Th-T with low affinity, suggesting major structural differences that include, among other features, a larger increase in the exposure of hydrophobic surfaces as revealed by ANS binding. These differences are probably because of the substitution of a key metal binding site (H50) in the H50Q mutant, which mainly coordinates Cu^{2+}, in comparison with other metal cations that may have higher affinities for other residues in the aSyn sequence (45). Despite the reported variability, all aSyn species share a common aggregation core with a less-compacted region mainly composed by the C terminus of the polypeptide. This finding is in agreement with previous data reporting the poor participation of the C-terminal domain in fibril formation, which indeed remains mainly disordered instead of acquiring the typical β-sheet structure (46). The definition of aSyn as a prion-like protein is intimately linked to an intrinsic seeding activity of, at least, some misfolded conformations. We observe that the distinctive structural features introduced by the presence of copper ions are mirrored in differential seeding activities. Among the species tested, the one formed by H50Q $+Cu^{2+}$ is the only one not acting as an efficient seed in our conditions. The aggregates formed by H50Q $+Cu^{2+}$ differ significantly

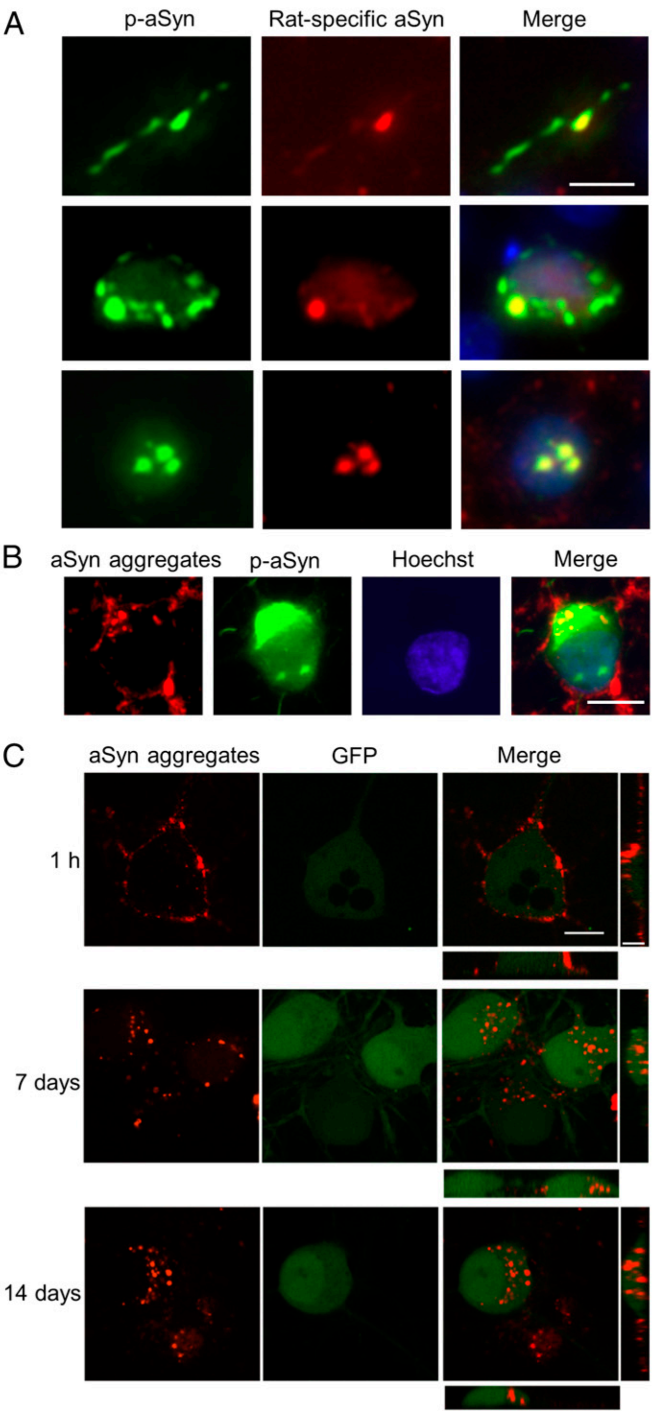


Fig. 6. Distribution of exogenous and endogenous aSyn in the S129 p-aSyn inclusions. (*A*) Neuritic (*Top*), perykarial (*Middle*), and nuclear (*Bottom*) p-aSyn inclusions show partial or total colocalization with rat-specific aSyn. (*B*) Presence and absence of labeled aSyn aggregates in perykarial and nuclear p-aSyn inclusions. (*C*) Confocal live cell imaging of cultures exposed to labeled aSyn aggregates and infected with a GFP-expressing lentivirus demonstrating the cellular distribution of the exogenous species at different time points after exposure. (Scale bars, 10 μm.)

from the other ones, being mainly amorphous, with low Th-T binding and lacking seeding activity. Taken together, these findings show that the combination of the H50Q mutation and Cu^{2+} promotes the formation of atypical aSyn aggregates, probably representing off-pathway species.

As mentioned above, H50 is crucial in the binding of Cu^{2+} to aSyn. Our quantum mechanics/molecular mechanics simulations suggest that the protein backbone and a water molecule complete the coordination of Cu^{2+}. The participation of H50 in metal binding may induce different folding of the protein through the formation of this macrochelate or may act as an intermolecular metal-anchoring point bridging different protein molecules, explaining how H50–Cu^{2+} complexation might impact on aSyn aggregation (33). Added to that, the independent, noninteractive H50–Cu^{2+} binding mode described in our study and those reported for Cu^{+} binding to aSyn (47) indicate that the H50 site constitutes a unique target for in vivo Cu^{2+}/Cu^{+} redox chemistry and oxidative damage, a reaction that might lead to a cascade of structural alterations promoting oligomerization and the subsequent amyloid aggregation of aSyn (33). Thus, H50 might play an important role in the structures adopted by soluble aSyn at the beginning of the aggregation reaction and, likely, of the polymerization rates and the final conformation in the resulting aggregates, which would explain the conformational diversity of WT $+Cu^{2+}$ and H50Q $+Cu^{2+}$ aggregates.

A relevant approach to assess aSyn propagation consists of the induction of aSyn aggregation, upon treatment with preformed fibrils or aggregates, in cultured cells without previous aSyn pathology (48–50). Here we used a similar approach with the objective to further characterize the aSyn species being studied, considering their distinct biophysical features. We used primary neuronal cultures as others have done in the past (51). We found that, although all types of recombinant aSyn aggregates promote cell death, the strongest effect occurs with those assembled in the absence of Cu^{2+}. The specific decrease in neurite length suggests deterioration of neuronal processes precedes cell death. The absence of significant alteration in Tau levels indicates that dendritic arborization is more susceptible to aSyn toxicity than axons. At the end point of our experiments, the axonal network is mature and complex (35), probably exhibiting increased resistance to external insults. The damage induced by aSyn aggregates is also evident in synaptic terminals. However, the toxicity linked to exogenous aSyn is still not fully established, probably because it depends on the cellular type, on the endogenous levels of aSyn, or on the type and properties of the aSyn species used (38, 49, 52–54). In the present study, two of four types of mature aSyn aggregates tested were significantly more toxic, strongly affecting neuronal structures. Interestingly, along with the neuronal damage observed, we also detected reactive astrocytes in our cultures. Our findings are consistent with studies demonstrating that incubation of astrocytic cultures with neuronal-derived aSyn triggered aSyn aggregation in these cells, along with changes in the levels of cytokines and chemokines, a sign of an inflammatory response (55). In our study, the inflammatory response we observe is intimately linked to neuronal damage but independent of inclusion formation.

Furthermore, we investigated the intrinsic capacity of the species tested to induce aSyn aggregation in the recipient cells. The use of an antibody against S129 p-aSyn reveals a basal nuclear aSyn phosphorylation in naïve neurons, compared with a more punctate signal for total aSyn that is mostly present in the neuronal processes. This finding is in agreement with the well-known presynaptic location of aSyn and the enrichment of the phosphorylated form in the nuclear compartment (4, 56). According to previous reports, we found that neurons exposed to aSyn aggregates develop S129 p-aSyn inclusions, displaying increased levels of insoluble aSyn (38, 49). However, in contrast to other studies (39), overexpression of the cellular aSyn or expression of the human form was not required. The inclusions formed could be classified depending on their localization in the cell. Neuritic and perykarial inclusions have also been observed by others (49, 51, 57) and colocalize with Hsp90, resembling pathology-related aggregates (40). These inclusions are formed by both endogenous aSyn and internalized recombinant species. On the other hand, we found nuclear inclusions, which are excluded from the nucleoli. Despite sporadic reports of the occurrence of aSyn aggregates in the nuclei (58, 59), the morphology we observe has not, to the best of our knowledge, been previously described. Nuclear inclusions are ubiquitinated and exclusively formed by endogenous aSyn, as recombinant species were not present in the nuclear compartment. Integrating our findings with data from previous reports (38), aSyn inclusion formation—upon stimulation from the outside—could follow a progression pattern in which exogenous species are internalized, leading to the appearance of inclusions in the projections, and then in the cell body, upon partial sequestration of the endogenous protein. Subsequently, nuclear inclusions appear, likely through a different mechanism because they do not contain detectable amounts of the exogenous protein. Given the apparent lack of seeding capacity of H50Q $+Cu^{2+}$ aggregates in vitro, additional studies will be required to unravel the cellular mechanisms by which endogenous S129 p-aSyn inclusions form in the nucleus, once the cytoplasm has been invaded by exogenous material.

In general, inclusion formation upon aSyn transmission is a phenomenon highly dependent on the type of exogenous species the cell encounters. We found that WT $-Cu^{2+}$ aggregates rarely induce inclusion formation in the cells, in contrast with previous studies (49, 51), possibly because of different experimental approaches. In contrast, cells treated with aggregates formed in the presence of Cu^{2+}, particularly H50Q $+Cu^{2+}$, promote the formation of the various types of inclusions we identified. However, we cannot explain why previous studies failed to observe nuclear aSyn inclusions. A possible explanation could be the actual nature of the H50Q $+Cu^{2+}$ aggregates, which display different biophysical features. Recent studies suggest that aSyn can adopt various conformations with distinctive characteristics, including distinct propagation capacity (53). These discoveries led to the assumption that aSyn could assemble as different strains in vitro, a hypothesis that has been recently tested in rats (58). In the present study, we demonstrate that structurally different aSyn species exert distinct effects when applied to neuronal cells, and that these effects might be because of the presence of Cu^{2+} during the aggregation process, a situation that may occur in the brain, a result of different environmental stimuli.

A final observation from our experiments is the dissociation between toxicity and inclusion formation, in agreement with other studies (52, 60, 61). WT $+Cu^{2+}$ and H50Q $+Cu^{2+}$ are the less-damaging aggregates and the ones that promote to a larger extent inclusion formation. Although so far we cannot identify the causes of this fact, similar observations were previously reported (62–66). Inclusion bodies were demonstrated to be protective in the case of the Huntington's disease-related protein huntingtin (67). This was also observed in the case of nuclear huntingtin inclusions (68). Importantly, the nuclear aSyn inclusions we describe herein resemble those found in Huntington's disease-affected brains, given that they are ubiquitinated and have similar size and morphology (69).

It is also possible that aSyn species exert toxicity from the outside, without the need to enter the recipient cell. In that case, the most aggressive species would induce cell death, avoiding the lag time necessary to be internalized and trigger inclusion formation. Indeed, toxicity exerted by exogenous aSyn species has been reported, particularly in the case of aSyn oligomers (13). The disruption of membrane integrity and permeability is one of the possible molecular mechanisms involved (70); this could promote calcium influx and deregulation of cellular homeostasis (52). Although unlikely, because we used mature aggregates at the end point of the aggregation reaction, it is possible that the presence of residual oligomers in our samples may contribute to the toxic effects observed. Finally, it is also possible that the intrinsic toxicity of the recombinant species hamper their own self-propagation. Recently, it was demonstrated that an increase in aSyn aggregation is not associated with stronger neuronal injury, arguing that efficient propagation of the protein is an active process, where healthy cellular metabolism and connections are necessary (61). In addition, injection of two different aSyn strains in vivo demonstrated that although both strains induce the formation of p-aSyn inclusions, this effect is reduced in the case of the most toxic strain, confirming that under certain conditions, there is an inverse correlation between seeding and damaging capacity (58).

In total, our study characterized the effects of various forms of aSyn aggregates in the intracellular milieu. Ultimately, the unambiguous dissociation between inclusion formation and toxicity will be essential to guide the development of future therapeutic approaches against synucleinopathies.

Materials and Methods

Purification of aSyn. aSyn WT and H50Q were produced as previously described (71, 72), using ammonium sulfate precipitation and ion-exchange chromatography. For further details, see *SI Appendix, SI Materials and Methods*.

Aggregation of aSyn. aSyn was aggregated at 70 μM in the presence or absence of $CuCl_2$ at 70 μM. Aggregation kinetics were conducted at 37 °C under agitation and followed by Th-T binding. Sedimentation assays were carried out by measuring the remaining soluble protein at the end time point of aggregation. Aggregation kinetics of aSyn in the presence of different Cu^{2+} concentrations and iron were performed in a 96-well plate with Teflon-balls, as previously described (73). For live cell imaging, aSyn aggregates were labeled with ATTO 590 NHS-ester (Atto-Tec) as previously described (50). For further details, see *SI Appendix, SI Materials and Methods*.

Circular Dichroism. CD spectra were obtained at a spectral resolution of 1 cm^{-1} and 15 $nm \cdot min^{-1}$ scan rate and collected over the 190- to 260-nm wavelength range at 25 °C using a Jasco 810 spectropolarimeter with a quartz cell of 0.1-cm path length. For the soluble state measurements, solutions at 10 μM were filtered through a 0.22-μM filter. At the end time point of aggregation, samples were diluted to 10 μM and vigorously homogenized in PBS before the measurements.

FTIR. ATR–FTIR spectra of the aSyn aggregates were obtained with Bruker Tensor 27 FTIR spectrometer (Bruker Optics). For further details, see *SI Appendix, SI Materials and Methods*.

ANS Binding Assays. The assays were performed by incubating 5 μM of aggregated samples with 20 μM of ANS in PBS. ANS fluorescence was measured on an ISS K2 spectrofluorometer. Samples were excited at 370 nm, and the emissions were collected from 400 to 600 nm using slits of 0.5 cm. The intensity at 374 nm was used to compare the extent of ANS bound to each sample. Assays with the soluble proteins were carried out by incubating 70 μM of the monomeric samples with 70 μM of ANS (soluble Transthyretin was used as positive control).

Transmission Electron Microscopy. Aggregated samples were adsorbed onto carbon-coated grids, rinsed with water, and stained with 2% (wt/vol) uranyl acetate. The samples were exhaustively scanned and representative fields were imaged in a Hitachi H-7000 TEM operating at an accelerating voltage of 75 kV.

PK Digestion Analysis. PK digestions of aSyn aggregates at the indicated time points were conducted at 37 °C. The resulting fragments were analyzed in Tricine-based protein electrophoresis. For further details, see *SI Appendix, SI Materials and Methods*.

NMR Spectroscopy. NMR experiments were recorded at 37 °C on a Bruker Avance 600 MHz spectrometer. To assess the seeding capabilities of different aggregates, a 4% concentration (vol/vol) of preformed aggregates, as described above WT $-Cu^{2+}$, WT $+Cu^{2+}$, H50Q $-Cu^{2+}$, H50Q $+Cu^{2+}$, were added to a 70-μM concentration of the fresh monomeric protein (in the absence or presence of an equimolar concentration of Cu^{2+}). The NMR buffer was 25 mM Hepes, 100 mM NaCl at pH 7.4, 90% H_2O/10% D_2O (vol/vol). ^1H NMR spectra were recorded every 2 h for at least 20 h after the addition of seeds. The reduction in signal intensity reports on the decrease in the concentration of monomeric protein.

Molecular Simulations. Details on the methodology are found in *SI Appendix, SI Molecular Simulations*.

Human Cell Cultures. Human H4 cells were treated, 24 h after transfection, with 1 μM of $CuCl_2$. For further details, see *SI Appendix, SI Materials and Methods*.

Primary Cell Cultures. Preparation of primary cortical cultures from WT rats was carried out as previously described (74). Cells were treated with 1 μM of aSyn aggregates. For live cell imaging, neurons were infected with a GFP-overexpressing lentivirus previous to the application of aSyn prelabeled aggregates. For further details, see *SI Appendix, SI Materials and Methods*.

Immunostaining Studies. Cells were fixed with paraformaldehyde and, when appropriate, with addition of 1% Triton X-100 to remove soluble structures (38). After permeabilization and blocking, cells were incubated with primary antibodies overnight at 4 °C and subsequently with Alexa Fluor-conjugated secondary antibodies. Leica DMI 6000B and Olympus XI81 microscopes were used for imaging. For further details, see *SI Appendix, SI Materials and Methods*.

Immunoblotting Analysis. Cells were lysed with a native buffer containing 1% Triton X-100 and cell debris was discarded by centrifugation. Cell lysates were separated in SDS/PAGE gels and transferred to blotting membranes, which were subsequently blocked with 5% (wt/vol) milk. Primary and secondary antibodies were incubated overnight at 4 °C and for 1 h at room temperature, respectively, and membranes were developed. For further details, see *SI Appendix, SI Materials and Methods*.

ACKNOWLEDGMENTS. We thank Prof. Dr. I. Ferrer for the nucleolin antibody; and Dr. N. Cook and Dr. N. Halbsgut for the assistance with confocal microscopy and image preparation, respectively. This research was supported by the European Community's Seventh Framework Programme FP7/2009 under Grant Agreement 238316. This study makes use of results or expertise provided by BioExcel CoE (www.bioexcel.eu), a project funded by the European Union Contract H2020-EINFRA-2015-1-675728. A.V.-P. was funded by the Dorothea Schlözer Programme of the Georg August University Göttingen; T.L. d.F. was funded by Fundação para a Ciência e Tecnologia (SFRH/BD/74881/2010); R.S. was funded by Comissao de Aperfeiçoamento de Pessoal de Nival Superior–Science without Borders and CNPq; R.P. was funded by Fundação para a Ciência e Tecnologia (SFRH/BD/80884/2011); É.M.S. was funded by the Dorothea Schlözer Programme of the Georg August University Göttingen; E.A. G. was funded by the Deutsche Forschungsgemeinschaft (DFG) grant "Copper binding to the physiological form of the α-synuclein protein"; I.M. was funded by an Emmy Noether/DFG grant; S.V. is supported by the Catalan Institution for Research and Advanced Studies-Academia program; and T.F.O. is supported by the DFG Center for Nanoscale Microscopy and Molecular Physiology of the Brain.

1. McCann H, Stevens CH, Cartwright H, Halliday GM (2014) α-Synucleinopathy phenotypes. *Parkinsonism Relat Disord* 20(Suppl 1):S62–S67.
2. Pasanen P, et al. (2014) Novel α-synuclein mutation A53E associated with atypical multiple system atrophy and Parkinson's disease-type pathology. *Neurobiol Aging* 35(9):2180.e1–2180.e5.
3. Xu W, Tan L, Yu J-T (2015) Link between the SNCA gene and parkinsonism. *Neurobiol Aging* 36(3):1505–1518.
4. Maroteaux L, Campanelli JT, Scheller RH (1988) Synuclein: A neuron-specific protein localized to the nucleus and presynaptic nerve terminal. *J Neurosci* 8(8):2804–2815.
5. Breydo L, Wu JW, Uversky VN (2012) Alpha-synuclein misfolding and Parkinson's disease. *Biochim Biophys Acta* 1822(2):261–285.
6. Burré J, Sharma M, Südhof TC (2014) α-Synuclein assembles into higher-order multimers upon membrane binding to promote SNARE complex formation. *Proc Natl Acad Sci USA* 111(40):E4274–E4283.
7. Spillantini MG, Crowther RA, Jakes R, Hasegawa M, Goedert M (1998) Alpha-synuclein in filamentous inclusions of Lewy bodies from Parkinson's disease and dementia with Lewy bodies. *Proc Natl Acad Sci USA* 95(11):6469–6473.
8. Fujiwara H, et al. (2002) Alpha-synuclein is phosphorylated in synucleinopathy lesions. *Nat Cell Biol* 4(2):160–164.
9. Gómez-Tortosa E, Irizarry MC, Gómez-Isla T, Hyman BT (2000) Clinical and neuropathological correlates of dementia with Lewy bodies. *Ann N Y Acad Sci* 920:9–15.
10. Dickson DW (2012) Parkinson's disease and parkinsonism: Neuropathology. *Cold Spring Harb Perspect Med* 2(8):a009258.
11. Cookson MR, van der Brug M (2008) Cell systems and the toxic mechanism(s) of alpha-synuclein. *Exp Neurol* 209(1):5–11.
12. Eisbach SE, Outeiro TF (2013) Alpha-synuclein and intracellular trafficking: impact on the spreading of Parkinson's disease pathology. *J Mol Med (Berl)* 91(6):693–703.

13. Roberts HL, Brown DR (2015) Seeking a mechanism for the toxicity of oligomeric α-synuclein. *Biomolecules* 5(2):282–305.
14. Winner B, et al. (2011) In vivo demonstration that alpha-synuclein oligomers are toxic. *Proc Natl Acad Sci USA* 108(10):4194–4199.
15. Karpinar DP, et al. (2009) Pre-fibrillar alpha-synuclein variants with impaired beta-structure increase neurotoxicity in Parkinson's disease models. *EMBO J* 28(20):3256–3268.
16. Braak H, et al. (2003) Staging of brain pathology related to sporadic Parkinson's disease. *Neurobiol Aging* 24(2):197–211.
17. Li J-Y, et al. (2008) Lewy bodies in grafted neurons in subjects with Parkinson's disease suggest host-to-graft disease propagation. *Nat Med* 14(5):501–503.
18. Angot E, Brundin P (2009) Dissecting the potential molecular mechanisms underlying alpha-synuclein cell-to-cell transfer in Parkinson's disease. *Parkinsonism Relat Disord* 15(Suppl 3):S143–S147.
19. Masuda-Suzukake M, et al. (2013) Prion-like spreading of pathological α-synuclein in brain. *Brain* 136(Pt 4):1128–1138.
20. Recasens A, et al. (2014) Lewy body extracts from Parkinson disease brains trigger α-synuclein pathology and neurodegeneration in mice and monkeys. *Ann Neurol* 75(3):351–362.
21. Gorell JM, et al. (1999) Occupational exposure to manganese, copper, lead, iron, mercury and zinc and the risk of Parkinson's disease. *Neurotoxicology* 20(2-3):239–247.
22. Willis AW, et al. (2010) Metal emissions and urban incident Parkinson disease: A community health study of Medicare beneficiaries by using geographic information systems. *Am J Epidemiol* 172(12):1357–1363.
23. Santner A, Uversky VN (2010) Metalloproteomics and metal toxicology of α-synuclein. *Metallomics* 2(6):378–392.
24. Kozlowski H, et al. (2009) Copper, iron, and zinc ions homeostasis and their role in neurodegenerative disorders (metal uptake, transport, distribution and regulation). *Coord Chem Rev* 253(21-22):2665–2685.
25. Uversky VN, Li J, Fink AL (2001) Metal-triggered structural transformations, aggregation, and fibrillation of human alpha-synuclein. A possible molecular NK between Parkinson's disease and heavy metal exposure. *J Biol Chem* 276(47):44284–44296.
26. Rasia RM, et al. (2005) Structural characterization of copper(II) binding to alpha-synuclein: Insights into the bioinorganic chemistry of Parkinson's disease. *Proc Natl Acad Sci USA* 102(12):4294–4299.
27. Jarrett JT, Lansbury PT, Jr (1993) Seeding "one-dimensional crystallization" of amyloid: A pathogenic mechanism in Alzheimer's disease and scrapie? *Cell* 73(6):1055–1058.
28. Rose F, Hodak M, Bernholc J (2011) Mechanism of copper(II)-induced misfolding of Parkinson's disease protein. *Sci Rep* 1:11.
29. Sóvágó I, Kállay C, Várnagy K (2012) Peptides as complexing agents: Factors influencing the structure and thermodynamic stability of peptide complexes. *Coord Chem Rev* 256(19–20):2225–2233.
30. Kowalik-Jankowska T, Rajewska A, Jankowska E, Grzonka Z (2006) Copper(II) binding by fragments of alpha-synuclein containing M1-D2- and -H50-residues; a combined potentiometric and spectroscopic study. *Dalton Trans* (42):5068–5076.
31. Rossetti G, et al. (2016) Conformational ensemble of human α-synuclein physiological form predicted by molecular simulations. *Phys Chem Chem Phys* 18(8):5702–5706.
32. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. *J Chem Phys* 81(8):3684–3690.
33. Valensin D, Dell'Acqua S, Kozlowski H, Casella L (April 11, 2016) Coordination and redox properties of copper interaction with α-synuclein. *J Inorg Biochem*, 10.1016/j.jinorgbio.2016.04.012.
34. Nosé S (1984) A molecular dynamics method for simulations in the canonical ensemble. *Mol Phys* 52(2):255–268.
35. Kosik KS, Finch EA (1987) MAP2 and tau segregate into dendritic and axonal domains after the elaboration of morphologically distinct neurites: An immunocytochemical study of cultured rat cerebrum. *J Neurosci* 7(10):3142–3153.
36. Sacino AN, et al. (2014) Brain injection of α-synuclein induces multiple proteinopathies, gliosis, and a neuronal injury marker. *J Neurosci* 34(37):12368–12378.
37. Pekny M, Pekna M (2014) Astrocyte reactivity and reactive astrogliosis: Costs and benefits. *Physiol Rev* 94(4):1077–1098.
38. Volpicelli-Daley LA, et al. (2011) Exogenous α-synuclein fibrils induce Lewy body pathology leading to synaptic dysfunction and neuron death. *Neuron* 72(1):57–71.
39. Sacino AN, et al. (2013) Conformational templating of α-synuclein aggregates in neuronal-glial cultures. *Mol Neurodegener* 8:17.
40. Uryu K, et al. (2006) Convergence of heat shock protein 90 with ubiquitin in filamentous alpha-synuclein inclusions of alpha-synucleinopathies. *Am J Pathol* 168(3):947–961.
41. Hasegawa M, et al. (2002) Phosphorylated alpha-synuclein is ubiquitinated in alpha-synucleinopathy lesions. *J Biol Chem* 277(50):49071–49076.
42. De Ricco R, et al. (2015) Remote His50 acts as a coordination switch in the high-affinity N-terminal centered copper(II) site of α-synuclein. *Inorg Chem* 54(10):4744–4751.
43. Wang X, Moualla D, Wright JA, Brown DR (2010) Copper binding regulates intracellular alpha-synuclein localisation, aggregation and toxicity. *J Neurochem* 113(3):704–714.
44. Xiang W, et al. (2015) Posttranslational modification and mutation of histidine 50 trigger alpha synuclein aggregation and toxicity. *Mol Neurodegener* 10(1):8.
45. Carboni E, Lingor P (2015) Insights on the interaction of alpha-synuclein and metals in the pathophysiology of Parkinson's disease. *Metallomics* 7(3):395–404.
46. Qin Z, Hu D, Han S, Hong D-P, Fink AL (2007) Role of different regions of alpha-synuclein in the assembly of fibrils. *Biochemistry* 46(46):13322–13330.
47. Binolfi A, et al. (2011) Exploring the structural details of Cu(I) binding to α-synuclein by NMR spectroscopy. *J Am Chem Soc* 133(2):194–196.
48. Danzer KM, Krebs SK, Wolff M, Birk G, Hengerer B (2009) Seeding induced by alpha-synuclein oligomers provides evidence for spreading of alpha-synuclein pathology. *J Neurochem* 111(1):192–203.
49. Luk KC, et al. (2009) Exogenous alpha-synuclein fibrils seed the formation of Lewy body-like intracellular inclusions in cultured cells. *Proc Natl Acad Sci USA* 106(47):20051–20056.
50. Hansen C, et al. (2011) α-Synuclein propagates from mouse brain to grafted dopaminergic neurons and seeds aggregation in cultured human cells. *J Clin Invest* 121(2):715–725.
51. Volpicelli-Daley LA, Luk KC, Lee VM-Y (2014) Addition of exogenous α-synuclein preformed fibrils to primary neuronal cultures to seed recruitment of endogenous α-synuclein to Lewy body and Lewy neurite-like aggregates. *Nat Protoc* 9(9):2135–2146.
52. Danzer KM, et al. (2007) Different species of alpha-synuclein oligomers induce calcium influx and seeding. *J Neurosci* 27(34):9220–9232.
53. Bousset L, et al. (2013) Structural and functional characterization of two alpha-synuclein strains. *Nat Commun* 4:2575.
54. Aulić S, et al. (2014) Defined α-synuclein prion-like molecular assemblies spreading in cell culture. *BMC Neurosci* 15:69.
55. Lee HJ, et al. (2010) Direct transfer of alpha-synuclein from neuron to astroglia causes inflammatory responses in synucleinopathies. *J Biol Chem* 285(12):9262–9272.
56. Schell H, Hasegawa T, Neumann M, Kahle PJ (2009) Nuclear and neuritic distribution of serine-129 phosphorylated α-synuclein in transgenic mice. *Neuroscience* 160(4):796–804.
57. Osterberg VR, et al. (2015) Progressive aggregation of alpha-synuclein and selective degeneration of Lewy inclusion-bearing neurons in a mouse model of parkinsonism. *Cell Reports* 10(8):1252–1260.
58. Peelaerts W, et al. (2015) α-Synuclein strains cause distinct synucleinopathies after local and systemic administration. *Nature* 522(7556):340–344.
59. Zhang S, et al. (2015) LK6/Mnk2a is a new kinase of alpha synuclein phosphorylation mediating neurodegeneration. *Sci Rep* 5:12564.
60. Outeiro TF, et al. (2008) Formation of toxic oligomeric alpha-synuclein species in living cells. *PLoS One* 3(4):e1867.
61. Ulusoy A, et al. (2015) Neuron-to-neuron α-synuclein propagation in vivo is independent of neuronal injury. *Acta Neuropathol Commun* 3(1):13.
62. Tanaka M, et al. (2004) Aggresomes formed by alpha-synuclein and synphilin-1 are cytoprotective. *J Biol Chem* 279(6):4625–4631.
63. Paine MG, Babu JR, Seibenhener ML, Wooten MW (2005) Evidence for p62 aggregate formation: Role in cell survival. *FEBS Lett* 579(22):5029–5034.
64. Bodner RA, et al. (2006) Pharmacological promotion of inclusion formation: A therapeutic approach for Huntington's and Parkinson's diseases. *Proc Natl Acad Sci USA* 103(11):4246–4251.
65. Smith WW, et al. (2010) Synphilin-1 attenuates neuronal degeneration in the A53T alpha-synuclein transgenic mouse model. *Hum Mol Genet* 19(11):2087–2098.
66. Fox LM, et al. (2011) Soluble tau species, not neurofibrillary aggregates, disrupt neural system integration in a tau transgenic model. *J Neuropathol Exp Neurol* 70(7):588–595.
67. Arrasate M, Mitra S, Schweitzer ES, Segal MR, Finkbeiner S (2004) Inclusion body formation reduces levels of mutant huntingtin and the risk of neuronal death. *Nature* 431(7010):805–810.
68. Saudou F, Finkbeiner S, Devys D, Greenberg ME (1998) Huntingtin acts in the nucleus to induce apoptosis but death does not correlate with the formation of intranuclear inclusions. *Cell* 95(1):55–66.
69. Sieradzan KA, et al. (1999) Huntington's disease intranuclear inclusions contain truncated, ubiquitinated huntingtin protein. *Exp Neurol* 156(1):92–99.
70. Butterfield SM, Lashuel HA (2010) Amyloidogenic protein-membrane interactions: Mechanistic insight from model systems. *Angew Chem Int Ed Engl* 49(33):5628–5654.
71. Volles MJ, Lansbury PT, Jr (2007) Relationships between the sequence of alpha-synuclein and its membrane affinity, fibrillization propensity, and yeast toxicity. *J Mol Biol* 366(5):1510–1522.
72. Chutna O, et al. (2014) The small GTPase Rab11 co-localizes with α-synuclein in intracellular inclusions and modulates its aggregation, secretion and toxicity. *Hum Mol Genet* 23(25):6732–6745.
73. Giehm L, Lorenzen N, Otzen DE (2011) Assays for α-synuclein aggregation. *Methods* 53(3):295–305.
74. Tönges L, et al. (2014) Alpha-synuclein mutations impair axonal regeneration in models of Parkinson's disease. *Front Aging Neurosci* 6:239.
75. Susi H, Byler DM (1983) Protein structure by Fourier transform infrared spectroscopy: Second derivative spectra. *Biochem Biophys Res Commun* 115(1):391–397.

Lázaro *et al. Acta Neuropathologica Communications* (2016) 4:128
DOI 10.1186/s40478-016-0402-8

Acta Neuropathologica Communications

RESEARCH Open Access

The effects of the novel A53E alpha-synuclein mutation on its oligomerization and aggregation

Diana F. Lázaro[1†], Mariana Castro Dias[1†], Anita Carija[2], Susanna Navarro[2], Carolina Silva Madaleno[3], Sandra Tenreiro[3], Salvador Ventura[2] and Tiago F. Outeiro[1,3,4*]

Abstract

α-synuclein (aSyn) is associated with both sporadic and familial forms of Parkinson's disease (PD), the second most common neurodegenerative disorder after Alzheimer's disease. In particular, multiplications and point mutations in the gene encoding for aSyn cause familial forms of PD. Moreover, the accumulation of aSyn in Lewy Bodies and Lewy neurites in disorders such as PD, dementia with Lewy bodies, or multiple system atrophy, suggests aSyn misfolding and aggregation plays an important role in these disorders, collectively known as synucleinopathies. The exact function of aSyn remains unclear, but it is known to be associated with vesicles and membranes, and to have an impact on important cellular functions such as intracellular trafficking and protein degradation systems, leading to cellular pathologies that can be readily studied in cell-based models. Thus, understanding the molecular effects of aSyn point mutations may provide important insight into the molecular mechanisms underlying disease onset.
We investigated the effect of the recently identified A53E aSyn mutation. Combining in vitro studies with studies in cell models, we found that this mutation reduces aSyn aggregation and increases proteasome activity, altering normal proteostasis.
We observed that, in our experimental paradigms, the A53E mutation affects specific steps of the aggregation process of aSyn and different cellular processes, providing novel ideas about the molecular mechanisms involved in synucleinopathies.

Keywords: Alpha-synuclein, Parkinson's disease, Oligomerization, Aggregation, Neurodegeneration

Introduction

Parkinson's disease (PD) is a highly debilitating and progressive neurodegenerative disorder affecting around seven million people worldwide. PD is typically known as a movement disorder, due to the characteristic motor manifestations associated with the loss of dopaminergic neurons from the *substantia nigra*, although it also affects other areas of the brain. PD and other neurodegenerative disorders, such as demential with Lewy bodies, and multiple system atrophy, are also characterized by the accumulation of aggregated alpha-synuclein (aSyn) in proteinaceous inclusions known as Lewy bodies (LBs) or Lewy neurites [54]. Together, these diseases are known as synucleinopathies [17, 55]. However, it is still unclear whether LBs are themselves toxic or protective [7, 43], with smaller oligomeric species of aSyn being the culprits as recent studies suggest [31, 45, 58, 63]. aSyn is a disordered and abundant neuronal protein whose normal function is still elusive. Familial forms of PD associated with duplication and triplication of the *SNCA* gene [53], along with studies of aSyn overexpression, in cellular and animal models, suggest the protein may acquire a toxic function. The cellular pathologies associated with increased levels and accumulation of aSyn include disruption of vesicular transport [6, 42], mitochondrial dysfunction, impairment of autophagy and proteasome, and oxidative stress [2, 21], suggesting aSyn plays a multitude of roles in the cell, perhaps due

---

* Correspondence: touteir@gwdg.de
†Equal contributors
[1]Department of Neurodegeneration and Restorative Research, University Medical Center Göttingen, Waldweg 33, 37073 Göttingen, Germany
[3]Chronic Disease Research Center (CEDOC), NOVA Medical School, Campo dos Mártires da Pátria, 130, 1169-056 Lisbon, Portugal
Full list of author information is available at the end of the article

to its intrinsically disordered nature. Under physiological conditions, aSyn is considered to be a pre-synaptic protein [37] that associates with vesicles and membranes [11].

According to the "Braak hypothesis", PD pathology is thought to start from the periphery (gut or nose), and progress until it reaches the brain [4, 5, 49], spreading in a prion-like manner [20, 29, 36]. However, this hypothesis is still controversial, and the molecular mechanisms underlying this phenomenon are not fully understood [25].

The vast majority of PD cases are sporadic but single point mutations in the gene encoding for aSyn (*SNCA*) cause familial forms of the disease [10]. The most recently identified aSyn mutation causes the substitution of alanine at position 53 by a glutamate residue (A53E), identified in a 36 year-old Finnish patient with atypical PD. The patient displayed a dense accumulation of *SNCA* inclusions in the striatum and a severe cortical pathology, affecting both the superficial and deep laminae [3]. In vitro, the A53E mutation was shown to reduce aSyn aggregation and fibril formation without changing the secondary structure content of the protein, when compared to WT aSyn [15]. These data suggest that the negatively charged glutamate residue may affect the folding and, consequently, the aggregation process of the protein.

In our study, we conducted a detailed study of the effects of the A53E mutation on aSyn using a combination of in vitro and cellular models of aSyn oligomerization and aggregation [40, 45]. Our results showed that the A53E mutation modulates aSyn aggregation in vitro and in vivo and impacts on distinct cellular pathways.

Altogether, the study of specific aSyn mutants provides novel insight into the spectrum of functions and cellular pathologies associated, opening novel avenues for the design of therapeutic strategies for PD and other synucleinopathies.

Materials and methods

Protein expression and purification

pET21a vectors (Novagen) encoding for WT aSyn and the A53E mutant were transformed into *E. coli* BL21 (DE3) cells. For protein expression, 10 ml overnight culture of transformed cells was used to inoculate 1 L of LB medium with 100 μg/mL ampicillin, which was further incubated at 37 °C and 250 rpm. At an OD_{600} 0.6, protein expression was induced with 1 mM of isopropyl-1-thio-β-D-galactopyranoside (IPTG) for 4 h at 37 °C. Afterwards, the cultures were centrifuged and the cell pellet frozen at −80 °C.

For cell lysis, pellets were resuspended in 15 mL of lysis buffer (50 mM Tris · HCl, 150 mM NaCl, 1 μg/mL Pepstatin A, 20 μg/ml Aprotinin, 1 mM Benzamidine, 1 mM PMSF and 1 mM EDTA) and sonicated on ice. The lysate was boiled for 10 min at 95 °C and soluble and insoluble fractions were separated by centrifugation at 48,384 × *g*, for 15 min. 136 μL/mL of 10% streptomycin sulfate and 228 μL/mL of glacial acetic acid was added to the supernatant. The resulting solution was centrifuged for 5 min at 48,384 × *g*, and the supernatant was collected and precipitated with saturated ammonium sulphate at 4 °C in a ratio 1:1 (v/v) with the supernatant. The pellets were washed with a 1:1 (v/v) solution of ammonium sulphate and water. The pellets were resuspended in 900 μL of ammonium acetate 100 mM and the same volume of 100% ethanol was added to precipitated aSyn.

Purification protocol was as adapted from [59]. Briefly, ethanol precipitated aSyn was resuspended in starting buffer (25 mM Tris · HCl at pH 8.0) and filtered through a Millex-HP filter syringe-driven filter unit (0.45 μm, Millipore). Anion exchange high-performance liquid-chromatography was carried out on an AKTA-FPLC (GE Healthcare). The sample was loaded, bounded to Hi-Trap column (GE Healthcare) and eluted with a NaCl linear gradient of elution buffer (25 mM Tris · HCl at pH 8.0, 1 M NaCl). The fractions containing aSyn were collected and buffer was exchanged for 20 mM ammonium acetate. The identity, and purity of the recombinant proteins was assessed by Mass Spectrometry and SDS-PAGE being higher than 99%.

Aggregation assays

The protein stocks were prepared by resuspending lyophilized protein in native buffer (10 mM sodium phosphate at pH 7.0) to a concentration of approximately 100 μM. Then, the protein stocks were filtered through 0.22 μm filter unit. The integrity of the protein and the absence of soluble oligomers at the beginning of the reaction was confirmed by gel filtration chromatography in a Superdex 75 10/300 column (GE Healthcare Life Sciences). Three aliquots of 300 μL of aSyn WT and A53E mutant were prepared from the protein stocks, in native buffer to a final concentration of 60 μM. Samples were incubated in an Eppendorf Thermomixer Comfort (Eppendorf, USA) with 0.02% sodium azide at 600 rpm and 37 °C.

Light scattering spectroscopy

The transition of aSyn from initial soluble monomeric form to aggregated state was determined by measuring light scattering in a Jasco FP-8200 spectrofluorometer (Jasco Inc, MD, USA) with an excitation wavelength of 330 nm and emission range from 320 to 340 nm at 25 °C. Final protein concentration was 10 μM in native buffer. Solutions without protein were used as negative controls. All experiments were carried out in triplicates.

Thioflavin T binding assay

Th-T binding to amyloid fibrils was recorded using a JASCO FP-8200 spectrofluorometer (Jasco Inc, MD, USA) with an excitation wavelength of 445 nm and emission range from 460 to 600 nm at 25 °C, using a slit width of 5 nm for excitation and emission. The final concentration of Th-T was 25 μM and final protein concentration was 10 μM in native buffer. Solutions without protein were used as negative controls. All experiments were carried out in triplicates.

Congo red binding assay

CR binding to amyloid fibrils was tested using a Cary-400 Varian spectrophotometer (Varian Inc., Palo Alto, CA, USA) by recording the absorbance spectra from 375 to 675 nm, at 25 °C. A final concentration of 10 μM CR was added to 10 μM protein samples in native buffer. Protein solutions in the presence and absence of CR were used to calculate the differential CR spectra. All experiments were carried out in triplicates.

Transmission electron microscopy (TEM) assays

For negative staining, incubated samples were diluted in Mili-Q water to 10 μM and 10 μL were then placed on carbon-coated copper grids, and left to stand for 5 min. The grids were washed with distilled water and stained with 2% (w/v) uranyl acetate for 1 min. The morphology of aggregates of WT and A53E aSyn was observed using a JEOL JEM 1400 transmission electron microscope (JEOL, USA) at an accelerating voltage of 120 kV. The width of fibrils for WT and necklace-like structures for A53E was measured using ImageJ (250 measurements). Results are provided as Mean ± SE.

ATR-FTIR spectroscopy

Attenuated total reflectance Fourier transform infrared spectroscopy (ATR FT-IR) analysis of amyloid fibrils was performed using a Bruker Tensor 27 FTIR Spectrometer (Bruker Optics Inc.) with a Golden Gate MKII ATR accessory. Incubated samples were centrifuged and the insoluble fraction was resuspended in water. Each spectrum consists of 16 independent scans, measured at a spectral resolution of 4 cm^{-1} within the 1800–1500 cm^{-1} range. Second derivatives of the spectra were used to determine the frequencies at which the different spectral components were located. FTIR spectra were fitted to overlapping Gaussian curves using PeakFit package software (Systat Software).

Aggregation kinetics

Aggregation of aSyn, departing from soluble monomeric form, was monitored by measuring the transition from non-aggregated to aggregated state according to the Th-T fluorescence at 486 nm on a 96-wells microplate reader for 72 h at 37 °C (Victor Microplate reader, Perkin Elmer, USA). The reactions were carried out with 70 μM soluble purified WT and A53E aSyn in native buffer. Experiments were carried out in triplicates.

Sedimentation assay

aSyn aggregation was measured by sedimentation using centrifugation. The incubated samples of WT and A53E mutant were centrifuged at 48,384 × g for 30 min, and the supernatants were carefully removed. The amount of soluble aSyn was measured by absorbance at 280 nm using aSyn extinction coefficient $\mathcal{E}_{280} = 5960\ M^{-1}/cm^{-1}$, before and after centrifugation. All measurements were carried out in triplicates.

Primer design

The primers were designed according with the manufacturer's instructions (Table 1).

Generation of A53E aSyn constructs for expression in mammalian cells

A53E was inserted in the Venus-BiFC system [45] or SynT [40] by site-directed mutagenesis (QuickChange II Site-Directed Mutagenesis Kit, Agilent Technologies, SC, USA) following the manufacturer's instructions. All constructions were confirmed by sequencing.

Cell culture

Human Embryonic Kidney 293 (HEK) cells were grown in Dulbecco's Modified Eagle Medium (DMEM, Life Technologies- Invitrogen, Carlsbad, CA, USA), and human neuroglioma cells (H4) in Opti-MEM I with Glutamax (Life Technologies- Gibco, Carlsbad, CA, USA), both supplemented with 10% Fetal Bovine Serum Gold (PAA, Cölbe, Germany) and 1% Penicillin-Streptomycin (PAN, Aidenbach, Germany). Cells were grown at 37 °C, with 5% of CO2.

Cell transfection

HEK cells

The day before transfection, 100 000 cells were plated in 12-well plates (Costar, Corning, New York, USA). The cells were transfected with equimolar amounts of the plasmids using Metafectene (Biotex, Munich, Germany) as specified by the manufacturer. After twenty-four hours, the cells were collected or stained for further analysis.

Table 1 Primers used to perform site-directed mutagenesis and generate the A53E mutant

A53E forward	5′ GAGTGGTGCATGGTGTGGAAACAGTGGCTGAGAAGAC 3′
A53E reverse	5′ GTCTTCTCAGCCACTGTTTCCACACCATGCACCACTC 3′

H4 cells

Eighty thousand cells were plated in 12-well plates (Costar, Corning, New York, USA). After 24 h, equal amount of SynT and Synphilin-1 were transfected using FuGENE6 Transfection Reagent (Promega, Madison, USA) in a ratio of 1:3 according to the manufacturer's recommendation. Forty-eight hours after transfection, the cells were processed for different assays.

Immunocytochemistry

After transfection, cells were fixed with 4% paraformaldehyde at room temperature (RT), followed by a permeabilization with 0.5% Triton X-100 (SigmaAldrich, St. Louis, MO, USA). The cells were blocked in 1.5% normal goat serum (PAA, Cölbe, Germany)/1xPBS (1.37 M NaCl, 27 mM KCl, 101.4 mM $Na_2HPO_47 \cdot H_2O$, 16.7 mM KH_2PO_4), and then incubated with primary antibody. Primary antibodies used in this study were: mouse Syn1 (1:1000, BD Transduction Laboratory, New Jersey, USA) or rabbit anti-aSyn (1:1000, Abcam, Boston, USA), anti-Giantin (1:1000, Abcam, Boston, USA), aSyn-S129 1:1000 (Wako Chemicals USA, Inc., Richmond, USA) overnight, and secondary antibody (Alexa Fluor 488 donkey anti-mouse IgG and/or Alexa Fluor 555 goat anti rabbit IgG, (Life Technologies- Invitrogen, Carlsbad, CA, USA)) for 2 h at RT. Cells were finally stained with Hoechst 33258 (Life Technologies- Invitrogen, Carlsbad, CA, USA) (1:5000 in DPBS) for 5 min, and maintained in 1xPBS for imaging.

Yeast transformation and plasmids

The p426GAL-aSyn-GFP plasmid carries the human gene of aSyn with a C-terminal fusion to GFP, under the regulation of *GAL1* inducible promoter [44]. This plasmid was used to generate aSyn A53E by site directed mutagenesis. The yeast cells W303-1A (*MAT***a**; *can1-100*; *his3-11,15*; *leu2-3,112*; *trp1-1*; *ura3-1*; *ade2-1*) were transformed with the indicated plasmids using lithium acetate standard method [16].

Yeast growth

For all experiments, an inoculum was prepared to obtain cells in *log* growth phase, using synthetic complete (SC) medium [0.67% (w/v) yeast nitrogen base without amino acids (Difco), 1% (w/v) raffinose and 0.79 $g.L^{-1}$ complete supplement mixture (CSM) (QBiogene)], 200 rpm, 30 °C, as we described [57]. To induce aSyn expression, in liquid medium, cells were grown ($OD_{600\ nm}$ 0.2) in SC selective medium 1% (w/v) galactose (aSyn ON) for 7 h at 30 °C, 200 rpm. aSyn cytotoxicity was evaluated by spotting assays. $OD_{600\ nm}$ was set to 0.1 ± 0.005 and 1:10 serially dilutions of each sample were prepared [57]. Then, 4 μL of each dilution was spotted in solid SC selective medium containing 2% glucose (aSyn OFF) or 1% (w/v) galactose (aSyn ON) and incubated at 30 °C for 36–42 h. Images were acquired using ChemiDoc Touch (Bio-Rad).

Fluorescence microscopy

HEK cells expressing the aSyn Venus-BiFC assay were visualized using the Olympus IX81-ZDC microscope system, with a 20× objective. One hundred images were randomly taken out of four independent experiments. Total intensity was measured using the Olympus Scan^R Image Analysis Software.

In order to determine the percentage of yeast cells with aSyn inclusions, cells were grown as described above and GFP fluorescence was visualized with a Zeiss Z2 Widefield Fluorescence microscope. The percentage of cells presenting aSyn inclusions was then determined by counting at least 300 cells for each treatment using ImageJ software.

Quantification of aSyn inclusions

Cells expressing aSyn were scored according with the presence or absence of inclusion on transfected cells. Three independent experiments were performed, and the results were expressed as the percentage of the total number of transfected cells.

Thioflavin S staining

Freshly prepared, 0.5% of Thio-S (Sigma-Aldrich, St. Louis, MO, USA) was incubated within the cells for 5 min. The cells were washed three times with 80% ethanol, and maintained in 1xPBS for fluorescence microscopy.

Quantification of Golgi fragmentation

Golgi morphology was assessed in transfected HEK and H4, and scored into three groups, was we previously published [32] (normal, diffused and fragmented). Three independent experiments were performed.

Western blot analysis

HEK and H4 cells were lysed with Radio-Immunoprecipitation Assay (RIPA) lysis buffer (50 mM Tris pH 8.0, 0.15 M NaCl, 0.1% SDS, 1% NP40, 0.5% Na-Deoxycholate), 2 mM EDTA and a Protease Inhibitor Cocktail (1 tablet/10 mL) (Roche Diagnostics, Mannheim, Germany). Protein concentration was determined by Bradford assay (BioRad Laboratories, Hercules, CA, USA), and the sample were denaturation for 5 min at 100 °C in protein sample buffer (125 mM of 1 M Tris HCl pH 6.8, 4% SDS 0,5% Bromophenol blue, 4 mM EDTA, 20% Glycerol, 10% b-Mercapto ethanol). The gels were loaded with 80 μg protein, and the samples separated on 12% SDS-polyacrylamide gels. The gel was transferred to a PVDF membrane using a Trans-Blot Turbo transfer system (BioRad), according to the manufacturer's instructions. Membranes were blocked with 5% (w/v) skim milk (Fluka, Sigma-Aldrich, St. Louis, MO, USA), and incubated with Syn1 (1:1000, BD Biosciences, San Jose, CA, USA), and 1:2000 anti-b-actin (Sigma-Aldrich, St. Louis, MO, USA) overnight at 4 °C. After washing, the membranes were incubated for 1 h with secondary antibody, anti-mouse IgG, or anti-rabbit IgG, horseradish peroxidase labeled secondary antibody (GE Healthcare, Bucks, UK) at 1:10,000. Proteins were detected by ECL chemiluminescent detection system (Millipore, Billerica, MA, USA) in Fusion FX (Vilber Lourmat). The band intensity was estimated using the ImageJ software (NIH, Bethesda, MD, USA) and normalized against b-actin.

For aSyn quantification total yeast protein extraction and western blot was performed following standard procedures as described before [57]. Antibodies used: aSyn (BD Transduction Laboratories, San Jose, CA, USA), pS129-aSyn (Wako Chemicals USA, Inc., Richmond VA, USA) PGK (Life Technologies, PaisleyUK). Triton soluble and insoluble fractions were processed and analyzed as described before [56].

Native PAGE

For native PAGE, HEK cells were lysed in 1xPBS pH 7.4 with Protease Inhibitor Cocktail tablet and separated in 4–16% gradient Native pre-cast gel (SERVA Electrophoresis GmbH, Heidelberg, Germany). Gels were run according to the manufacturer's instructions, and transferred as previously described.

Proteinase K digestion

H4 cells samples were digested with Proteinase K (2.5 μg/mL) (Roth, Carlsbad, Germany) for 1, 3, and 5 min at 37 °C. The enzyme reaction was stopped with protein sample buffer, and the samples were separated in a SDS-page gel, as described above.

Flow cytometry (FCM)

FCM was performed in a BD FACSCanto II. To analyze cell viability yeast cells transformed with the indicated plasmids were incubated with 5 $\mu g.mL^{-1}$ PI, for 15 min at 30 °C, 200 rpm and protected from light. Cells were then washed with PBS and used to FCM. A minimum of 10,000 events were collected for each experiment. Results were expressed as median fluorescence intensity (MFI) of a molecule.

Measurement of 26S Proteasome Catalytic Activity

The chymotrypsin-like activity of the 26S proteasome was determined has previously described [27]. Briefly, after 48 h transfected cells were collected in lysis buffer (50 mM Tris, pH 7.5, 250 mM Sucrose, 5 mM MgCl2, 1 mM DTT, 0.5 mM EDTA, 0.025% Digitonin, 2 mM ATP). The reaction was initiated with 15 μg from the total protein lysates, together with the addition of reaction buffer (50 mM Tris (pH 7.5), 40 mM KCl2, 5 mM MgCl2, 1 mM DTT, 0.5 mM ATP, 100 μM Suc-LLVYAMC) were mix in 100 μl final volume. The fluorescence of AMC (380 nm ex-Citation and 460 nm emission) was monitored in a microplate fluorometer (Infinite M1000, Tecan) at 37 °C. As control, proteasome was inhibited with 20 μM MG132 (Sigma, Hamburg, Germany) prior to the measurements.

Statistical analyses

Data were analyzed using GraphPad Prism 6 (San Diego California, USA) software and were expressed as the mean + − SD. Statistical differences from WT aSyn were calculated using unpaired Student *t*-test and one-way ANOVA with post-hoc Tukey's test. Significance was assessed for, where * corresponds to $p < 0.05$, ** corresponds to $p < 0.01$ and *** corresponds to $p < 0.001$.

Results

A53E mutant forms protofibrils by reducing aSyn fibrilization

To understand the effect of the A53E substitution in the aggregation process of aSyn (Fig. 1a), we assembled a pipeline of both in in vitro and cellular studies, in human and yeast cells (Fig. 1b), and conducted a battery of experiments to characterize the behavior of this recently identified aSyn mutant form.

We started to compare the in vitro aggregation properties of WT and A53E aSyn using synchronous light scattering, sedimentation and binding to amyloid-binding dyes. For these assays, the soluble forms of both proteins were incubated at 60 μM under agitation at 37 °C for two weeks. The light scattering signal was 6-fold higher for WT aSyn than for the A53E mutant, suggesting differences in aggregation process (Fig. 2a). Using spectrophotometry, we quantified the levels of aggregated protein in both solutions after separating aSyn in two fractions (soluble and insoluble), by centrifugation. The amount of protein in the insoluble fraction in the WT aSyn preparation was two times higher than with A53E (*$p < 0.05$, Fig. 2b).

The presence of amyloid fibrils can be detected in vitro using Thioflavin T (Th-T), a dye that specifically binds to amyloid fibrils [35, 51]. In agreement with light scattering and sedimentation data, we found that the Th-T fluorescence signal was 10 times lower for A53E than for WT aSyn (Fig. 2c). The presence of amyloid fibrils can be further detected by monitoring the increase of the absorbance of Congo Red (CR) and the red shift of the dye absorbance maximum [28]. The binding of A53E to CR was negligible, since the peak of absorbance in the presence of A53E aSyn was similar to that of free CR (Fig. 2d). In contrast, we observed a dramatic spectral change in CR for WT aSyn (Fig. 2d). To confirm the different amyloidogenic propensities of WT and A53E aSyn, we analyzed the morphological features of the aggregates formed using

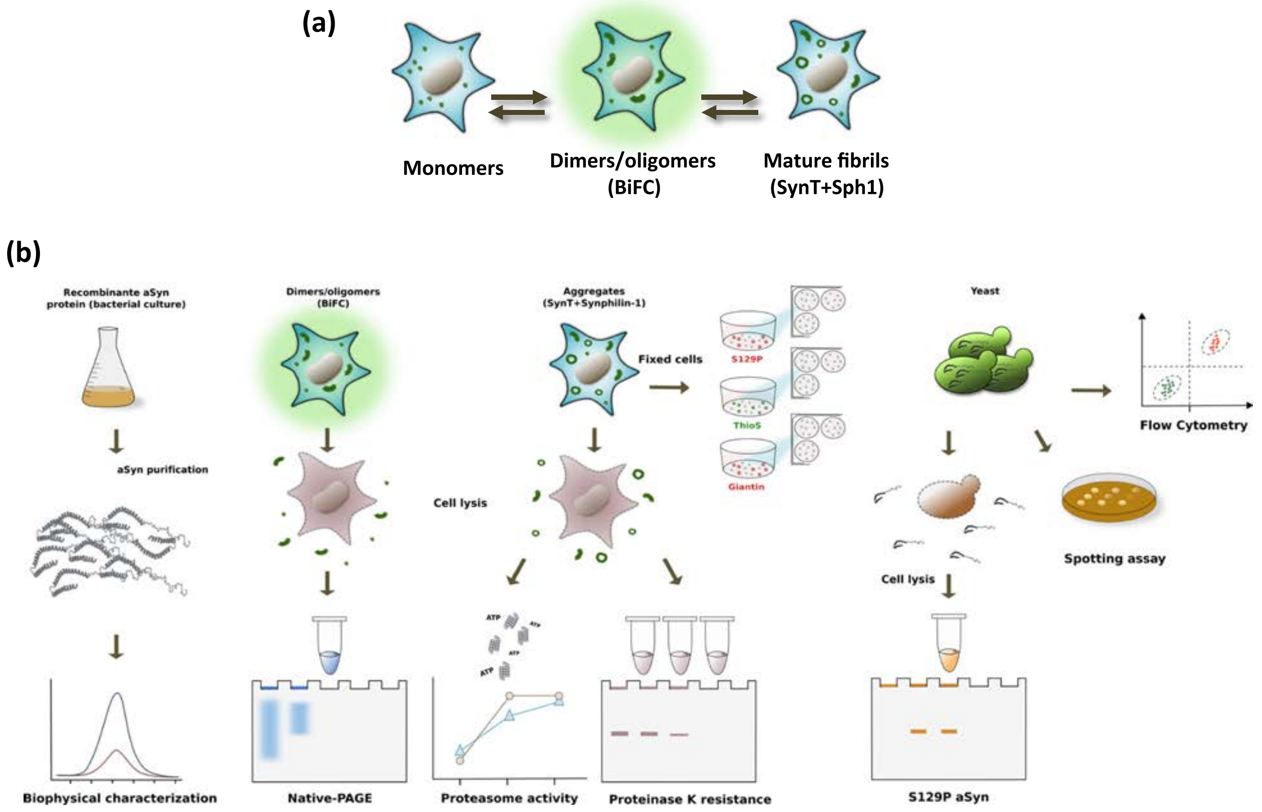


Fig. 1 Aggregation process and study design. **a** aSyn aggregation process in cell models. **b** Experimental design used in the study. In vitro studies and studies in cell models were used to assess the effect of the A53E mutation on aSyn

transmission electron microscopy (TEM). Although we detected the presence of higher order complexes in both preparations, their size and morphology was different. For WT aSyn, we observed the typical long and unbranched amyloid fibrils (Fig. 2e), 11.6 ± 0.4 nm with a width of 11.6 ± 0.4 nm (Fig. 2e). In contrast, the structures formed by A53E aSyn exhibited a protofibrillar appearance, with small round oligomeric structures that seemed to be linked in a necklace fashion, with a width of 28.5 ± 0.7 nm (Fig. 2f-g).

To assess the secondary structure content of the assemblies formed by WT and A53E aSyn, we analyzed the amide I region of the FTIR spectrum (1700–1600 cm^{-1}). This region of the spectrum corresponds to the absorption of the carbonyl peptide bond of the main amino acid chain of the protein, and is a sensitive marker of the protein secondary structure. After deconvolution of the FTIR spectra of the aSyn solutions, we were able to assign the individual secondary structure elements and their relative contribution to the main absorbance signal at the end of the aggregation reaction (Fig. 2g and h and Table 2). The absorbance spectra were radically different for WT and A53E aSyn. While the spectrum of WT aSyn was dominated by a peak at 1625 cm^{-1}, attributable to the presence of amyloid-like inter-molecular β–sheet structure (Fig. 2g), the spectrum of the A53E mutant was dominated by a peak at 1649 cm^{-1} corresponding to disordered/random coil conformation (Fig. 2h).

Next, we monitored how the mutation impacted on the aggregation kinetics of aSyn by continuously monitoring the changes in Th-T binding over time for WT and A53E variants. The kinetics of amyloid fibril formation usually follows a sigmoidal curve that reflects a nucleation-dependent growth mechanism. The aggregation of both proteins followed this pattern, with an apparent lag phase of 8 h (Fig. 2i). After this lag phase, the two aggregation reactions diverged significantly, with an exponential increase for WT aSyn that plateaued at around 55 h, and a steady and much slower increase for A53E aSyn, reaching a 3.5-times lower fluorescence intensity. Altogether, our data demonstrates that the A53E mutation reduces aSyn amyloid formation in vitro.

The A53E mutation decreases aSyn oligomerization in cellular models

We next investigated the effects of the A53E mutation on the behavior of aSyn in the context of living human cell models. First, we used the Bimolecular Fluorescence

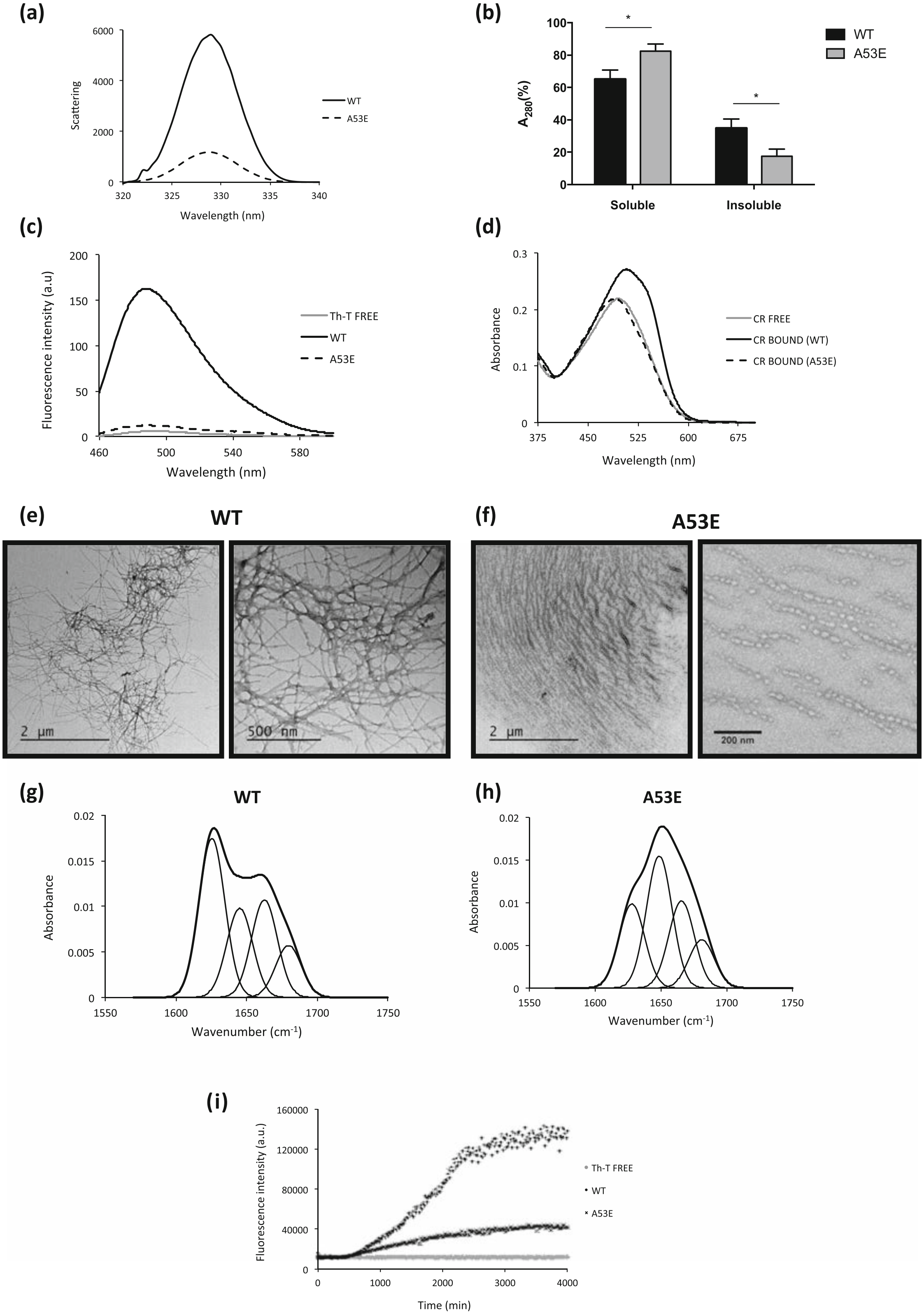


Fig. 2 (See legend on next page.)

(See figure on previous page.)
Fig. 2 Aggregation properties of WT and A53E aSyn variants. aSyn WT and A53E mutant, prepared at 60 μM in 10 mM sodium phosphate, pH 7.0, were incubated for 2 weeks under agitation at 37 °C. **a** Static light scattering of 10 μM aSyn in 10 mM sodium phosphate, WT (*solid line*) and A53E mutant (*dashed line*). **b** Distribution of aSyn between the soluble and insoluble fractions. **c** Fluorescence emission spectra of Th-T upon incubation with 10 μM aSyn WT (*solid line*) and A53E mutant (*dashed line*). Free Th-T emission spectrum is represented in grey. **d** CR absorbance spectra in the presence of 10 μM aSyn WT (solid line) and A53E mutant (*dashed line*). Free CR absorbance spectrum is represented in grey. **f**-**g** Morphology of WT and A53E aSyn aggregates TEM micrographs. Negatively stained aggregates formed by aSyn WT (left panel) and A53E mutant (right panel) incubated for two weeks. **h**-**i** Secondary structure of WT and A53E aSyn aggregates. Secondary structure content of the aSyn WT and A53E mutant after two weeks incubation. ATR-FTIR absorbance spectra in the amide I region was acquired (thick line) and the fitted individual bands after Gaussian deconvolution are shown (*thin lines*). **i** Aggregation kinetics of WT and A53E aSyn. Aggregation kinetics of aSyn were monitored by following the changes in relative ThioT fluorescence emission. Concentration of protein was 70 μM WT aSyn (crosses) and A53E mutant (*dots*) in a final volume of 150 μL. The evolution of Th-T fluorescence in the absence of protein is represented in *grey*, $n = 3$

Complementation (BiFC) assay to monitor aSyn oligomerization, as we previously described [45]. Briefly, non-fluorescence Venus fragments are fused to either the N- or C-terminus of aSyn and, upon dimerization/oligomerization of the protein, the fluorophore is reconstituted resulting in fluorescence signal. While this assay involves the tagging of aSyn with fragments of fluorescent proteins, it constitutes a powerful paradigm to assess aSyn oligomerization. We observed that A53E aSyn, similarly to WT, formed dimers/oligomers (Fig. 3a). However, the fluorescence signal was lower than that observed with WT aSyn, suggesting differences in the dimerization/oligomerization process ($^{**}p < 0.01$) (Fig. 3b), since the levels of expression of WT and A53E aSyn were identical (Fig. 3c and d, Additional file 1: Figure S2.1 and Additional file 1: Figure S2.2). We also found that the A53E mutant produced high molecular weight species, similar to WT aSyn (Fig. 3e). Thus, we refer to the species formed as oligomers, for simplicity.

The A53E mutation alters the biochemical properties of aSyn inclusions

Next, we investigated if the later stages of the aSyn aggregation process were altered by the A53E mutation. For that, we used a well-established cell-based aggregation model that consists in the co-expression of SynT (C-terminally modified aSyn) and Synphilin-1 [40], since expression of aSyn alone does not result in inclusion formation. In this aSyn aggregation model, aSyn inclusions are readily detected by immunocytochemistry using antibodies against aSyn, allowing the characterization of different types of inclusions [32], and screening modulators of aSyn aggregation [41]. 48 h after transfection, we analyzed inclusion formation in the cells (Fig. 4a), and observed that the A53E mutation did not alter inclusion formation when compared to WT aSyn (Fig. 4b). Again, no differences in the levels of aSyn or Synphilin-1 were detected between WT and A53E mutant aSyn (Fig. 4c-e).

aSyn is phosphorylated on Serine 129 (pS129) in LBs found in the brains of PD patients, linking this post-translational modification with disease [14, 52]. We previously showed that specific aSyn mutations, such as the E46K, can alter pS129 on aSyn [39]. Thus we investigated the effect of the A53E mutation on S129 phosphorylation. We found that both WT and A53E aSyn inclusions were positive for pS129 (Fig. 4f).

To further investigate the biochemical nature of the inclusions formed by the A53E aSyn mutant, we stained the cells with Thioflavin S (Th-S), a dye that binds to β-sheet rich amyloid structures [33]. We observed that the larger inclusions formed by WT and A53E aSyn stained positive for Th-S (Fig. 4g). In addition, we also used proteinase K (PK) resistance as a marker of aggregate formation, as protein inclusions tend to be more resistance to PK digestion. Interestingly, we found that the A53E mutant was less resistant when compared to WT aSyn (Fig. 4h and i), suggesting the inclusions formed by the A53E mutant have a less-compact nature than those formed by the WT protein.

The ubiquitin-proteasome system (UPS) is the major non-lysosomal pathway for selective protein degradation. In cell models, it has been shown that aSyn accumulation can affect the activity of the UPS system [61]. In our experimental conditions, we observed that cells expressing A53E aSyn mutant display increased proteolytic activity of the proteasome (Fig. 4j and k).

Table 2 Assignment of secondary structure components of aSyn variants in the amide I region of the FTIR spectra

	WT			A53E		
	Band (cm^{-1})	Area (%)	Structure	Band (cm^{-1})	Area (%)	Structure
1	1625	40	β -sheet (inter)	1628	24	β -sheet (inter)
2	1645	22		1649	38	
3	1663	25	Loop/β-turn/bend/α-helix	1666	25	Loop/β-turn/bend/α-helix
4	1680	13	β-turn	1681	14	β-turn

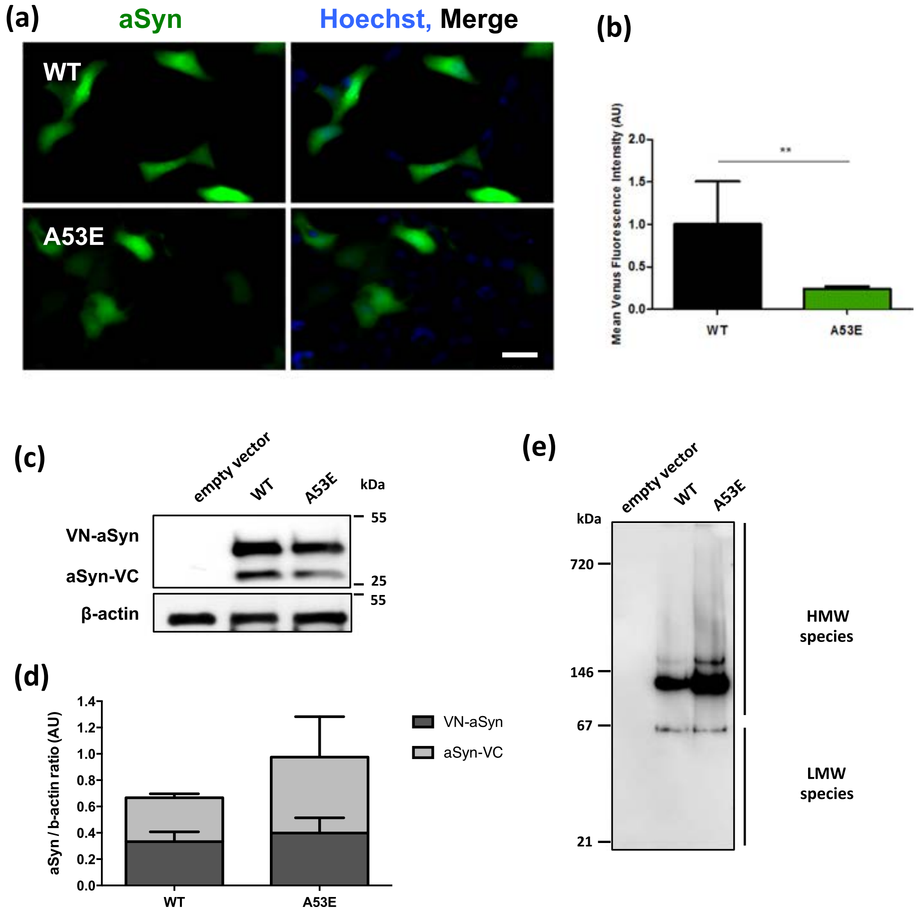


Fig. 3 A53E reduces aSyn oligomerization. **a** Fluorescent cells, expressing VN-aSyn and aSyn-VC constructs, as a result of the aSyn interaction. Scale bar: 30 μm. **b** Mean fluorescence intensity of cells were assessed 24 h post-transfection using an Olympus IX81-ZDC microscope. For each condition, 100 pictures were acquire in 4 independent experiments were conducted. Student's *t* test ($^{**}p < 0.01$). **c**-**d** aSyn protein levels were assessed by immunoblot analysis and were found to similar between WT aSyn and the A53E mutant. $n = 3$. **e** Native-PAGE gel showed that the A53E mutant forms high molecular weight species similar to WT aSyn

aSyn aggregation leads to Golgi fragmentation

aSyn induces several cellular pathologies that have been documented over the years and are routinely used to assess the effect of specific mutations or genetic interactors [60]. One particular type of cellular pathology associated with aSyn toxicity is the fragmentation of the Golgi apparatus [13]. This is also evident in other neurodegenerative diseases [18], suggesting it might be a more general response to the proteotoxicity associated with protein misfolding and aggregation. To assess whether expression of A53E mutant aSyn affected the integrity of the Golgi, we analyzed the morphology of this organelle in both the aSyn oligomerization and aggregation models (Fig. 5a and c). We classified the morphology of the Golgi as normal, diffuse and fragmented, as we previously described [32].

In the oligomerization model, we observed that WT aSyn reduced the percentage of cells exhibiting normal Golgi morphology (~40%, $^{***}p < 0.001$ Fig. 5a and b and Additional file 1: Figure S3.1–3.3), as we previously reported [32]. However, in cells expressing the A53E mutant the effects were not as pronounced as with WT, and the phenotype was more similar to that of cells carrying an empty vector (~70% and 80% of the transfected cells displayed normal Golgi morphology for A53E, and empty vector, respectively) ($^{*}p < 0.05$) (Fig. 5b and Additional file 1: Figure S3.1).

In the aSyn aggregation model, we observed the opposite effect. Around 50% of the cells expressing the A53E SynT displayed normal Golgi morphology whereas around 70% of the cells expressing WT SynT displayed normal Golgi (Fig. 5c and d and Additional file 1: Figure S2.4–2.6; $^{***}p < 0.001$ and $^{*}p < 0.01$ for empty vector, and WT, respectively). This suggests that the aggregation of A53E aSyn induces Golgi alterations.

A53E aSyn behaves identically to WT aSyn in yeast cells

Yeast cells have been extremely useful to assess cellular pathologies associated with the expression of aSyn. Thus,

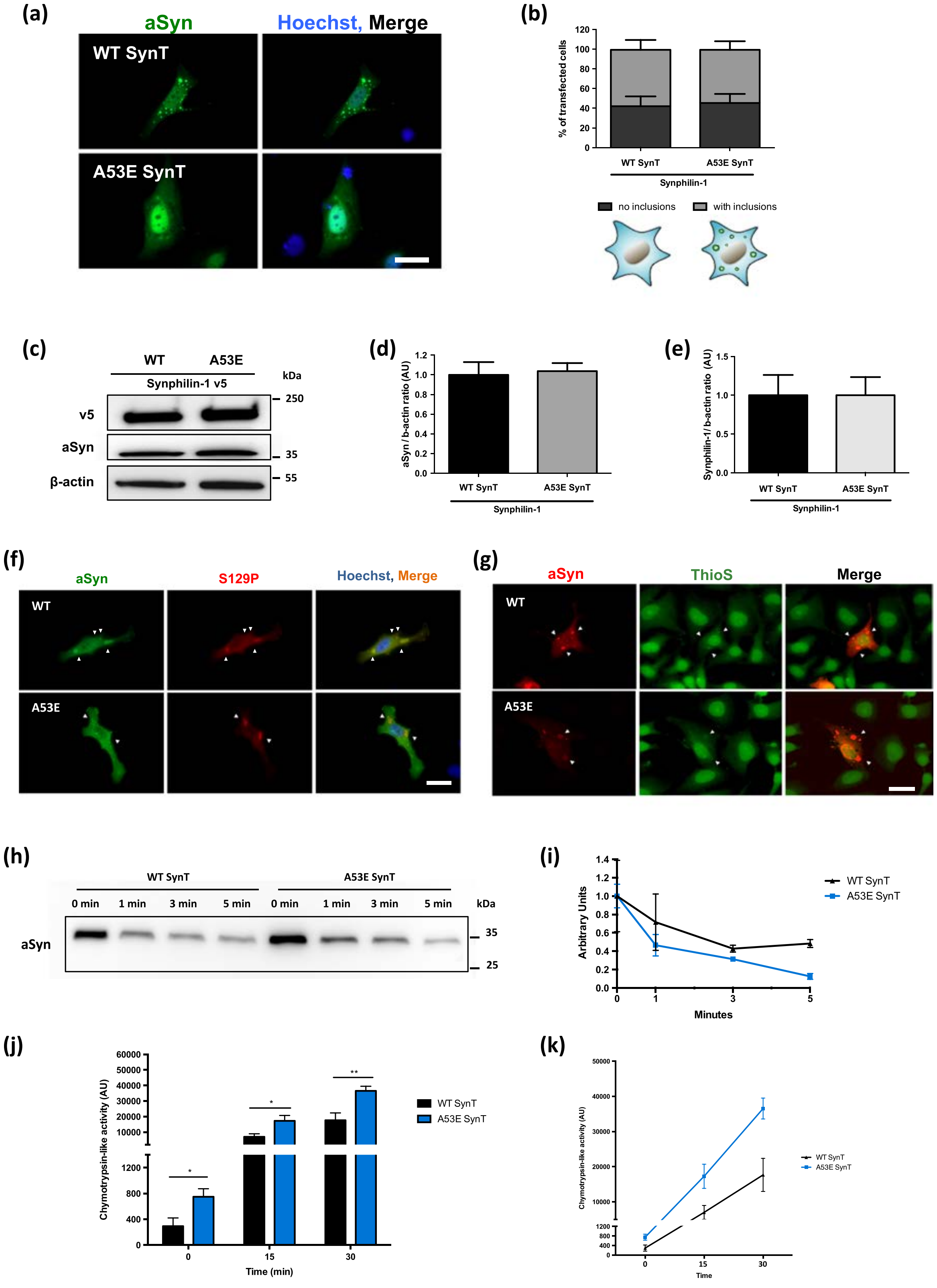


Fig. 4 (See legend on next page.)

(See figure on previous page.)
Fig. 4 A53E does not change the inclusion pattern. **a**-**b** At least 50 cells *per* condition were classified according to the pattern formed. We observed that A53E did not change the number of inclusions per cells. $n = 3$. Scale bar: 30 μm. **c**-**e** Immunoblot analysis of the aSyn and Synphilin-1 levels showed no significant differences in expression of WT or A53E aSyn. $n = 3$ (**f**) Inclusions formed by WT and A53E are positive for pS129. Positive inclusions are indicated with *white arrows*. Scale bar: 30 μm. **g** Inclusions were stained with Th-S and analyzed *via* fluorescence microscopy. As indicated with *arrow* heads, we observed that some inclusions displayed amyloid-like properties, by staining positive with Th-S. Scale bar: 30 μm. **h**-**i** WT and A53E aSyn protein lysates were digested with PD for different times (1, 3 and 5 min). After normalization of the values to the undigested condition, we observed that A53E inclusions are less resistant to PK-digestion. $n = 2$. **j**-**k** 48 h post-transfection the cells were collected and we assessed the activity of the proteasome. We observed that cells expressing A53E mutant increased proteolytic activity of the proteasome in comparison with WT. $n = 3$. Student's *t* test (*$p < 0.05$, **$p < 0.01$)

in order to assess if the A53E mutation alters the cytotoxicity and aggregation of aSyn, we expressed this mutant in *S. cerevisiae* and monitored phenotypes previously established [44]. We expressed WT or mutant A53E aSyn fused to eGFP using multi-copy (2 μ) plasmids and under the regulation of a galactose-inducible promoter (*GAL1*). aSyn cytotoxicity was first evaluated by a spotting assay. The growth of the cells expressing A53E aSyn was compared to that of cells expressing WT aSyn (Fig. 6a). As described before, expression of WT aSyn is toxic and results in reduced cell growth (Fig. 6a). We found that cells expressing A53E aSyn display a similar phenotype, suggesting that this familial mutation does not significantly affect aSyn toxicity in yeast (Fig. 6a). To further dissect the cytotoxicity of the A53E mutant, we performed propidium iodide (PI) staining and flow cytometry analysis, a readout of plasma membrane integrity (Fig. 6b). We observed that, 7 h after induction of aSyn expression, no significant differences were observed between cells expressing either WT or the A53E aSyn (Fig. 6b).

Next, we evaluated the subcellular distribution of the A53E mutant aSyn, using fluorescence microscopy. 7 h after induction of aSyn expression we assessed inclusion formation in the cells (Fig. 6c). The expression of A53E aSyn resulted in the formation of cytoplasmic inclusions that looked similar to those formed in cells expressing aSyn WT (Fig. 6c). In addition, no significant differences were observed in the percentage of cells with inclusions between WT and the A53E (Fig. 6c). We also evaluated the levels of aSyn expression by immunoblot analyses and found that both proteins were expressed at similar levels (Fig. 6d). The levels of phosphorylation on serine 129 were also indistinguishable between WT and A53E mutant aSyn (Fig. 6d).

Discussion

aSyn plays a major role in the pathological processes involved in neurodegenerative diseases, like PD or Dementia with Lewy Bodies [26]. When overexpressed in cells, to mimic familial forms of PD associated with multiplications of the aSyn gene, aSyn can promote cytotoxicity and impair vital processes, thereby contributing to cell death [9, 38, 44]. However, the precise molecular mechanisms underlying aSyn toxicity are still unclear, compromising our ability to intervene therapeutically. Both mutations and multiplications of the *SNCA* gene cause familial forms of PD [8, 22, 64]. Currently, six missense mutations in aSyn have been associated with autosomal dominant forms of parkinsonism (A30P, E46K, H50Q, G51D, A53E, A53T) [1, 30, 34, 46–48, 65]. Of these, the A53E was the last one to be identified and, therefore, has been less investigated.

In this study, we aimed to investigate the effect of the substitution of the alanine at position 53 by a glutamic acid residue that introduces an additional negative charge in the protein. For this purpose, we used in vitro techniques to characterize the biophysical effects of the mutation, and exploited cell-based models to assess the effects of the expression of the A53E mutant on the distribution, aggregation, and toxicity of the protein.

In vitro, we observed that A53E attenuates aSyn aggregation and reduces amyloid fibril formation when compared to WT aSyn. In fact, the formation of amyloid structures by the A53E mutant is marginal, since A53E is unable to bind CR. Overall, these observations are in line with a previous report showing that the presence of a negatively-charged residue can reduce the intrinsic aggregation propensity of aSyn [15]. Nevertheless, because the change in net charge in aSyn is small, −9 and −10 for the WT and A53E proteins, respectively, the effect this mutation has on aggregation has more likely a local origin. We used the Amylpred2 consensus aggregation predictor to analyze if the A53E mutation might have an impact in the intrinsic aggregation propensity of the aSyn sequence. Amylpred2 identifies a hot spot of aggregation corresponding to the 49–55 sequence stretch (VHGVATV), including Ala53. The A53E mutation shortens the aggregation region now including only residues 49–53 (VHGVE). The AGGRESCAN algorithm enables the comparison of the aggregation propensity of the two regions, showing that it is 2.1 lower for A53E than for WT aSyn.

In the context of a cell, the behavior of a protein is subjected not only to the crowded environment but also to the action of various protein quality control mechanisms. In human cells, we observed that the A53E mutation reduces aSyn oligomerization without changing the aggregation pattern. Interestingly, the inclusions formed

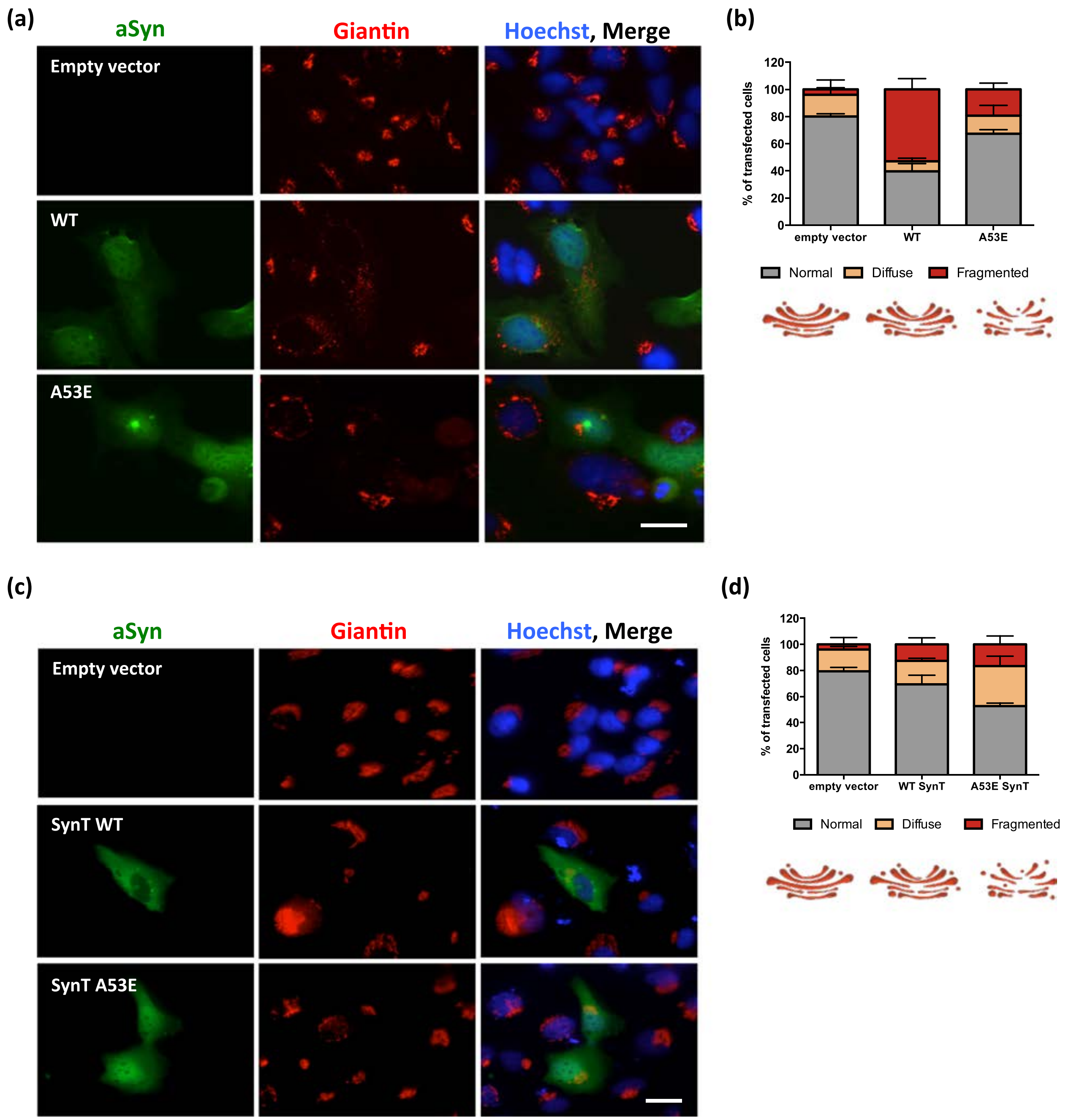


Fig. 5 Golgi morphology in the oligomerization and aggregation models. **a**-**b** Representative pictures of transfected cells with the BiFC system. We analyzed and categorized transfected cells in different categories. Scale bar: 30 μm. **b** WT aSyn resulted in an increased percentage of cells with fragmented Golgi morphology when compared with the empty vector and with cells expressing the A53E mutant. **c**-**d** Representative pictures of transfected cells with SynT + Synphilin-1. Scale bar: 30 μm. **d** In the presence of A53E SynT, the Golgi is more diffuse when compared with the control and with WT SynT. *n* = 3

by A53E aSyn are more sensitive to PK digestion than those formed by WT aSyn, suggesting that the inclusions formed are less compact or, possibly, more immature. The negative charge introduced by the glutamate residue can perhaps alter the intermolecular interactions between aSyn molecules, and prevent the formation of tighter inclusions. This may also correlate with the increase in proteasome activity that we observed, since this is an important degradation system to eliminate soluble proteins and smaller assemblies that are not degraded by autophagy.

In our previous studies, we did not observe major differences when comparing another PD-associated aSyn mutant

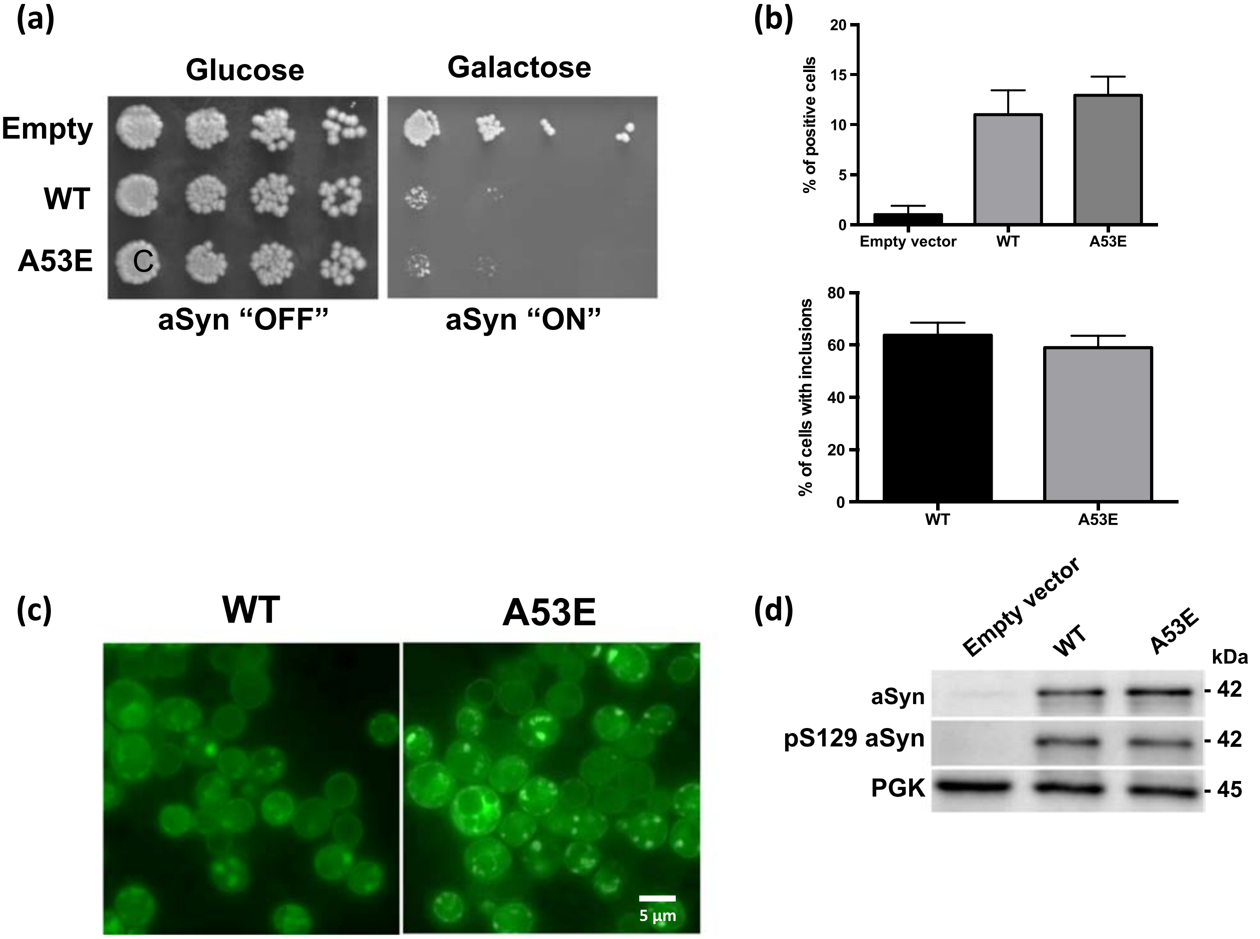


Fig. 6 Phenotypic characterization of yeast cells expressing the A53E aSyn mutant. **a** Cytotoxicity of WT and A53E aSyn in yeast cells compared to the empty vector, assessed by spotting assay. Photos were taken 3 days after incubation at 30 °C. **b** Frequency of PI positive cells assessed by flow cytometry, after 7 h of induction of expression of WT and A53E aSyn. **c** Fluorescence microscopy visualization (*left panel*) and percentage of cells with WT and A53E aSyn inclusions (*right panel*). **d** Expression levels of WT and A53E aSyn-GFP in yeast cells assessed by western blot analysis of total protein extracts. Results shown are from one representative experiment from at least three independent experiments. Values represent the mean ± SD of at least three independent measurements

at position 53 (A53T) with WT protein. We found that the aggregation of A53T aSyn was identical to that of WT aSyn [32], suggesting again that the change in charge at position 53 may influence the initial steps of the aggregation process (dimerization/oligomerization) of aSyn, and not the later steps that culminate with the formation of the mature inclusions. Also, the differences might result by difference spatial sequestration of aSyn. Misfolded proteins can be targetted to specific compartments, like aggresomes, to facilitate their degradation [23, 24, 62]. These compartments rely on filament network proteins like vimentin, actin, and tubulin, or proteins that can assist in the transportation of misfolded proteins through the microtubule network, such as p62 (24843142, 12093283, 14623963, 24086678). For example, Tubulin Polymerization Promoting Protein (TPPP/p25), belongs to the microtubule network, and associates with aSyn in pathological conditions, possibly affecting aSyn aggregation (17123092).

The Golgi apparatus plays a determinant role in the intracellular flow of several endogenous proteins and exogenous macromolecules, regulating the trafficking to their final destination inside or outside cells [12]. Fragmentation of the Golgi apparatus is a characteristic feature in several neurodegenerative diseases, including PD [13, 18, 19]. Interestingly, in our cell-based aggregation models, we observed a striking loss of the typical Golgi morphology in cells expressing A53E aSyn. Previous studies about the effect of this recently identified PD mutation in aSyn showed that A53E is as toxic as WT aSyn [15, 50]. While this was also the trend we observed in our cellular models, our findings suggest that the A53E mutant may cause specific alterations in cell physiology that are still not understood and demand additional investigation.

Conclusions

Overall, our study demonstrates that the A53E aSyn mutation influences the ability of aSyn to aggregate in vitro and in vivo, and that it may induce different cellular pathologies that should be further investigated in vivo. Ultimately, a deeper understanding of the effect of aSyn mutations on the behavior of the protein will enable the design of novel models to assess the potential value of future therapeutic strategies for PD and other synucleinopathies.

Additional file

Additional file 1: Figure S1.1. Biochemical characterization of WT and A53E recombinant aSyn. Analysis of purified WT and A53E aSyn variants by SDS-PAGE (18%) stained with Coomassie Brilliant Blue. **S1.2.** MALDI-TOF mass spectrometry analysis of purified WT (up) and A53E (down) aSyn variants. A MW of 14462.98 Da was obtained for WT (theoretical of 14460.16 da) and of 14519.47 Da for A53E (theoretical of 14518.19 Da) aSyn variants. The peaks of 7231.82 and 7260.97 Da correspond to the M^{+2} ions of WT and A53E aSyn, respectively. **Figure S2.1 and S2.2.** Immunoblot quantifications. Levels of VN-aSyn (S2.1) and aSyn-VC (S2.1). *n*=3. **Figure S3.1-S3.3.** Morphological analysis of Golgi apparatus in the aSyn BiFC system. The morphology of the Golgi was analyzed as normal (S3.1), diffuse (S3.2) and fragmented (S3.3). One-way ANOVA with post-hoc Tukey's test (**p*<0.05, ***p*<0.01, ****p*<0.001). *n*=3. **Figure S3.4-3.6** Morphological analysis of Golgi apparatus in the aSyn aggregation model. Transfected cells were analyzed according to the morphology of the Golgi: normal (S3.4), diffuse (S3.5) and fragmented (S3.6) One-way ANOVA with post-hoc Tukey's test (**p*<0.05, ***p*<0.01, ****p*<0.001). *n*=3. (PDF 389 kb)

Abbreviations

A53E: Substitution of alanine at position 53 by a glutamate residue; aSyn: alpha-synuclein; ATR FT-IR: Attenuated total reflectance Fourier transform infrared spectroscopy; BiFC: Bimolecular fluorescence complementation assay; CR: Congo red; H4: Human Neuroglioma cells; HEK: Human embryonic kidney cells; IPTG: isopropyl-1-thio-β-D-galactopyranoside; LBs: Lewy bodies; PD: Parkinson's disease; PI: Propidium iodide; PK: Proteinase K; RT: Room temperature; SC: synthetic complete medium; TEM: Transmission electron microscopy; Th-S: Thioflavin S; Th-T: Thioflavin T; UPS: Ubiquitin-proteasome system

Acknowledgements

TFO is supported by the DFG Center for Nanoscale Microscopy and Molecular Physiology of the Brain, by the German-Israeli Foundation for Scientific Research and Development (GIF), and by a grant from the Niedersachsisches Ministerium fur Wissenschaft und Kultur (MWK). TFO and SV are supported by a grant from Fundación La Marato de TV3 (Ref. 20144330). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Authors' contributions

Conceived and designed the experiments: DFL, MD, SV, TFO. Performed the experiments: DFL, MD, AC, SN, CSM, ST. Analyzed the data: DFL, MD, SV, ST, TFO. Wrote the paper: DFL, MD, SV, TFO. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Department of Neurodegeneration and Restorative Research, University Medical Center Göttingen, Waldweg 33, 37073 Göttingen, Germany. [2]Institut de Biotecnologia i Biomedicina and Departament de Bioquímica i Biologia Molecular, Universitat Autònoma de Barcelona, 08193, Bellaterra, Barcelona, Spain. [3]Chronic Disease Research Center (CEDOC), NOVA Medical School, Campo dos Mártires da Pátria, 130, 1169-056 Lisbon, Portugal. [4]Max Planck Institute for Experimental Medicine, Göttingen, Germany.

Received: 29 October 2016 Accepted: 3 December 2016

References

1. Appel-Cresswell S, Vilarino-Guell C, Encarnacion M, Sherman H, Yu I, Shah B, Weir D, Thompson C, Szu-Tu C, Trinh J et al (2013) Alpha-synuclein p.H50Q, a novel pathogenic mutation for Parkinson's disease. Mov Disord 28:811–813
2. Auluck PK, Caraveo G, Lindquist S (2010) alpha-Synuclein: membrane interactions and toxicity in Parkinson's disease. Annu Rev Cell Dev Biol 26:211–233
3. Barona-Lleo L, Zulueta-Santos C, Murie-Fernandez M, Perez-Fernandez N (2014) Recent onset disequilibrium mimicking acute vestibulopathy in early multiple sclerosis. Am J Otolaryngol 35:529–534
4. Braak H, Del Tredici K, Rub U, de Vos RA, Jansen Steur EN, Braak E (2003) Staging of brain pathology related to sporadic Parkinson's disease. Neurobiol Aging 24:197–211
5. Braak H, Ghebremedhin E, Rub U, Bratzke H, Del Tredici K (2004) Stages in the development of Parkinson's disease-related pathology. Cell Tissue Res 318:121–134
6. Cai H, Reinisch K, Ferro-Novick S (2007) Coats, tethers, Rabs, and SNAREs work together to mediate the intracellular destination of a transport vesicle. Dev Cell 12:671–682
7. Chandra S, Gallardo G, Fernandez-Chacon R, Schluter OM, Sudhof TC (2005) Alpha-synuclein cooperates with CSPalpha in preventing neurodegeneration. Cell 123:383–396
8. Chartier-Harlin MC, Dachsel JC, Vilarino-Guell C, Lincoln SJ, Lepretre F, Hulihan MM, Kachergus J, Milnerwood AJ, Tapia L, Song MS et al (2011) Translation initiator EIF4G1 mutations in familial Parkinson disease. Am J Hum Genet 89:398–406
9. Cooper AA, Gitler AD, Cashikar A, Haynes CM, Hill KJ, Bhullar B, Liu K, Xu K, Strathearn KE, Liu F et al (2006) Alpha-synuclein blocks ER-Golgi traffic and Rab1 rescues neuron loss in Parkinson's models. Science 313:324–328
10. Coppede F (2012) Genetics and epigenetics of Parkinson's disease. Sci World J 2012:489830
11. Dikiy I, Eliezer D (2014) N-terminal acetylation stabilizes N-terminal helicity in lipid- and micelle-bound alpha-synuclein and increases its affinity for physiological membranes. J Biol Chem 289:3652–3665
12. Fan J, Hu Z, Zeng L, Lu W, Tang X, Zhang J, Li T (2008) Golgi apparatus and neurodegenerative diseases. Int J Dev Neurosci 26:523–534
13. Fujita Y, Ohama E, Takatama M, Al-Sarraj S, Okamoto K (2006) Fragmentation of Golgi apparatus of nigral neurons with alpha-synuclein-positive inclusions in patients with Parkinson's disease. Acta Neuropathol 112:261–265
14. Fujiwara H, Hasegawa M, Dohmae N, Kawashima A, Masliah E, Goldberg MS, Shen J, Takio K, Iwatsubo T (2002) alpha-Synuclein is phosphorylated in synucleinopathy lesions. Nat Cell Biol 4:160–164
15. Ghosh D, Sahay S, Ranjan P, Salot S, Mohite GM, Singh PK, Dwivedi S, Carvalho E, Banerjee R, Kumar A, Maji SK (2014) The newly discovered Parkinson's disease associated Finnish mutation (A53E) attenuates alpha-synuclein aggregation and membrane binding. Biochemistry 53:6419–6421
16. Gietz D, St Jean A, Woods RA, Schiestl RH (1992) Improved method for high efficiency transformation of intact yeast cells. Nucleic Acids Res 20:1425
17. Goedert M, Spillantini MG, Del Tredici K, Braak H (2013) 100 years of Lewy pathology. Nat Rev Neurol 9:13–24
18. Gonatas NK, Stieber A, Gonatas JO (2006) Fragmentation of the Golgi apparatus in neurodegenerative diseases and cell death. J Neurol Sci 246:21–30
19. Gosavi N, Lee HJ, Lee JS, Patel S, Lee SJ (2002) Golgi fragmentation occurs in the cells with prefibrillar alpha-synuclein aggregates and precedes the formation of fibrillar inclusion. J Biol Chem 277:48984–48992
20. Hardy J (2005) Expression of normal sequence pathogenic proteins for neurodegenerative disease contributes to disease risk: 'permissive templating' as a general mechanism underlying neurodegeneration. Biochem Soc Trans 33:578–581

21. Hernandez-Vargas R, Fonseca-Ornelas L, Lopez-Gonzalez I, Riesgo-Escovar J, Zurita M, Reynaud E (2011) Synphilin suppresses alpha-synuclein neurotoxicity in a Parkinson's disease Drosophila model. Genesis 49:392–402
22. Ibanez P, Bonnet AM, Debarges B, Lohmann E, Tison F, Pollak P, Agid Y, Durr A, Brice A (2004) Causal relation between alpha-synuclein gene duplication and familial Parkinson's disease. Lancet 364:1169–1171
23. Johnston JA, Ward CL, Kopito RR (1998) Aggresomes: a cellular response to misfolded proteins. J Cell Biol 143:1883–1898
24. Kaganovich D, Kopito R, Frydman J (2008) Misfolded proteins partition between two distinct quality control compartments. Nature 454:1088–1095
25. Kannarkat GT, Boss JM, Tansey MG (2013) The role of innate and adaptive immunity in Parkinson's disease. J Park Dis 3:493–514
26. Kim WS, Kagedal K, Halliday GM (2014) Alpha-synuclein biology in Lewy body diseases. Alzheimers Res Ther 6:73
27. Kisselev AF, Goldberg AL (2005) Monitoring activity and inhibition of 26S proteasomes with fluorogenic peptide substrates. Methods Enzymol 398:364–378
28. Klunk WE, Pettegrew JW, Abraham DJ (1989) Quantitative evaluation of congo red binding to amyloid-like proteins with a beta-pleated sheet conformation. J Histochem Cytochem 37:1273–1281
29. Kordower JH, Brundin P (2009) Propagation of host disease to grafted neurons: accumulating evidence. Exp Neurol 220:224–225
30. Kruger R, Kuhn W, Muller T, Woitalla D, Graeber M, Kosel S, Przuntek H, Epplen JT, Schols L, Riess O (1998) Ala30Pro mutation in the gene encoding alpha-synuclein in Parkinson's disease. Nat Genet 18:106–108
31. Lashuel HA, Overk CR, Oueslati A, Masliah E (2013) The many faces of alpha-synuclein: from structure and toxicity to therapeutic target. Nat Rev Neurosci 14:38–48
32. Lazaro DF, Rodrigues EF, Langohr R, Shahpasandzadeh H, Ribeiro T, Guerreiro P, Gerhardt E, Krohnert K, Klucken J, Pereira MD et al (2014) Systematic comparison of the effects of alpha-synuclein mutations on its oligomerization and aggregation. PLoS Genet 10:e1004741
33. Lee HJ, Lee SJ (2002) Characterization of cytoplasmic alpha-synuclein aggregates. Fibril formation is tightly linked to the inclusion-forming process in cells. J Biol Chem 277:48976–48983
34. Lesage S, Anheim M, Letournel F, Bousset L, Honore A, Rozas N, Pieri L, Madiona K, Durr A, Melki R et al (2013) G51D alpha-synuclein mutation causes a novel parkinsonian-pyramidal syndrome. Ann Neurol 73:459–471
35. LeVine H 3rd (1993) Thioflavine T interaction with synthetic Alzheimer's disease beta-amyloid peptides: detection of amyloid aggregation in solution. Protein Sci 2:404–410
36. Li JY, Englund E, Holton JL, Soulet D, Hagell P, Lees AJ, Lashley T, Quinn NP, Rehncrona S, Bjorklund A et al (2008) Lewy bodies in grafted neurons in subjects with Parkinson's disease suggest host-to-graft disease propagation. Nat Med 14:501–503
37. Maroteaux L, Campanelli JT, Scheller RH (1988) Synuclein: a neuron-specific protein localized to the nucleus and presynaptic nerve terminal. J Neurosci 8:2804–2815
38. Martin LJ, Pan Y, Price AC, Sterling W, Copeland NG, Jenkins NA, Price DL, Lee MK (2006) Parkinson's disease alpha-synuclein transgenic mice develop neuronal mitochondrial degeneration and cell death. J Neurosci 26:41–50
39. Mbefo MK, Fares MB, Paleologou K, Oueslati A, Yin G, Tenreiro S, Pinto M, Outeiro T, Zweckstetter M, Masliah E, Lashuel HA (2015) Parkinson disease mutant E46K enhances alpha-synuclein phosphorylation in mammalian cell lines, in yeast, and in vivo. J Biol Chem 290:9412–9427
40. McLean PJ, Kawamata H, Hyman BT (2001) Alpha-synuclein-enhanced green fluorescent protein fusion proteins form proteasome sensitive inclusions in primary neurons. Neuroscience 104:901–912
41. Moree B, Yin G, Lazaro DF, Munari F, Strohaker T, Giller K, Becker S, Outeiro TF, Zweckstetter M, Salafsky J (2015) Small molecules detected by second-harmonic generation modulate the conformation of monomeric alpha-synuclein and reduce its aggregation in cells. J Biol Chem 290:27582–27593
42. Murphy DD, Rueter SM, Trojanowski JQ, Lee VM (2000) Synucleins are developmentally expressed, and alpha-synuclein regulates the size of the presynaptic vesicular pool in primary hippocampal neurons. J Neurosci 20:3214–3220
43. Olanow CW, Perl DP, DeMartino GN, McNaught KS (2004) Lewy-body formation is an aggresome-related process: a hypothesis. Lancet Neurol 3:496–503
44. Outeiro TF, Lindquist S (2003) Yeast cells provide insight into alpha-synuclein biology and pathobiology. Science 302:1772–1775
45. Outeiro TF, Putcha P, Tetzlaff JE, Spoelgen R, Koker M, Carvalho F, Hyman BT, McLean PJ (2008) Formation of toxic oligomeric alpha-synuclein species in living cells. PLoS One 3:e1867
46. Pasanen P, Myllykangas L, Siitonen M, Raunio A, Kaakkola S, Lyytinen J, Tienari PJ, Poyhonen M, Paetau A (2014) Novel alpha-synuclein mutation A53E associated with atypical multiple system atrophy and Parkinson's disease-type pathology. Neurobiol Aging 35:2180.e2181–2185
47. Polymeropoulos MH, Lavedan C, Leroy E, Ide SE, Dehejia A, Dutra A, Pike B, Root H, Rubenstein J, Boyer R et al (1997) Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. Science 276:2045–2047
48. Proukakis C, Dudzik CG, Brier T, MacKay DS, Cooper JM, Millhauser GL, Houlden H, Schapira AH (2013) A novel alpha-synuclein missense mutation in Parkinson disease. Neurology 80:1062–1064
49. Rey NL, Steiner JA, Maroof N, Luk KC, Madaj Z, Trojanowski JQ, Lee VM, Brundin P (2016) Widespread transneuronal propagation of alpha-synucleinopathy triggered in olfactory bulb mimics prodromal Parkinson's disease. J Exp Med 213:1759–1778
50. Rutherford NJ, Giasson BI (2015) The A53E alpha-synuclein pathological mutation demonstrates reduced aggregation propensity in vitro and in cell culture. Neurosci Lett 597:43–48
51. Sabate R, Rodriguez-Santiago L, Sodupe M, Saupe SJ, Ventura S (2013) Thioflavin-T excimer formation upon interaction with amyloid fibers. Chem Commun 49:5745–5747
52. Saito Y, Kawashima A, Ruberu NN, Fujiwara H, Koyama S, Sawabe M, Arai T, Nagura H, Yamanouchi H, Hasegawa M et al (2003) Accumulation of phosphorylated alpha-synuclein in aging human brain. J Neuropathol Exp Neurol 62:644–654
53. Singleton AB, Farrer M, Johnson J, Singleton A, Hague S, Kachergus J, Hulihan M, Peuralinna T, Dutra A, Nussbaum R et al (2003) alpha-Synuclein locus triplication causes Parkinson's disease. Science 302:841
54. Spillantini MG, Schmidt ML, Lee VM, Trojanowski JQ, Jakes R, Goedert M (1997) Alpha-synuclein in Lewy bodies. Nature 388:839–840
55. Spillantini MG, Crowther RA, Jakes R, Hasegawa M, Goedert M (1998) alpha-Synuclein in filamentous inclusions of Lewy bodies from Parkinson's disease and dementia with lewy bodies. Proc Natl Acad Sci U S A 95:6469–6473
56. Tenreiro S, Reimao-Pinto MM, Antas P, Rino J, Wawrzycka D, Macedo D, Rosado-Ramos R, Amen T, Waiss M, Magalhaes F et al (2014) Phosphorylation modulates clearance of alpha-synuclein inclusions in a yeast model of Parkinson's disease. PLoS Genet 10:e1004302
57. Tenreiro S, Rosado-Ramos R, Gerhardt E, Favretto F, Magalhaes F, Popova B, Becker S, Zweckstetter M, Braus GH, Outeiro TF (2016) Yeast reveals similar molecular mechanisms underlying alpha- and beta-synuclein toxicity. Hum Mol Genet 25:275–290
58. Tsigelny IF, Sharikov Y, Wrasidlo W, Gonzalez T, Desplats PA, Crews L, Spencer B, Masliah E (2012) Role of alpha-synuclein penetration into the membrane in the mechanisms of oligomer pore formation. FEBS J 279:1000–1013
59. Volles MJ, Lansbury PT Jr (2007) Relationships between the sequence of alpha-synuclein and its membrane affinity, fibrillization propensity, and yeast toxicity. J Mol Biol 366:1510–1522
60. Wales P, Pinho R, Lazaro DF, Outeiro TF (2013) Limelight on alpha-synuclein: pathological and mechanistic implications in neurodegeneration. J Park Dis 3:415–459
61. Webb JL, Ravikumar B, Atkins J, Skepper JN, Rubinsztein DC (2003) Alpha-Synuclein is degraded by both autophagy and the proteasome. J Biol Chem 278:25009–25013
62. Weisberg SJ, Lyakhovetsky R, Werdiger AC, Gitler AD, Soen Y, Kaganovich D (2012) Compartmentalization of superoxide dismutase 1 (SOD1G93A) aggregates determines their toxicity. Proc Natl Acad Sci U S A 109:15811–15816
63. Winner B, Jappelli R, Maji SK, Desplats PA, Boyer L, Aigner S, Hetzer C, Loher T, Vilar M, Campioni S et al (2011) In vivo demonstration that alpha-synuclein oligomers are toxic. Proc Natl Acad Sci U S A 108:4194–4199
64. Xu W, Tan L, Yu JT (2015) Link between the SNCA gene and parkinsonism. Neurobiol Aging 36:1505–1518
65. Zarranz JJ, Alegre J, Gomez-Esteban JC, Lezcano E, Ros R, Ampuero I, Vidal L, Hoenicka J, Rodriguez O, Atares B et al (2004) The new mutation, E46K, of alpha-synuclein causes Parkinson and Lewy body dementia. Ann Neurol 55:164–173

Data in Brief 7 (2016) 143–147

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Data on correlation between Aβ42 structural aggregation propensity and toxicity in bacteria

Anita Carija [1], Susanna Navarro [1], Salvador Ventura

Institut de Biotecnologia i Biomedicina, Departament de Bioquimica i Biologia Molecular, Universitat Autònoma de Barcelona, Bellaterra, 08193 Barcelona, Spain

ARTICLE INFO

Article history:
Received 13 October 2015
Received in revised form
13 December 2015
Accepted 4 February 2016
Available online 12 February 2016

ABSTRACT

Protein aggregation and amyloid formation is a hallmark of an increasing number of human disorders. Because protein aggregation is deleterious for the cell physiology and results in a decrease in overall cell fitness, it is thought that natural selection acts to purify aggregating proteins during evolution. This data article contains complementary figures and results related to the research article entitled “Selection against toxic aggregation-prone protein sequences in bacteria” (Navarro et al., 2014) [1]. Here, we used the AGGRESCAN3D (A3D) server, a novel in house predictor that forecasts protein aggregation properties in protein structures to illustrate a striking correlation between the structure-based predictions of aggregation propensities for Alzheimer's Aβ42 peptide variants and their previously reported deleterious effects in bacteria.

Specifications Table

Subject area	*Biology*
More specific subject area	*Protein aggregation*
Type of data	*Table, Figures*

E-mail address: Salvador.Ventura@uab.es (S. Ventura).

[1] Both authors contributed equally.

http://dx.doi.org/10.1016/j.dib.2016.02.017

How data was acquired	*Aggrescan (bioinf.uab.es/*aggrescan*) and Aggrescan3D (http://biocomp.chem.uw.edu.pl/A3D) predictions.*
Data format	*Analyzed*
Experimental factors	*Aggregation propensities of Aβ42 peptide and two generated mutants F19D and F19D-L34P were analyzed with predictors based on the analysis of the linear sequence and the three dimensional structure.*
Experimental features	*A3D protein predictions are indicated in a table containing the total and average score for A3D prediction, shown as surfaced structures colored according to A3D score and related to biological properties.*
Data source location	*Not applicable*
Data accessibility	*Aβ42 structures* correspond to PDB: 2OTK, PDB: 2BEG, PDB: 2MXU, PDB: 2LMN.

Value of the data

- The data show that AGGRESCAN3D (A3D) is able to forecast Aβ42 intracellular protein aggregation propensity and its associated toxicity, while allowing visualizing and dissecting the contribution of the regions responsible for this undesired properties in the 3D space.
- The methodology used here to generate data on Aβ42 aggregation properties could be used for the study of the aggregation of other proteins involved in conformational disorders.
- These data are valuable to researchers interested in the relationship between the intrinsic aggregation properties of disease-linked proteins and its associated cytotoxic effect.

1. Data

Aβ peptide variants (wild type, F19D and F19D/L34P) aggregation propensities were calculated according to AGGRESCAN [2,3], which uses protein sequences as input and AGGRESCAN3D (A3D) [4], which instead uses 3D structures. The structures with PDB codes 2OTK, 2BEG, 2MXU, 2LMN, all corresponding to the Alzheimer's Aβ42 peptide were modeled.

Table 1
The aggregation propensity data obtained by AGGRESCAN and AGGRESCAN3D are represented for Aβ42wt peptide and variants F19D, F19D/L34P. Linear sequences were used to obtain Na4vSS (Normalized a4v Sequence Sum for 100 residues) values. To obtain data on the aggregation propensities of 3D-structures, A3D was used in either Static or Dynamic Modes and the indicated PDB files were used as input structures.

Protein	**AGGRESCAN**	**AGGRESCAN3D**									
		Static Mode								**Dynamic Mode**	
	Na4vSS	**2OTK**		**2BEG**		**2MXU**		**2LMN**		**2OTK**	
		Average score	Total score	Average score	Total score	Average score	Total score	Average score	Total score	Average score	Total score
Aβ42wt	6.4	0.8	21.0	1.3	33.6	0.8	26.9	0.6	18.0	1.6	40.8
Aβ42F19D	−2.2	0.4	9.3	0.9	23.5	0.5	16.6	0.3	8.9	1.2	30.0
Aβ42F19D/L34P	−6.3	0.1	2.3	0.7	18.6	0.4	11.3	0.1	4.3	0.8	20.6

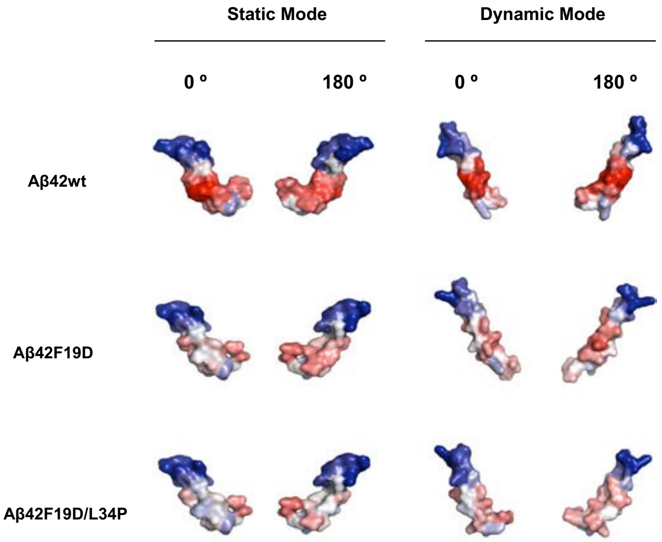


Fig. 1. Aβ42wt peptide (PDB: 2OTK:C) and variants F19D, F19D/L34P were modeled and analyzed using A3D in both Static and Dynamic Mode. The protein surfaces shown at 0° and 180° are colored according to A3D score in a gradient from: red (high-predicted aggregation propensity) to white (negligible impact on protein aggregation) to blue (high-predicted solubility). Both Static and Dynamic prediction modes show reduced surface-aggregation propensity in the designed variants when compared with the Aβ42wt.

AGGRESCAN protein aggregation prediction data is provided as the global protein aggregation propensity of the sequence (Na4vSS). With regard to A3D prediction, the total and the average scores corresponding to the overall and average aggregation propensities of the analyzed protein structures are provided. Both in AGGRESCAN and A3D predictions the smallest the score is the highest it is the predicted solubility of the variant (Table 1). The Aβ42 peptide structures corresponding to PDB 2OTK and its variants were modeled using the static and dynamic modes. In Fig. 1 residues are colored according to their Aggrescan3D score. Table 1 and Fig. 1 illustrate the increasing solubilizing effect of the introduced mutations. Because in the used Aβ42 peptide structures the mutated side chains expose to solvent more than 25% of their surface there is a good correlation between AGGRESCAN and A3D scores.

A3D aggregation propensity data were compared with previously obtained biological data (Fig. 2), observing a striking correlation between the predicted and the experimentally determined solubility, measured as the total intracellular fluorescence of the GFP fused to the specific peptide variant [1]. Not surprisingly, the best correlation with A3D was found for the monomeric 2OTK structure, which in static mode exhibited an $R^2=0.994$, superior to the correlation found for AGGRESCAN predictions, with $R^2=0.960$. In the same manner, the A3D predicted aggregation propensity exhibits an excellent correlation with the impact the different peptides have on both cell metabolism and viability [1] (Fig. 2), with $R^2=0.998$ and $R^2=0.999$ for the 2OTK structure, respectively; being again more accurate than AGGRESCAN, which predictions exhibit correlation coefficients of $R^2=0.978$ and $R^2=0.988$ with the impacts the peptides cause on cell metabolism and viability, respectively.

2. Experimental design, materials and methods

2.1. Aggregation propensity predictions: AGGRESCAN vs. AGGRESCAN3D

We used two algorithms developed by our group to test their ability to predict the relative aggregation propensities of the Alzheimer's related Aβ42wt peptide and of two mutants with

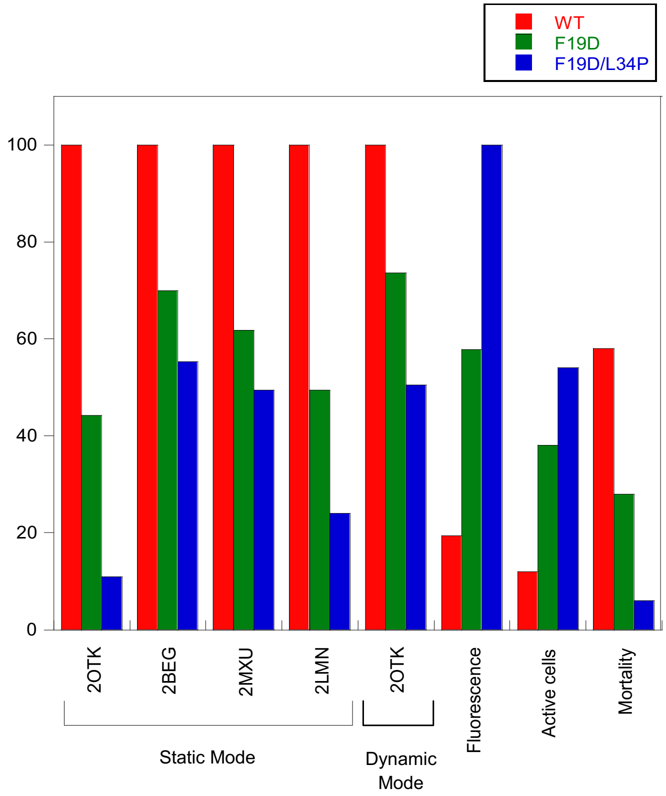


Fig. 2. Bar graph comparing the relative predicted aggregation propensities, GFP mean fluorescence as a reporter of protein solubility, metabolic activity and cell viability of variants F19D (green bars) and F19D/L34P (blue bars) with regards to Aβ42wt (red bars). Normalized total scores were obtained by A3D analysis of the indicated PDB files in Static and Dynamic Mode. Aggregation propensities of Aβ42 peptide 3D-structures can be correlated with previous experimental data reflecting the solubility of the protein and their impact in metabolic activity and cellular mortality in the bacterial population.

increased experimental solubility (Aβ42F19D and Aβ42F19D/L34P). AGGRESCAN [2] is a widely used algorithm that employs linear sequence as an input, while AGGRESCAN3D (A3D) [4] is a recently developed algorithm that implements a structure-based approach, uses as input protein 3D-structures derived from X-ray diffraction, solution NMR or modeling approaches and predicts aggregation propensity of initially folded states; this approach resembles that of the previously described Spatial Aggregation Propensity (SAP) suite [5].

Aβ42wt, F19D and F19D/L34P peptide sequences were submitted to AGGRESCAN in FASTA format and Na4vSS (Normalized a4v Sequence Sum for 100 residues) values were selected to compare the predictions. This value is obtained dividing the average aggregation propensity by the number of residues in the input amino acid sequence and multiplying it by 100. Aβ42wt structures corresponding to both the aggregated fibrillar state (PDB files: 2BEG:A, 2MXU:A and 2LMN:A) and the monomeric structure (PDB file: 2OTK:C) were used to analyze the aggregation propensity using A3D. For the fibrils structures the aggregation propensity of a single monomer in the fibrillar conformation was analyzed after energy minimization using the FoldX algorithm [6] integrated in A3D. All PDB files were submitted to A3D in 'Static Mode', while only the PDB file (2OTK:C) was also submitted in 'Dynamic Mode', since it corresponds to a real monomer in solution and not to a conformer dissected from the fibrillar structure. 10 Å was selected as a distance for aggregation analysis (default sphere radius). The following data were obtained from the output interfaces: average score and total score. The average score allows comparing the solubility of different protein structures. It also allows assessing changes in solubility promoted by amino acid substitutions in a particular protein structure. The total score is a global indicator of the aggregation propensity/solubility of the protein structure. It depends on the protein size. It allows assessing changes in solubility promoted by amino acid substitutions in a particular protein structure as long as they do not result in changes in protein size. The F19D and F19D/L34P mutants were modeled using the FOLDX force field implemented in A3D and analyzed subsequently. Pictures were made using the PyMOL software. The A3D server is available at: http://biocomp.chem.uw.edu.pl/A3D/.

Acknowledgments

Work in our lab is supported by Ministerio de Economia y Competividad, Spain [BFU2013-44763-P to S.V.].

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.dib.2016.02.017.

References

[1] S. Navarro, A. Villar-Piqué, S. Ventura, Selection against toxic aggregation-prone protein sequences in bacteria, Biochim. Biophys. Acta – Mol. Cell. Res. 1843 (2014) 866–874.

[2] O. Conchillo-Solé, N.S. de Groot, F.X. Avilés, J. Vendrell, X. Daura, S. Ventura, AGGRESCAN: a server for the prediction and evaluation of "hot spots" of aggregation in polypeptides, BMC Bioinforma. 8 (2007) 65. http://dx.doi.org/10.1186/1471-2105-8-65.

[3] N.S. de Groot, V. Castillo, R. Graña-Montes, S. Ventura, AGGRESCAN: method, application, and perspectives for drug design, Methods Mol. Biol. 819 (2012) 199–220. http://dx.doi.org/10.1007/978-1-61779-465-0-14.

[4] R. Zambrano, M. Jamroz, A. Szczasiuk, J. Pujols, S. Kmiecik, S. Ventura, AGGRESCAN3D (A3D): server for prediction of aggregation properties of protein structures, Nucleic Acids Res. (2015), http://dx.doi.org/10.1093/nar/gkv359.

[5] N. Chennamsetty, V. Voynov, V. Kayser, B. Helk, B.L. Trout, Design of therapeutic proteins with enhanced stability, Proc. Natl. Acad. Sci. U.S.A. 106 (2009) 11937–11942. http://dx.doi.org/10.1073/pnas.0904191106.

[6] L. Serrano, J. Schymkowitz, J. Borg, F. Stricher, R. Nys, F. Rousseau, The FoldX web server: an online force field, Nucleic Acids Res. 33 (2005) W382–W388. http://dx.doi.org/10.1093/nar/gki387.