Universitat Autònoma de Barcelona

Departament de Bioquímica i Biologia Molecular

Institut de Biotecnologia i Biomedicina

Connection between protein disorder, folding and aggregation: Physiological and Pathological implications

Doctoral thesis presented by Jordi Pujols Pujol for the degree of Ph. D. in Biochemistry, Molecular Biology and Biomedicine from the Autonomous University of Barcelona

The work of the present thesis has been performed in the Department of Biochemistry and Molecular Biology and the Institute of Biotechnology and Biomedicine, supervised by Prof. Francesc Xavier Avilés Puigvert and Prof. Salvador Ventura Zamora

Jordi Pujols Pujol

Prof. Francesc Xavier Avilés Puigvert   Prof. Salvador Ventura Zamora

Bellaterra, January 2021

English Summary

Intrinsically Disordered Proteins (IDPs) and Regions (IDRs) are a class of polypeptides that lack defined three-dimensional structures. Instead, they populate a dynamic ensemble of flexible conformers that endorse them with unique properties to interact with multiple partners and mediate in signal transduction. Environmental factors, post-translational modifications, and binding partners impact the IDPs' conformational space, promoting structural transitions that regulate their functions.

This thesis has addressed the disorder-to-order transition phenomenon from two perspectives: its physiological function and its implication in disease.

First, we studied the redox-driven conformational transition of two small disulfide-rich mitochondrial proteins. We demonstrate that both polypeptides follow different folding pathways despite sharing a common fold. The reaction is determined by the primary sequence of amino acids and the reduced and unfolded species' conformational state. Reduced COX17 is mostly disordered, and its folding is directed by the local packing of a loop region. In contrast, TRIAP1 reduced state resembles a molten globule. This early loosely packed conformer biases and slows down the folding pathway, but its population is unavoidable, because it arises from functional constraints in the folded state.

Second, we addressed the sequential overlap between amyloidogenic and interaction regions in the context of disordered proteins, to demonstrate that the extended C-terminus of the proteostasis guardian protein, Hsf1, contains a cryptic prion-like domain with the potential to assemble into amyloid fibrils.

Finally, we focused our attention on the aggregation of α-synuclein, the main culprit of the onset and progression of Parkinson's Disease. This protein's disordered nature precludes the use of rational strategies to design effective drugs that block its transition to ordered and deleterious amyloids. Therefore, we implemented a blind high-throughput screening methodology to identify novel small molecules that might succeed in that purpose. This effort rendered three promising candidates: Syunclean-D, ZPD-2, and ZPDm. Their anti-amyloidogenic and neuroprotective properties were characterized, and their pharmacologic potential was discussed.

Resum en català

Les Proteïnes o Regions Intrínsecament Desordenades (IDPs, IDRs) son una classe de polipèptids que són incapaços d'adoptar una estructura tridimensional definida en el seu estat natiu. La seva funció depèn de la flexibilitat estructural i de la fluctuació entre un conjunt de conformacions diferents. Aquest fet els permet interaccionar amb una gran varietat de macromolècules i regular la transmissió de senyals cel·lulars. Els canvis en els factors ambientals, les modificacions post-traduccionals i la interacció amb altres proteïnes influeixen en el seu estat estructural, induint-ne el plegament condicional i, per tant, modificant-ne la seva funció.

En aquesta tesi enfoquem la transició desordre-ordre des de dues perspectives diferenciades. D'una banda, el seu paper en les funcions fisiològiques i, per l'altra, la seva implicació en el desenvolupament de malalties.

En primer lloc, estudiem com l'estat redox afecta en la transició estructural de dues proteïnes riques en disulfur que localitzen a l'espai intermembranós del mitocondri. Tot i compartir la mateixa conformació nativa, demostrem que ambdues proteïnes segueixen vies de plegament independents. Tant la seqüència primària d'aminoàcids com l'estat conformacional inicial defineixen la via de plegament. Per una banda, COX17 es troba desordenada en l'estat reduït i el tancament de la regió pròxima al *loop* estructural dirigeix el plegament. Per l'altra, l'estat reduït de TRIAP1 es comparable a un *molten globule*. Aquesta conformació empaquetada limita la reacció de plegament, però es indispensable per a que TRIAP1 pugui realitzar la seva funció cel·lular.

En segon lloc, estudiem la correlació seqüencial entre regions d'interacció i la tendència a formar agregats amiloides, en el context de les regions desordenades. Demostrem que Hsf1, una proteïna reguladora de la proteostasis, conté un domini críptic d'agregació *prion-like* en la seva regió desordenada del C-terminal. Efectivament, aquest domini agrega i forma fibres amiloides.

Finalment, ens centrem en l'agregació de la proteïna α-synuclein, responsable de la malaltia del Parkinson. El desordre estructural d'aquesta proteïna impedeix l'ús d'estratègies racionals que permetin dissenyar molècules que inhibeixin la seva agregació. En aquest context, implementem un cribratge massiu com a metodologia per identificar noves molècules que puguin ser efectives. Mitjançant aquesta tècnica, hem

identificat tres candidats amb activitat inhibidora: Syunclean-D, ZPD-2, and ZPDm. En aquesta tesi, mostrem les seves propietats anti-amiloides i neuroprotectores, al mateix temps que discutim el seu potencial farmacològic.

List of abbreviations

Alanine	Ala
1-anilinonaphtalene 8-sulfonate	ANS
Aggregation Prone Region	APR
Congo Red	CR
Carboxyl Terminus	Cter
Cysteine	Cys
Dalton	Da
Disulfide-Rich Domain	DRD
Dithiothreitol	DTT
Fourier-Transform Infrared Spectroscopy	FTIR
Oxidized/Reduced Glutathione	GSSG/GSH
Intrinsically Disordered Protein	IDP
Inner Membrane of the Mitochondria	IMM
Intermembrane Space of the Mitochondria	IMS
Lewi Body	LB
Lewi Neurite	LN
Time-of-Flight Matrix-Assisted Laser Desorption/Ionization	MALDI-TOF
Mass Spectrometry	MS
Nuclear Magnetic Ressonance	NMR
Amino Terminus	Nter
Outer Membrane of the Mitochondria	OMM
Parkinson's Disease	PD
Radius of Gyration	Rg
Reverse Phase - High-Performance Liquid Chromatography	RP-HPLC

Small-Angle X-Ray Scattering	SAXS
Size Exclusion Chromatography	SEC
Serine	Ser
Transmission Electron Microscopy	TEM
Thioflavine-T	Th-T
Translocase of the Inner Membrane	TIM
Translocase of the Outer Membrane	TOM
Triptophane	Trp
Alpha-synuclein	α-syn

TABLE OF CONTENTS

CHAPTER 1

Introduction

Introduction

The foundations of the structure-function paradigm

Understanding how proteins fold and how protein structures translate into the diverse range of physiological functions exerted by these macromolecules is a central topic of biology.

In 1958 and 1960, John Kendrew and Max Perutz solved myoglobin[1] and hemoglobin[2] structure in two independent studies using X-ray crystallography, becoming the first descriptions of how proteins arrange their linear strings of amino acids in the three-dimensional space and connecting functions to protein architectures. For a long time, proteins were conceived as symmetric and semi-crystalline blocks. However, myoglobin and hemoglobin illustrated how the nature of protein structures relies on significantly asymmetric conformations with a characteristic backbone bending, stabilized by a series of weak, but very abundant and cooperative non-covalent interactions. Indeed, previous studies on protein denaturation already demonstrated that protein function's inactivation correlated with unfolding events rather than with the lysis of protein chains into monomeric free amino acids[3,4]. Further publications of high-resolution structures for the hen egg-white lysozyme[5], α-chymotrypsin[6], ribonuclease-S[7], carboxypeptidase[8], and staphylococcus aureus nuclease[9] provided concluding evidence supporting that protein functions derive from their characteristic three-dimensional structures, referred to as the native state.

In 1961, Christian B. Anfinsen and coworkers demonstrated that the information required to attain the native state is indeed encoded in the primary amino acid sequence. In a series of elegant experiments, they reported how reduced and unfolded RNAse could refold spontaneously, oxidize its eight cysteines and form the subsequent four native disulfides[10]. Accordingly, RNAse recovered almost 100% of the catalytic activity. Since the random pairing of 8 cysteines raises a total of 105 different disulfide bond possibilities, the RNAse chain should adopt preferentially native-like conformations that facilitate the oxidation of the correct disulfides when they are close in the spaces. Based on these discoveries, Anfinsen hypothesized that unfolded protein chains fold through an energetically driven reaction and that the native state corresponds to the thermodynamically most stable conformation[11].

The early works Kendrew, Perutz, and Anfinsen settled the bases for so called structure-function paradigm. Essentially, it states that each particular polypeptidic sequence folds into a unique specific three-dimensional structure – the native state –, which carries out the protein function. Under this premise, structural contexts and deviations from native states has been exploited to rationalize proteins' mechanisms of action in health and malfunction in disease. Not surprisingly, the number of reported structures in the Protein Data Bank[12], the virtual reservoir of three-dimensional protein coordinates, has been growing exponentially and comprises more than 170.000 accumulated entries by the end of 2020.

Disorder in protein structures

While attaining a defined conformation becomes essential for globular proteins – such as enzymes, transporters or structural proteins –, several lines of evidence suggested that a significant number of proteins and polypeptide segments were incapable of folding into defined conformations, yet they seemed to be functional[13,14]. Intrinsically Disordered Proteins (IDPs) and regions (IDRs) are a growing class of polypeptides whose native states are composed of an ensemble of multiple and dynamic conformations that are in a permanent interconversion[15,16].

Like globular folds, the structural disorder is also imprinted in the primary sequence of amino acids and determined by their individual physicochemical properties[17]. Compositional analysis of IDP sequences in proteomes showed that they are depleted in hydrophobic (Isoleucine, Leucine, and Valine) and aromatic amino acids (Tryptophan, Tyrosine, and Phenylalanine), while being enriched in prolines, polar (Serine and Glutamine) and ionizable residues (Arginine, Aspartic acid, Glutamic acid, and Lysine) (Fig 1).
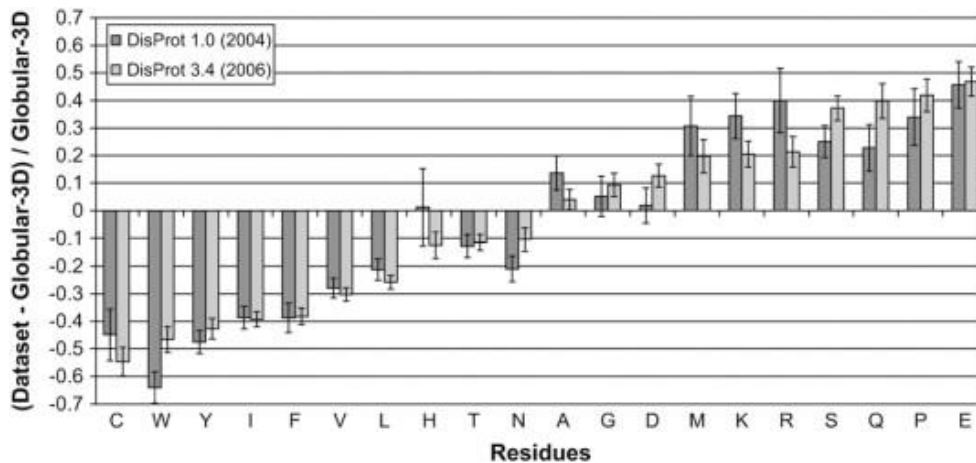
Figure 1. Amino acid frequencies in disordered proteins, relative to globular protein structures. Reproduced from Ref. 14.

This compositional bias provides disordered segments with an overall low mean hydrophobicity and a relatively high net charge, both features acting as opposing forces to structural packing and folding[14]. First, packing of sequentially distant hydrophobic stretches constitutes the so-called hydrophobic effect, an early and strong driver of protein folding, in which hydrophobic residues cluster together at high density within globular protein cores, avoiding energetically unfavorable water solvation. Second, unbalanced charges of ionizable residues promote electrostatic repulsion, precluding backbone packing, and long-distance interactions. In addition to these effects, prolines constitute the standard regular secondary structure breaker and are often usually located in the terminal regions of helices in globular proteins or in beta-turns, delimiting the secondary structure's extension. Accordingly, proline-rich sequences tend to be disordered[18], except when they are clustered together forming polyproline helices[19,20]. Although IDPs are usually seen as fully unstructured chains, they comprise different levels of compaction and indeed they are not complete free of conformational restrictions[21]. IDPs can be found as extended and highly dynamic conformations, as partially packed pre-molten globules that contain transient secondary structure elements, and non-native collapsed molten globules with secondary structure elements[22,23] (Fig. 2). The net charge, clustered hydrophobicity, and secondary structure propensity determine their specific degree of compaction, although all conformations entail significant dynamism. Disordered and ordered regions can coexist in the same polypeptide, which prompts to an additional level of structural heterogeneity and functional modulation[24].
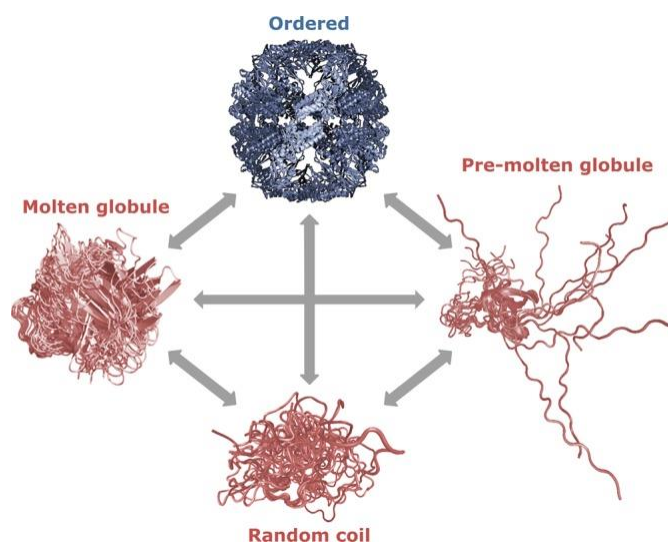
Figure 2: Different conformational states of functional proteins. Ordered, molten globule, pre-molten globule, and random coils illustrate protein conformations' versatility, ranging from well-defined and endurable structures to poorly defined and highly dynamic states. Reproduced from ref. 22.

The particular amino acid bias of IPDs has been exploited to develop a series of algorithms aimed to predict disordered regions in protein sequences and complete proteomes[25,26]. Rather than being anecdotic, as initially thought, protein disorder is a ubiquitous structural phenomenon widespread in all organisms[27], being more prevalent as the complexity of the proteome increases[13]. For instance, eukaryotes possess an average protein disorder content of more than 32%, while in prokaryotes it is below 27%[28]. IDPs' natural functions can explain, at least in part, this unbalanced distribution across different proteomes, since a significant fraction of these polypeptides act as regulators of protein function and as central components of protein interaction networks[29]. IDPs take advantage of their inherent plasticity to establish interactions with multiple substrates, accommodating their labile structures to the binding interface's topology[30,31] (Fig. 3). As well, they may act as common binding motifs for different proteins. Their structural promiscuity endorses IDP interactions with low affinity but fast interaction and disassociation rates, which allow the rapid transduction of signals[32].

Overall, IDPs complement the spectra of functions that can be performed by rigid structures and provide new ways of modulating protein activity.

Disorder-to-order transitions

Protein structural disorder relies on a series of physicochemical properties of the constituent amino acids that preclude the compaction and folding of the protein. As a consequence of the lack of structural protection, these residues are highly exposed to solvent and therefore highly sensitive to the environmental conditions[33,34]. Changes in an IDP surrounding can shift its biophysical properties and induce conformational transitions, an effect that can be also exerted by the presence of defined interaction interfaces or post-translational modifications[35,36].

Deviation from the physiological pH unbalances the protonated/deprotonated equilibrium of charged residues and might contribute to the packing of specific IDRs since the electrostatic repulsion is no longer effective[37]. A similar effect can be exerted by changes in the medium ionic strength, since it shields/unshields IDPs ionizable amino acids, impacting the effective global or local protein net charge[38]. Protein-protein interfaces can induce the transient folding of the disordered interacting segment into structures such as a helix or an extended strand, in a process known as folding-upon-binding, in which a recognition and a binding step take place[39]. Generally, it seems that binding precedes IDPs folding; however, the formation helices may precede the interaction. Short IDPs bind following apparent two-state reactions, with conformational selection and induced fit being two extreme mechanisms behind their protein-protein interactions. In contrast, longer proteins use to exhibit far more complex, multi-step binding reactions[40].
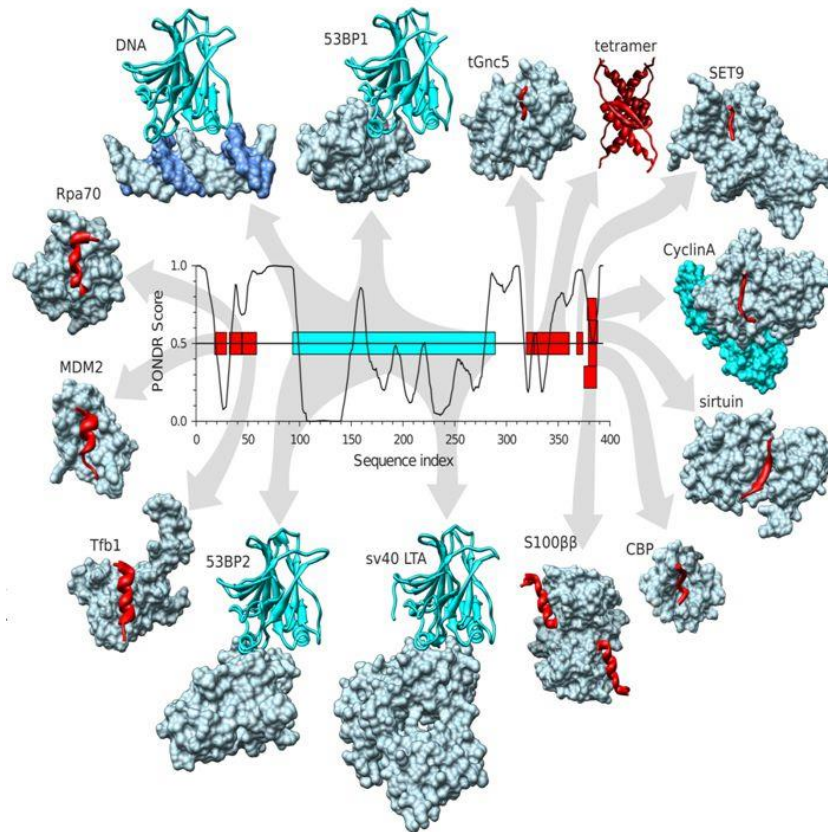
Figure 3: Interactome of p53. Prediction of disordered regions with PONDR algorithm. Adapted from ref. 16.

Cysteines are classically considered order-promoter amino acids, because in globular folds they are often engaged in disulfide bonds, which are often protected from the solvent in the well-structured rigid core of the protein, as such Cys are strongly underrepresented in IDPs. However, the order promoting state is associated to the oxidized form of the residue in disulfides, whereas reduced Cys exhibit properties that are closer to those of serine, a disorder promoting residues. This dual character of Cys is exploited by several proteins for redox-dependent conformational transitions, where massive structural rearrangements can occur[41,42]. In the first section of the thesis, we study how disulfide bonds regulate the conditional folding of mitochondrial proteins.

Remarkably, conformational transitions can cause the opposite effect and promote unfolding and the gain of structural flexibility[43]. In globular proteins, even partial and transient unfolding can lead to severe diseases[44]; in contrast, conditional unfolding has evolved as a way to regulate the function of a specific subset of proteins[45].

IDPs and amyloid aggregation

In addition to structural transitions towards locally ordered motifs, IDPs conformational changes can be associated with the formation of macromolecular amyloid assemblies[46,47]. Amyloids are a type of remarkably ordered protein aggregates that were initially discovered as the origin of several degenerative human diseases[48]. However, it is now clear that their particular physical properties and their self-assembly capability are exploited for functional purposes by a wide range of organisms, from bacteria to humans[44]. Structurally, they are characterized by a fibrillar and unbranched nature, with a regular β-sheet stacking of individual proteins along the fibril, where the β-strands run perpendicularly to the fibril axis to yield a common cross-β fold[49] (Fig. 4). Amyloids result from a self-templating reaction, in which monomer units are incorporated autonomously to the fibril ends. However, the process is complicated, and, very often, the reaction responds to a nucleation-dependent polymerization mechanism [50], which comprises three different phases. The first step is the nucleation phase, during which monomers associate to form the nuclei from which aggregation is propagated. This stage is thermodynamically disfavored and characterized by a lag phase and a slow increase of the aggregates signal. This stage depends on the protein concentration since it determines monomers' probability of interacting[51]. The nucleation phase can be shortened or even abrogated by the addition of pre-formed nuclei or seeds, in a process known as seeding. Once the critical concentration is reached, nucleation becomes favorable, and the polymerization reaction proceeds spontaneously, leading to the elongation phase. Monomers are rapidly incorporated into the nuclei edges in this stage, building-up protofibrils in an exponential reaction, which will later assemble to form amyloid fibrils. As the monomer concentration in the solution decreases, polymerization's velocity slows down, reaching the stationary phase, in which fibrils are remodeled to acquire their mature structure. The few monomers that remain in the solution are in a continuous exchange equilibrium between the assembled and the disassembled states[52], leading to molecular recycling in the fibrils.

It appears that not all the protein sequence is equally relevant for amyloid assembly and the molecular determinants that govern this process are often concentrated in the form of short linear regions of low net charge and high hydrophobicity, known as Aggregation Prone Regions (APR) or "hot spots"[53–55]. APRs establish non-native inter-molecular contacts and oligomerize to nucleate the formation of amyloids ultimately. Because they

establish the initial contacts, these non-polar regions are disposed inside the mature fibrils' core, in many cases forming a steric zipper that sustains the overall structure[56].
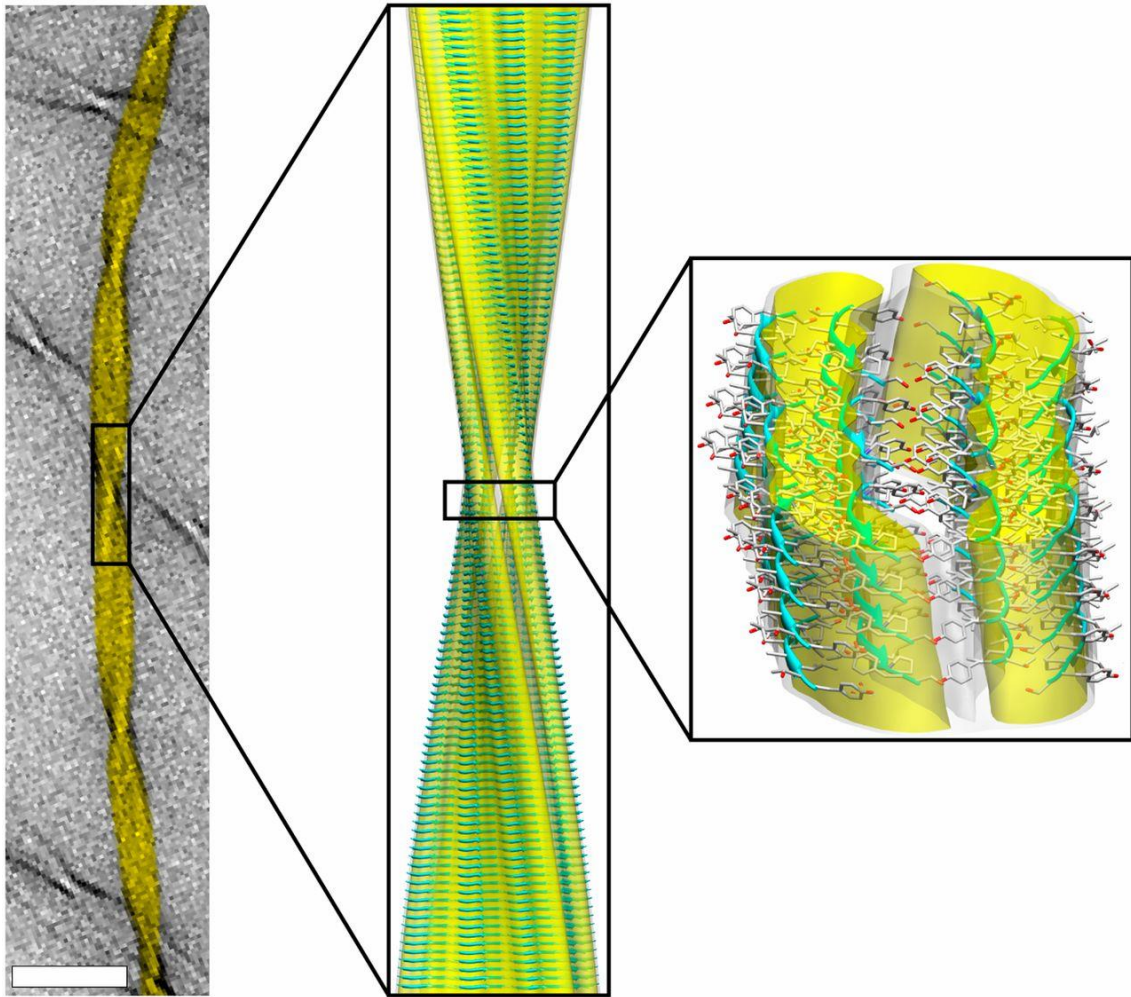
Figure 4: Three-dimensional model of an amyloid fibril captured by cryo-EM. Reproduced from Ref. 56.

While globular proteins hide APR at the hydrophobic core or mask them in regular secondary elements, IDPs' dynamic conformation permanently exposes them to spurious interactions. Therefore, globular proteins require destabilization and major misfolding events to progress towards amyloid formation[57], whereas in IDPs, the transition to this supramolecular state can occur spontaneously. It is not surprising that many amyloid-associated proteinopathies are ultimately caused by an IDP[46].

Prions and prion-like proteins are a particular subset of proteins exhibiting extended IDRs, which are essential for their biological function. They constitute a special type of amyloids that can propagate their amyloid conformation to soluble counterparts in other cells or organisms[58,59] (Fig. 5A). The term prion was coined by Prof. Stanley B. Prusiner, who discovered the amyloid state of the prion protein (PrP) as the causative agent of

diverse spongiform neurodegenerative diseases[60]. Despite no other cross-organism transmissible prion has been discovered so far, an increasing number of proteins are being shown to share a prion-like templating mechanism[61]. In yeast, prion-like proteins are involved in inheritance and survival, working as an adaptive response to environmental changes[62] (Fig. 5B). These proteins are composed of one or more globular domains, often involved in RNA/DNA and a long IDR of low complexity. This last segment is both necessary and sufficient to promote the prion behavior and is known as Prion Forming Domain (PFD). PFDs are compositionally biased, being enriched in polar and uncharged amino acids (Gln, Asn, Tyr, Ser, and Gly) and depleted in hydrophobic residues[63,64]. Since the description of the first prion-like proteins in yeast, several bioinformatics tools exploited their particular sequential traits to identify PFDs in other proteomes[65]. Similar regions, known as named prion-like domains (PrLD), have been found both in prokaryotes and eukaryotes[66], including mammals and humans, where prion-like proteins seem to account for almost 1 % of the total proteome[67]. As a general trend, these proteins use their dual disordered/assembled character to regulate gene transcription, splicing, and translation[68]. In the 4th chapter of the present thesis, we propose that the proteostasis guardian protein, Heat shock factor 1 (HSF1), belongs to the prion-like protein class.
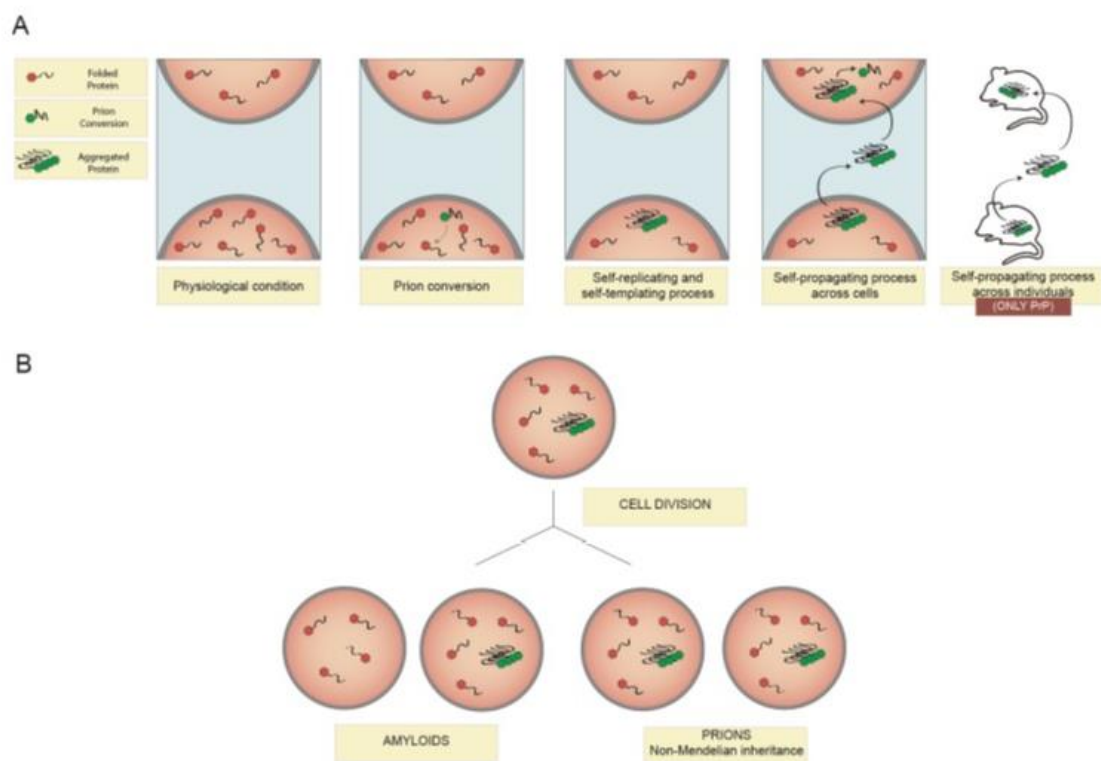
Figure 5. Prion-like amyloid behavior. A. Schematic representation of prion-like self-replication and transmissible properties. B. Prion-like mediated inheritance. Reproduced from ref. 65.

The therapeutic treatment of conformational diseases, and more precisely of amyloid-connected pathologies, like Alzheimer's or Parkinson's diseases, constitute one of the major challenges for biomedical research. Despite intensive efforts, our understanding of the particular mechanisms by which amyloids provoke cellular toxicity and tissue damage is still incomplete. It has been proposed that the intermediate oligomers formed along self-assembly reaction are the major drivers of toxicity, rather than the mature rigid fibrils[69]. The transient, dynamic, and heterogeneous nature of these assemblies constitutes the first bottleneck for pharmacologic entities' design against these proteins. Besides, the disordered nature of many amyloidogenic proteins precludes the use of rational structure-based strategies to target the native state and prevent the subsequent conformational transition. In the 5th chapter, we focus on the ambitious project of inhibiting the aggregation of the disordered protein α-synuclein with small chemical molecules. The identified chemical entities might find application in the fight against Parkinson's disease and related synucleinopathies, for which, so far, only symptomatic treatments are available.

CHAPTER 2

Aims and Objectives

Aims and Objectives

This thesis's global aim is to study disorder-to-order transitions, which are their sequential and structural determinants, and how to modulate them. To do so, we employ three different protein models in which we investigate the structural transitions from disordered states to either functional globular conformations or abnormal disease-related aggregates.

We list below the specific objectives for each chapter.

Chapter 3:

· To analyze the relationship between redox state and disorder in two small disulfide-rich proteins of the Intermembrane Space of the Mitochondria: COX17, and TRIAP1.

· To characterize the conformational, thermodynamic and kinetic determinants that govern the disorder-to-order transition of COX17 and TRIAP1 proteins.

· To study the role of the chaperone and oxidoreductase MIA40 in the catalyzed disorder-to-order transition of COX17 and TRIAP1 proteins.

Chapter 4:

· To address the overlap between, low complexity, disordered and amyloidogenic regions in the yeast inducible transcription factor HSF1.

· To characterize the cryptic amyloid potential of the C-terminal disordered region of HSF1.

Chapter 5:

· To set up a high-throughput screening platform to identify inhibitors of the amyloid aggregation of the intrinsically disordered and Parkinson-associated protein α-synuclein.

· To characterize biophysically the aggregation inhibitory properties of three top-ranked compounds: Synuclean-D, ZPD-2, and ZPDm.

· To study if Synuclean-D, ZPD-2, and ZPDm molecules can rescue from amyloid-associated toxicity using a simple animal model of Parkinson's Disease.

CHAPTER 3

Conditional folding of

cysteine-rich proteins

Conditional folding of cysteine-rich proteins

Cysteines and oxidative folding

Cysteine is probably the most singular amino acid among the pool of protein building blocks. With methionine, cysteines are the only amino acids containing a sulfur atom in their side chain. However, while the tioether (R-S-R) form of methionine's sulfur has poor reactivity, the thiol (-SH) group of cysteines can react in physiological conditions and suffer several chemical modifications, such as alkylation and oxidation[70]. Thiols' chemical properties and those of their deprotonated and more reactive thiolate anions (-S-) impact protein sequences and structures features. First, they can oxidize and form covalent bonds (S-S) with other cysteine thiols, known as disulfide bonds. Disulfide bridging can occur between cysteines comprised within the same polypeptide chain or with cysteines of other interacting proteins. As a consequence, they connect non-consecutive regions of the protein string and thermodynamically stabilize native states. The more distant the regions, the highest the energetic contribution of the disulfide. Second, they function as electronegative groups; thiolates may coordinate metals, preferentially bivalent cations such as Zinc (Zn2+) and Copper (Cu2+). Therefore, cysteine clusters can store and supply metals to other protein complexes. Third, as redox-sensitive chemical groups, they represent a molecular tool to cope with oxidative stress and reactive oxygen species. Cysteines are usually identified in highly conserved sequential patterns or clusters. Their irreplaceable physicochemical properties explain this conservation, because no other residue can perform their chemical or structural role. In the particular case of small proteins, disulfide bridging represents a major structural determinant of their conformations. Usually, small proteins lack compact hydrophobic cores and stable secondary structure elements that can effectively drive the folding reaction towards a stable and unique three-dimensional architecture. Thus, the linkage provided by cys-cys covalent bonds restrains the unfolded chain into energetically favored conformations and actively participates in the folding mechanism. Therefore, oxidation funnels secondary and tertiary structure formation, and on the contrary, disulfide reduction leads to unfolding. Both processes are referred to as oxidative folding and reductive unfolding, respectively, and small proteins that require of disulfide bonds to fold are named small disulfide-rich proteins or domains.

Cysteine's spontaneous oxidation into disulfide bonds is a relatively slow and error-prone process. Oxidative folding rates depend both on cysteines spatial proximities and on the inherent chemistry of redox reactions. Regarding the spatial constraints, both the sequential distance and the unfolded polypeptide's conformation impact the positioning of correct cysteine pairs. Cysteines that are more distant in sequence, are more unlikely to coincide in space. Conformational features of the unfolded protein, such as its intrinsic flexibility, transient native conformations, or hydrophobic packing, can approach two correct cysteines. In contrast, non-native conformations formed at early-stages or along the folding process result in miss-paired disulfide bonds, whose cysteines will require an additional stage of reshuffling – the process of disulfide interchange – to find their correct partners. As well, the burial of free cysteines inside structural cores of folding intermediates restricts their accessibility. For what respects cysteine chemistry, since the deprotonated form of thiols is more reactive, their ability to form disulfide bonds depends on the ratio of thiol/thiolate species. The specific pKa of a given cysteine and the solution's pH determine the equilibrium between thiols and thiolates, with lower pka values corresponding to more reactive cysteines. Electrophilic environments in protein sequences and structures, such as a compositional bias towards positively charged or polar amino acids, stabilize cysteine's thiolate forms and decrease their pKa. Finally, redox reactions require a donor and an acceptor of electrons. While cysteines stand as electron donors in oxidative folding, an acceptor must complement the chemical reaction. In the test tube, where protein folding occurs in isolation, molecular oxygen acts as a final electron acceptor. Molecular oxygen is a weak oxidant agent, which together with structural and the buffering constraints, forces the spontaneous folding of disulfide-rich proteins to follow extremely slow kinetics.

The seminal work of Anfinsen with RNase[10] and other landmark studies on folding intermediates in the 70s and 80s, evidenced that disulfide-rich proteins are very useful models to address the protein folding problem. By exploiting the slow and disulfide-limited progression towards the native state, the folding of these proteins can be monitored along time-lapses of hours. This behavior allows to overcome the speed limitations of normal fast-folding reactions and to infer folding with spectroscopic methods that otherwise can only be used to characterize fully folded or unfolded – by the addition of denaturing and reducing agents – states. Since pH acidification blocks cysteine reactivity, the folding process can be quenched by the simple addition of a strong acid, allowing to access snapshots of the folding reaction. Besides, alkylation, allows to

derivatize individual free cysteines and to identify their oxidized state and disulfide partners.

Bovine Pancreatic Trypsin Inhibitor (BPTI) and Hirudine are two canonic examples of disulfide-rich proteins whose oxidative folding goes through two extremely different pathways[71] (Fig. 6). They both contain three highly conserved disulfides, resulting in 15, 45, and 15 possibilities of species with 3, 2, and 1 disulfide, respectively, to a total of 75 possible forms. While BPTI folds through a reduced number of native-like intermediates, Hirudine folds through a chaotic ensemble of miss-paired disulfides and non-native conformations. The preferentially compact early-stage conformations of BPTI, driven by its primary sequence, promote the spatial proximity of native cysteines. In contrast, the extreme dynamism of Hirudine fails to lock the polypeptide chain in productive conformations.
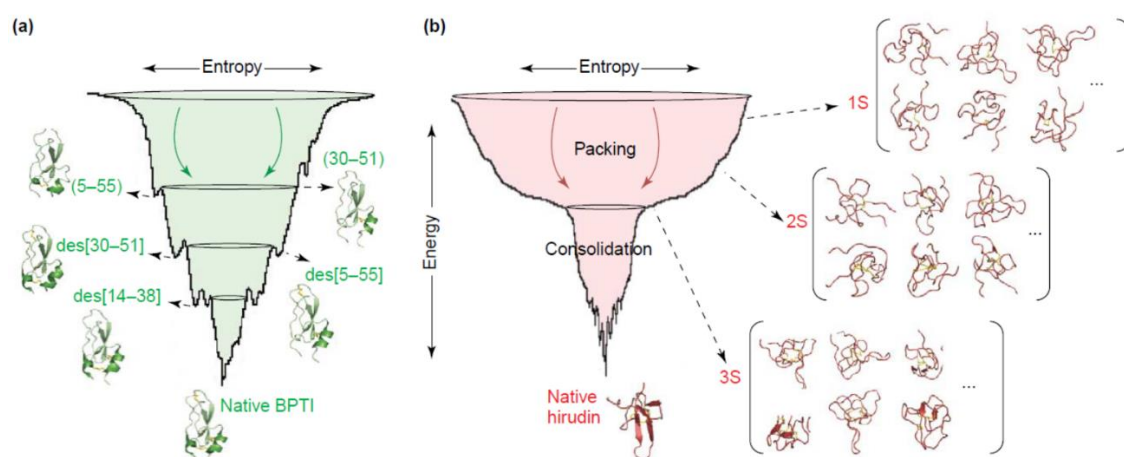
Figure 6. Schematic representation of the folding funnel of (a) BPTI and (b) Hirudin. Reproduced from ref. 71.

The study of small disulfide-rich proteins paved the way to understand the nature of folding pathways and intermediates shapes. However, from the physiological side, these slow rates of oxidative folding are far from being compatible with living-systems, which require fast transitions from unfolded to folded states. Cellular environments also disfavor oxidative folding. Most of the cellular compartments have a reductive character and contain antioxidants. For instance, for glutathione, one of the most prominent redox regulators, the reduced form (90%) overrides its oxidized counterpart (10%). With this strategy, nature can prevent spurious oxidative stress that might trigger fatal cellular damage. To cope with the duality between beneficial and harmful oxidation, in the cell, the formation of structural disulfides is catalyzed by a specific redox machinery, the thiol

oxidoreductases. In eukaryotic cells, disulfide formation takes place both in the endoplasmic reticulum, by the protein disulfide isomerase family (PDI) and Ero1p[72–74], and in the mitochondrial intermembrane space (IMS) by the oxidoreductase MIA40[75]. In prokaryotes, oxidation occurs in the periplasm by the Dsb (disulfide bond formation) family of oxidoreductases. Noteworthy, in both cases disulfide formation is compartmentalized and tightly controlled.

The physiologic function of disulfide-rich proteins is connected with the compartment where they are oxidized. In this way, most mature small disulfide proteins execute their functions in the extracellular environment – after their secretion from ER-based pathways –, in the periplasm, in the ER as membrane-anchored proteins, or the IMS. Extracellular disulfide proteins, such as inhibitors, hormones or toxins, are probably the most studied archetypes. BPTI and Hirudin belong to this class. Other examples are the mature form of insulin, ribonucleases, lysozyme, β-defensin, phospholipase A2, or epidermal growth factor. Of note, protein stabilization by disulfides stands as an essential structural strategy to resist harsh conditions. Thus, secreted small disulfide-rich proteins usually contain a higher percentage of cysteines in their sequences than other cysteine-containing proteins. The strong association between disulfides and structured regions leads to consider cysteine as an order promoting amino acid. This conception is strongly influenced by the fact that oxidized cysteines in disulfide bridges contribute to structure stabilization. Nevertheless, free thiols' physicochemical properties are more similar to that of a polar residue than to a bulky and hydrophobic side chain. Thus, the behavior of reduced thiols is expected to resemble that of serine hydroxyl (-OH), one of the strongest disorder promoting residues. Indeed, cysteine substitutions to serine are usually used in mutagenesis studies to knock out disulfides from structures. Remarkably, the bioinformatics analysis of a set of well-described representative disulfide-rich proteins reported that, when reduced, their inherent features resemble more to intrinsically disordered proteins than to globular domains[76].

To summarize, the inherent properties of disulfide-rich proteins are unavoidably determined by their function and cellular context. Synthesized as disordered polypeptides, they rely on the catalyzed formation of disulfide bonds to fold into functional forms. Their oxidation is compartmentalized and associated with their destiny. They are usually depleted in hydrophobic residues to prevent their aggregation before folding effectively occurs. Since disulfide bridges are reversible and can unlink attached regions after their reduction, they might mediate bidirectional folding/unfolding in response to particular

redox environments. Several examples of redox-sensitive, transiently disordered, proteins have been reported[42]. This conditional folding behavior is best illustrated by the small disulfide-rich proteins residing in the IMS, as discussed in the following section.

Oxidation in the intermembrane space of the mitochondria.

Constituting the antechamber of the mitochondrial matrix, the Intermembrane Space of the Mitochondria (IMS) is an extremely dynamic subcellular compartment that regulates the crosstalk between the mitochondria and the cytosol[77]. Despite its inherent function as a transitory space for a variety of ions, small molecules, and macromolecules; a growing number of cysteine-rich proteins have been reported to populate the IMS, where they function in essential processes such as assisting in the assembly of the respiratory chain, controlling apoptosis and in proteostasis by guiding protein folding and traffic[78].

IMS targeting proteins are coded in the genomic DNA and synthesized in the cytosolic ribosomes, and thus they must be imported through the Outer Membrane of the Mitochondria (OMM). Disulfide rich proteins lack the standard cleavable N-termini targeting sequences that direct most of the organelle's proteins[79]. Instead, their import relies on the disulfide relay system[80,81]. Disulfide rich proteins are recruited to the IMS as reduced polypeptides and recognized by the chaperone and oxidoreductase MIA40, which facilitates the formation of critical structural disulfides and their subsequent folding into functional conformations[82,83]. MIA40 substrates are usually small proteins carrying CX3C or CX9C twin motifs that, upon oxidation, form two disulfide bonds that stabilize an antiparallel helix-loop-helix fold[84,85]. They cannot be re-released to the cytosol as folded species, fueling a vectorial energy-free import of new reduced polypeptides[86–88].

Interestingly, MIA40 also contains itself a twin CX9C motif and shares its substrates' fold[89]. It is suggested that active Mia40 also triggers the oxidation of newly imported molecules of its reduced version[90,91]. MIA40 is present in almost all eukaryotes, and the folded domain is highly conserved, being cysteines the most conserved amino acids along the entire sequence, as expected. In stark contrast to PDI and Dsb oxidoreductase families, which contain several copies of the canonic thioredoxin catalytic motif (CXXC), MIA40 is a thioredoxin-unrelated oxidoreductase. Its singular catalytic motif (CPC) juxtaposes to the helix-loop-helix fold; in a flexible arm that facilitates substrate oxidation. In the

active form, the two cysteines of the catalytic motif are oxidized, forming a disulfide bond. MIA40 recognizes its substrates through hydrophobic interactions[92]. The helix-loop-helix structure contains a hydrophobic groove that accommodates the unfolded polypeptide of the substrate and aligns a reactive cysteine with the catalytic domain[89] (Fig. 7). Regarding the MIA40 substrate, the reactive cysteine vicinity encompasses a cryptic hydrophobicity, which allows the binding to MIA40's hydrophobic cavity and would be critical for the formation of a mixed intermolecular disulfide[93]. This signal is referred to as the Mitochondrial Intermembrane Sorting Signal (MISS) or the IMS targeting signal (ITS) and is highly conserved[94,95]. It is located up or downstream of the reactive cysteine and comprises two hydrophobic residues, positions ±3 and ±4, and an aromatic residue in position ±7 of the cysteine [Cys-X-X-Hy-Hy-X-X-Ar]. The IMS/MISS sequence usually forms the end part of one of the helices in the folded state, positioning the hydrophobic residues towards the helices' interface and the structural core.
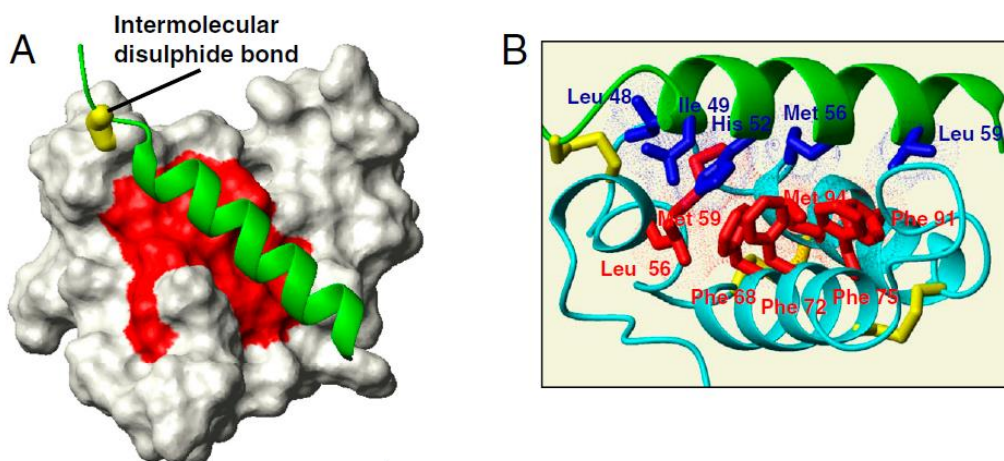
Figure 7. Solution structure of the MIA40-COX17 Interaction. A) COX17 induced helix (green), Mia40 residues involved in complex formation (red) and disulfide bridge (yellow). B) hydrophobic residues of COX17 (blue) and MIA40 (red) involved in the interaction Mia40-cox17. Adapted from ref. 92.

Subsequent to the binding, a reactive cysteine of the disulfide-rich protein performs a nucleophilic attack to the second cysteine of the oxidized CPC motif and forms an intermolecular disulfide bond[92,96]. The mixed specie MIA40-substrate is then resolved by a new nucleophilic attack from another substrate's cysteine, releasing a partially oxidized product with a native disulfide. The partially folded disulfide-rich protein either proceeds autonomously to the native state or enters a new round of MIA40 catalysis to oxidize the second disulfide. As a result of the redox reaction, Mia40's catalytic site becomes reduced

and should be recycled by a series of coupled redox reactions that conduct the electrons to the respiratory chain, via cytochrome C and ultimately, to molecular oxygen. Overall, the MIA40 pathway is an energy-free strategy to form disulfides de novo, a self-recycled circuit that produces water as a dead-end[97] (Fig. 8).
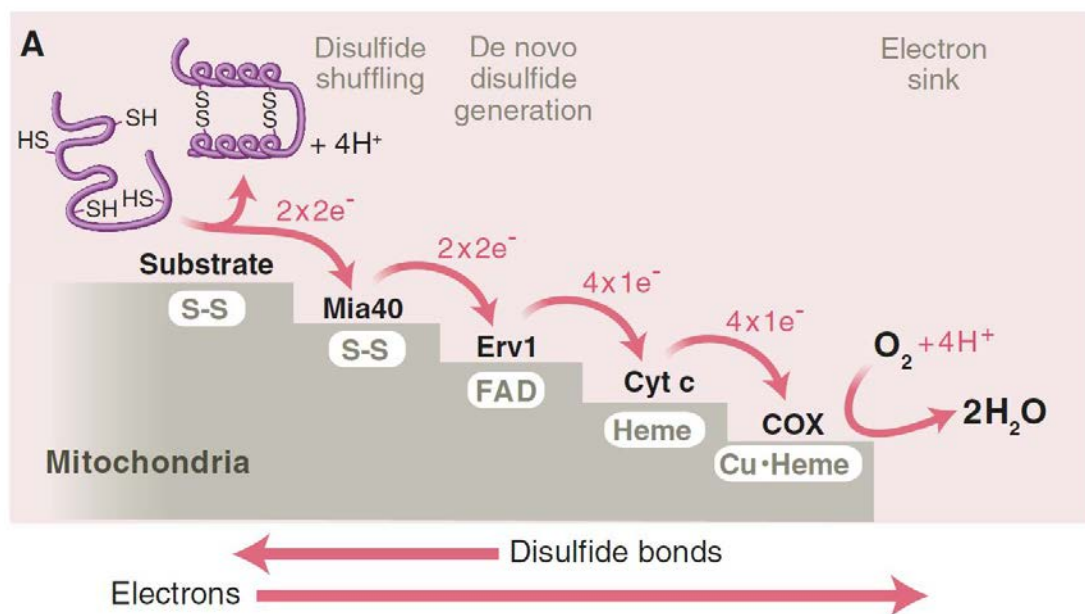
Figure 8. Schematic representation of the IMS cycle redox reactions for *de novo* disulfide formation. Reproduced from ref. 97.

MIA40 substrates are reduced and unfolded in the cytosol, featuring as intrinsically disordered proteins[98]. This behavior is consistent with the generic conformational state of disulfide-rich proteins in their respective reduced states[76]. Therefore, MIA40 substrates populate two different conformational states in the cell depending on the redox state of the protein and the cellular compartment; a reduced and significantly unfolded state, characteristic of the cytosol, and a well-defined and oxidized conformation, found in the IMS. Despite the physiologic function of disulfide-rich proteins is attributed to the oxidized form, the population of reduced and unfolded form is a requirement for the efficient disposal and import to the IM. Because they fold in the mitochondria, Mia40 substrates are exposed to cytosol's crowded milieu where spurious interactions or degradation events may occur; thus, they need to maximize their solubility[99], while remaining reduced, since oxidation of disulfide-rich proteins in the cytosol decreases their import rates[100]. Recent evidence shows that active MIA40 localizes close to the TOM exit, where it binds their substrates as they start translocating[101]. The current hypothesis

states that MIA40 functions as a pulling force towards the IMS instead of acting as a simple folding trap[102]. Therefore, partial oxidation in the cytosol hampers the action of MIA40[100]. Importantly, in the cytosol, non-translocated mitochondrial precursors and miss targeted species are degraded by the proteasome machinery[103]. The abnormal accumulation of miss targeted proteins perturbs homeostasis, inducing cell stress[104,105] and accordingly, the ubiquitin control system acts as a negative regulator of the disulfide relay system by degrading the excess of precursors in physiologic conditions[106]. Remarkably, the accumulation of non-productive MIA40-substrates isoforms can be proofread by glutathione, resulting in the formation of unfolded species inside the IMS[86,107]. Changes in mitochondria's redox environment might induce an unbalance between the conformational states of disulfide-rich proteins. In addition, reduced, partially folded, or misfolded species might undergo retrotranslocation to the cytosol, where they are degraded[108].

In this scenario, understanding the connection between the reduced and folded states of IMS disulfide-rich proteins will reveal crucial information on the functioning of the MIA40 import pathway. This is not only important in health, but also in disease[78]; since several mutations in MIA40 substrates are associated with human disorders, some of them impacting their fold and cellular distribution, rather than their function[105,109].

In this section we go back to oxidative folding techniques and apply them to study two CX$_9$C substrates of MIA40; COX17 and TRIAP1. They allowed us to characterize the reduced states of both proteins and study the structural determinants that guide their folding towards the attainment of folded and functional structures, both spontaneously and in the presence of MIA40. We demonstrate that despite sharing a common global fold, COX17 and TRIAP1 folding reactions are determined by completely different structural constraints. Our results shed new light on the function of MIA40 and on the nature of IMS disulfide-rich proteins.

Work 1

Disulfide driven folding for a conditionally

disordered protein

Disulfide driven folding for a conditionally disordered protein

Hugo Fraga[1,2,3‡], Jordi Pujols[1,2‡], Marcos Gil-Garcia[1,2], Alicia Roque[2], Ganeko Bernardo-Seisdedos[4], Carlo Santambogrio[5], Joan-Josep Bech-Serra[6], Francesc Canals[6], Pau Bernadó[7], Rita Grandori[5], Oscar Millet[4] and Salvador Ventura [1,2*]

[1] Institut de Biotecnologia i Biomedicina. Universitat Autònoma de Barcelona, 08193-Bellaterra, Spain.

[2] Departament de Bioquímica i Biologia Molecular. Universitat Autònoma de Barcelona, 08193-Bellaterra, Spain.

[3] Departamento de Bioquimica, Faculdade de Medicina da Universidade do Porto, Portugal

[4] Protein Stability and Inherited Diseases Laboratory, CIC bioGUNE, 48160-Derio, Spain.

[5] Department of Biotechnology and Biosciences, University of Milano-Bicocca, Italy

[6] Vall d'Hebron Institute of Oncology (VHIO), Barcelona, Spain

[7] Centre de Biochimie Structurale, INSERM-U1054, CNRS UMR-5048, Université de Montpellier. 29, rue de Navacelles, 34090-Montpellier, France

[‡] These authors contributed equally.

[*] Correspondence: salvador.ventura@uab.es; Tel.: (+34) 93 586 8956

Abstract

Conditionally disordered proteins are either ordered or disordered depending on the environmental context. The substrates of the mitochondrial intermembrane space (IMS) oxidoreductase Mia40 are synthesized on cytosolic ribosomes and diffuse as intrinsically disordered proteins to the IMS, where they fold into their functional conformations; behaving thus as conditionally disordered proteins. It is not clear how the sequences of these polypeptides encode at the same time for their ability to adopt a folded structure and to remain unfolded. Here we characterize the disorder-to-order transition of a Mia40 substrate, the human small copper chaperone Cox17. Using an integrated real-time approach, including chromatography, fluorescence, CD, FTIR, SAXS, NMR, and MS analysis, we demonstrate that in this mitochondrial protein, the conformational switch between disordered and folded states is controlled by the formation of a single disulfide bond, both in the presence and in the absence of Mia40. We provide molecular details on how the folding of a conditionally disordered protein is tightly regulated in time and space, in such a way that the same sequence is competent for protein translocation and activity.

Keywords: oxidative protein folding, conditional disorder, disulfide bonds, mitochondrial intermembrane space, hCox17

Introduction

Deciphering how an unstructured polypeptide folds into its biologically active three-dimensional structure has been one of the major challenges in biochemistry over the last 50 years[1]. A first problem in folding studies is that the biological unfolded state cannot be directly accessed, but only approximated, i.e. using strongly denaturing conditions[2]. A second limitation is that due the high speed at which proteins generally fold, most experiments measure the rate of appearance or disappearance of a specific signature for a particular species, typically on millisecond-second scales[3,4]. Spectroscopic probes like fluorescence[5] and circular dichroism (CD)[6] are usually used in these assays, since they display enough intrinsic time resolution to follow folding reactions. These data provide important information on the height of energetic barriers, but say little about how protein conformation evolves along the pathway. Accordingly, alternative techniques like Nuclear Magnetic Resonance (NMR)[7], infrared spectroscopy (FTIR)[8], small-angle X-ray scattering (SAXS)[9] or native mass spectrometry (MS)[10] have been implemented to track the folding of certain proteins, providing novel and wealthy structural information in these specific cases. Because no single method can provide an entire picture of the folding reaction, the integration of all those complementary approaches to characterize the complete folding trajectory of a given polypeptide is a long pursued objective.

A large number of cysteine-containing proteins encoded in the nucleus and synthesized on cytosolic ribosomes are targeted to the intermembrane space (IMS) of mitochondria as reduced species and they become oxidized into their functional forms only after entering this organelle compartment[11]. The IMS oxidoreductase Mia40 introduces disulfide bonds into these target proteins to facilitate their folding[11,12].

The unique properties of Mia40 substrates offer an opportunity to characterize folding reactions with unprecedented conformational detail. First, a biologically relevant unfolded state can be accessed, since in their reduced state these proteins remain essentially unstructured, both in vitro and inside cells[13]. Second, non-catalyzed folding of these proteins can take several hours to occur[14] which permits tracking the process in real-time by a battery of low temporal-resolution techniques reporting on different structural properties. Third, in these proteins the adoption of the native structure and the formation of the disulfide bonds are linked in a process known as oxidative folding, allowing an

accurate characterization of the major transient intermediates that populate the folding reaction[15].

Cox17 is one of the best-characterized substrates of Mia40. It acts in the IMS as the donor of Cu(I) to both Sco1 and Cox11[16,17], being also involved in the regulation of mitochondrial contact sites and cristae organization[18]. Mutations in Cox17 result in the dysfunction of cytochrome C oxidase, leading to respiratory defects[19]. Human Cox17 (hCox17) belongs to the CX9C family and is a 62-residue protein containing six conserved Cys. The NMR solution structure of this protein shows that it consists of a coiled-coil-helix-coiled-coil-helix domain (CHCH) stabilized by two disulfide bonds involving Cys26–Cys55 and Cys36–Cys45, preceded by a flexible and completely unstructured N-terminal tail (Fig. 1A). The Cu(I) ion is coordinated by two additional free cysteines at positions 23 and 24.

The folding of hCox17 can be considered as a disorder-to-order transition mediated by a post-translational modification: the formation of structural disulfide bonds[20,21]. In the present study, we integrate real-time chromatography, disulfide, fluorescence, CD, FTIR, SAXS, MS and NMR analyses in order to provide a complete conformational description of the mechanism governing this structural switch.

Results

hCox17 is predicted as an IDP in the reduced state

We have recently proposed that many small proteins containing disulfide bonds would resemble bona fide IDPs, when in their reduced states[22]. Despite there is no perfect mimic for reduced Cys residues, mutation of disulfide-forming Cys to Ser is considered the less disruptive change to simulate disulfide bond reduction[23]. We virtually mutated Cys to Ser in hCox17 sequence to mimic the reduced state of the protein and analyzed the degree of predicted disorder using FoldIndex[24], Espritz[25], IUPRED[26], PONDR-FIT[27] and RONN[28] algorithms. With the exception of a small stretch at the hCox17 C-terminus, all these programs consistently predict this sequence as mostly disordered (Fig. 1B).
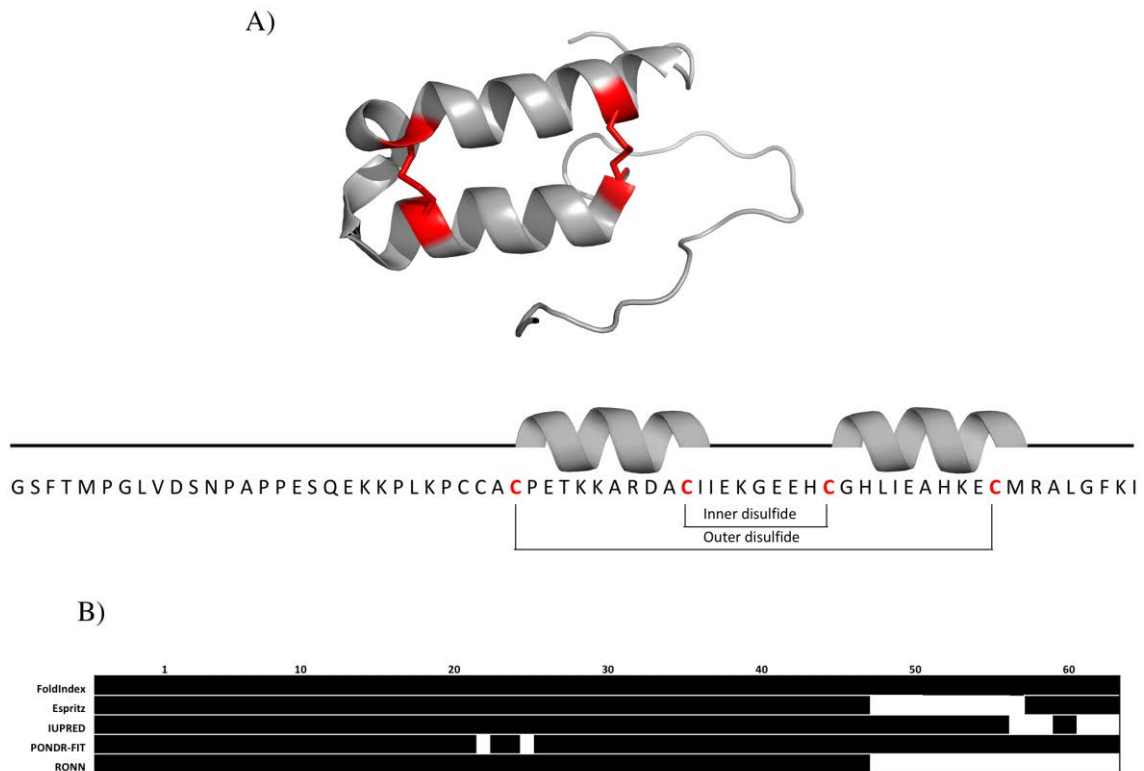
Figure 1. hCox17 structure and disorder prediction. A) Upper, 3D representation of hcox17 (pdb code: 2R9N) with red-colored cysteines and disulfides. Bottom, primary sequence and secondary structure of hCox17. B) Disorder predictions for hCox17 (C26S, C36S, C45S, C55S) sequence using Foldindex, Espritz, IUpred, Pondr-Fit and Ronn. Those amino acids above the methodology threshold and predicted as disordered are colored in white.

29

Reduced hCox17 is mostly disordered

Previous folding studies on yeast Cox17 (yCox17), which has only 35 % sequence identity with hCox17, have employed mutants where several Cys were changed to Ser[29]. In order to attain a comprehensive view of the complexity of hCox17 folding pathway, we used here instead the human WT protein, where all 6 Cys are present. For example, reducing this number to 4 Cys would decrease the potential disulfide isomers in the folding pathway from 75 (15 1S-S + 45 2S-S + 15 3S-S) to only 9 (6 1S-S + 3 2S-S).

It has been suggested that hCox17 is significantly unfolded when reduced and that this is probably pivotal for its diffusion across the mitochondrial outer membrane[30]. In order to confirm this point, hCox17 purified from *E.coli* Origami cells was reduced with 200 mM DTT for 2 hours. hCox17 reduction was clear from the shift in RP-HPLC retention time (Fig. 2A). The oxidation state of the protein was confirmed using MALDI-TOF after reaction with vinylpiridine. An increase of 637 Da (105 of vynilpiridine per free SH + 1 proton per SH) was observed when native hCox17 was reduced and treated with vinylpiridine indicating that all 6 Cys were accessible in the reduced state.

The NMR ^{1}H-^{15}N-HSQC spectrum of reduced hCox17 (Fig. 2B) is consistent with an IDP fingerprint: The amide groups of hCox17 protein are in a similar chemical environment which leads to a clustering of the 41 ^{15}N - ^{1}H distinguishable signals in the central part of the spectrum, with strong peak intensities and a characteristic up-field position for the ε-TRP signal. For folded hCox17, the chemical environments are more heterogeneous due to the presence of the two alpha-helices, thus creating a more disperse spectrum (Fig. 2B). We further investigated the structural properties of reduced hCox17 by recording the ^{1}H-^{15}N-HSQC and HNCACB spectra of the ^{15}N-^{13}C-labelled protein. Chemical shifts were assigned for Ala, Trp and Ser and their neighboring residues (Fig. S1), as they fall in isolated Cβ chemical shift regions. Reference chemical shift values for the same residues in the oxidized state were obtained from BMRB 11019 (Table S1). Protein flexibility and secondary structure was predicted by comparing the Random Coil index values from the backbone chemical shifts (HN, N, Cα and Cβ)[31] (Fig. 3). Based on the analyzed residues it can be deduced that reduction of hCox17 results in the loss of most of its secondary structure content.
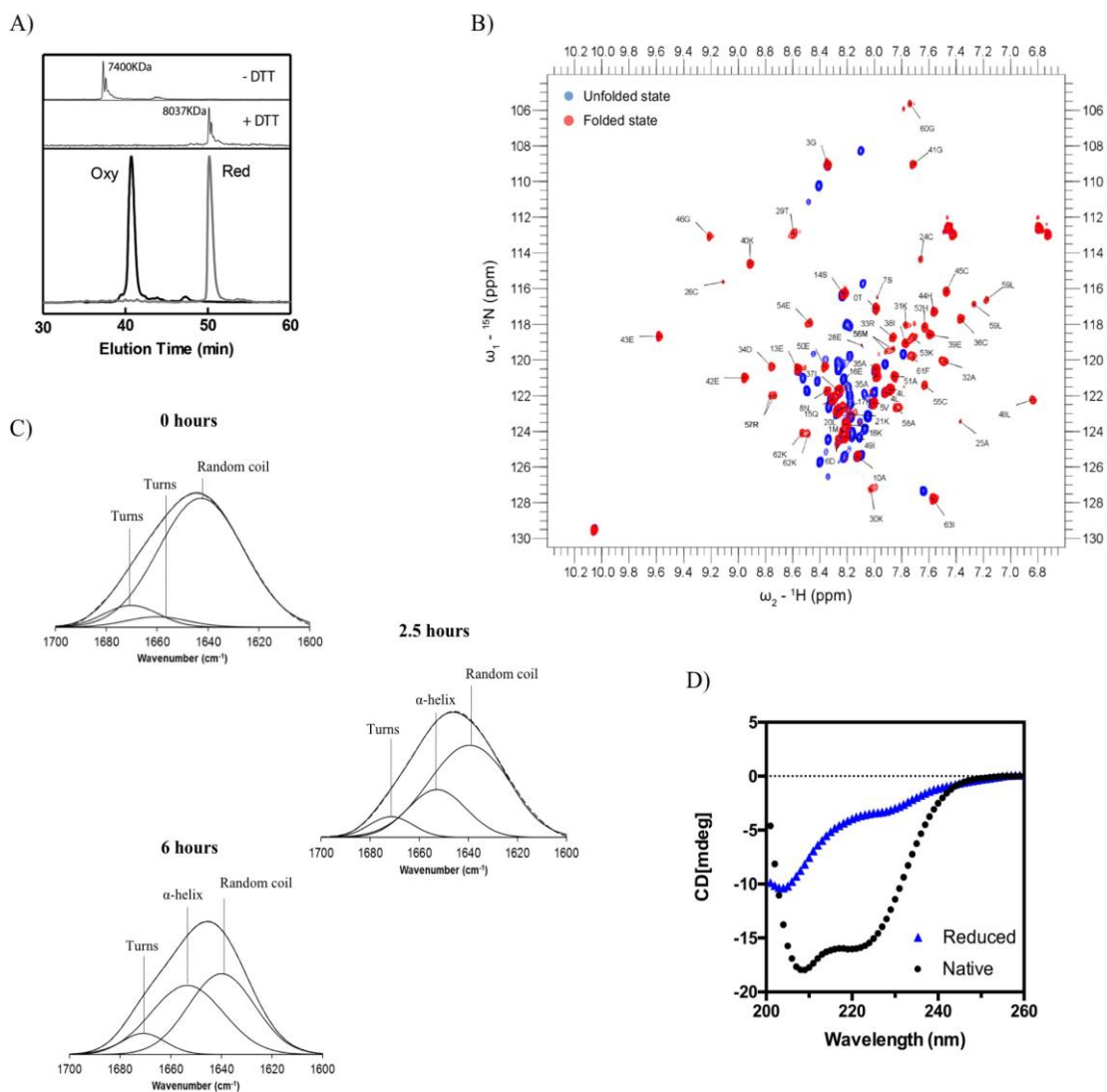
Figure 2. Conformational properties of reduced and native hCox17. A) Lower panel: HPLC chromatograms of native (black) and 2 hours DTT-treated hCox17 (grey). The upper panel shows MALDI-TOF spectra of both samples quenched with vinylpiridine. B) NMR 1H-15N-HSQC spectrum of reduced (blue) and native (red) hCox17 with the respective assignments. C) FTIR absorbance spectra of the reduced and native states of hCox17 in the amide I region. Experimental spectra and their fitting are shown in dashed and solid lines, respectively. The assignment of each structural component is indicated. D) Far-UV CD spectra of reduced (blue triangles) and native (black circles) hCox17.

These results agree with the Fourier transform infrared spectroscopy (FTIR) spectrum of DTT treated hCox17 in the amide I region, which evidences the lack of significant regular secondary structure (Fig. 2C), whereas the FTIR of spectrum of native hCox17 contains

a 44% α-helical content (Fig. 2C and Table S2). This percentage of α-helix is in perfect agreement with the solution structure of hCox17 (30/67 residues in α-helix = 44%).

The far-UV circular dichroism (CD) of native hCox17 is typical of an α-helix containing protein with characteristic minima at ≈210 nm and ≈222 nm (Fig. 2D). The far-UV CD spectra of reduced hCox17 displays much lower ellipticity and a minimum at 205 nm consistent with a mostly disordered structure (Fig. 2D). However, a shoulder at ≈222 nm was evident in the spectrum. To discard that this signal could be indicative of a certain residual α-helical content, we performed a thermal denaturation of reduced hCox17 and monitored the changes in α-helix content following the changes in the CD signal at 222 nm. We did not observe any change in ellipticity along the melting (Fig. S2A) and the CD spectra of the reduced protein at 25 and 95 °C were highly similar (Fig. S2B), in contrast to what happens with the oxidized protein (Fig. S2C). This lack of structural transition is in agreement with the NMR and FTIR data, supporting hCox17 being mostly disordered in the reduced state.

The disordered N-terminus of native hCox17 accounts for roughly 50 % of its sequence, this affects the hydrodynamic properties of the protein. Accordingly, native hCox17 elutes much earlier than expected for its molecular weight in a size exclusion chromatography (SEC) (apparent MW ≈14.000 Da/predicted MW = 7.400). Reduction of hCox17 further increases its hydrodynamic volume and the protein elutes earlier than the native form (apparent MW ≈ 20.500 Da), which is consistent with a larger fraction of the protein populating disordered conformations (Fig. S3).
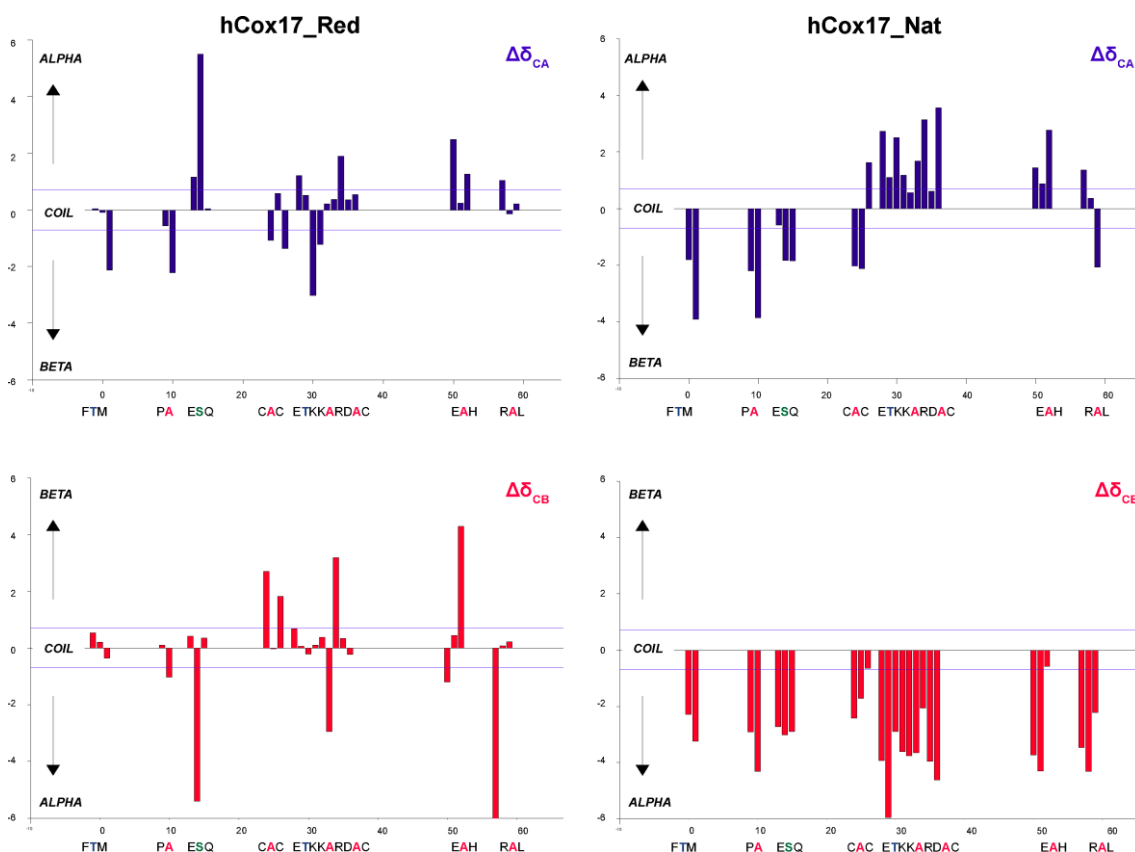
Figure 3. Random Coil Index values for Cα and Cβ chemical shifts. Values above 0.7 in Cα are indicative of α-helix whereas values below -0.7 are indicative of β-sheet and vice versa for those values of Cβ.

Cox17 folding proceeds by sequential disulfide formation

To monitor hCox17 folding, we first analyzed the progression of hCox17 secondary structure formation using CD. The reduced protein was loaded in a PD10 desalting column (*GE Healthcare*) to remove DTT, allowing the protein to re-oxidize its free thiol groups into disulfides as previously described[14]. As the reaction progresses the ellipticity of the protein increases. By plotting the changes in ellipticity at 222 nm versus time it can be observed that hCox17 folding is extremely slow, taking ≈ 16 h for completion at pH 7.0 (Fig. S4).

To further characterize hCox17 folding, we used FTIR spectroscopy coupled to RP-HPLC analysis. Time-resolved FTIR has proven to be a powerful technique to follow conformational changes during protein folding[9]. On the other hand, RP-HPLC, after quenching the folding reaction with acid, allows the resolution of the different isomers that populate the pathway[14].

hCox17 oxidative folding chromatograms are rather simple, considering the high number of potential disulfide isomers, 75, for a protein containing 6 Cys residues. Apart from the starting fully reduced form (peak R) we could only detect two major species, peak I and peak N (Fig. 4A), displaying decreasing hydrophobic surface accessible to solvent under chromatographic conditions, according to RP-HPLC retention times. Peak N corresponds to native hCox17, whereas, Peak I contained only the inner C36-C45 native disulfide (Table S3). Therefore, no off-pathway intermediates are populated significantly along the folding reaction. The pathway involves exactly the same species independently of whether the reaction is performed at pH 7.0 or at pH 8.4, where the thiols are more reactive (Cys side chain pKa = 8.37) and the reaction proceeds faster, taking only about 6 h to complete (Fig. 4A). This provides us a mean to match the reaction kinetics with the time scale of the different techniques used in this study, just by controlling the pH of the reaction. MALDI- TOF analysis of the oxidative folding mixtures excluded the formation of any sulfenic or sulfonic species.

The quantitative analysis of disulfide species along the reaction (Fig. 4B) indicates that peak I reaches a steady level before the formation of the native protein (Peak N) is complete; suggesting that the transition between the disordered and native states of hCox17 proceeds via the sequential oxidation of single disulfide bonds, in which the formation of the inner disulfide precedes the formation of the outer one. In order to confirm this point, we performed a stop/go experiment. The inner disulfide acid-trapped intermediate was isolated from oxidative folding mixtures, dried and dissolved in standard buffer at pH 8.4 to resume folding. Peak I leads directly and efficiently to the formation of peak N, without the accumulation of any other alternative species (Fig. 4C).
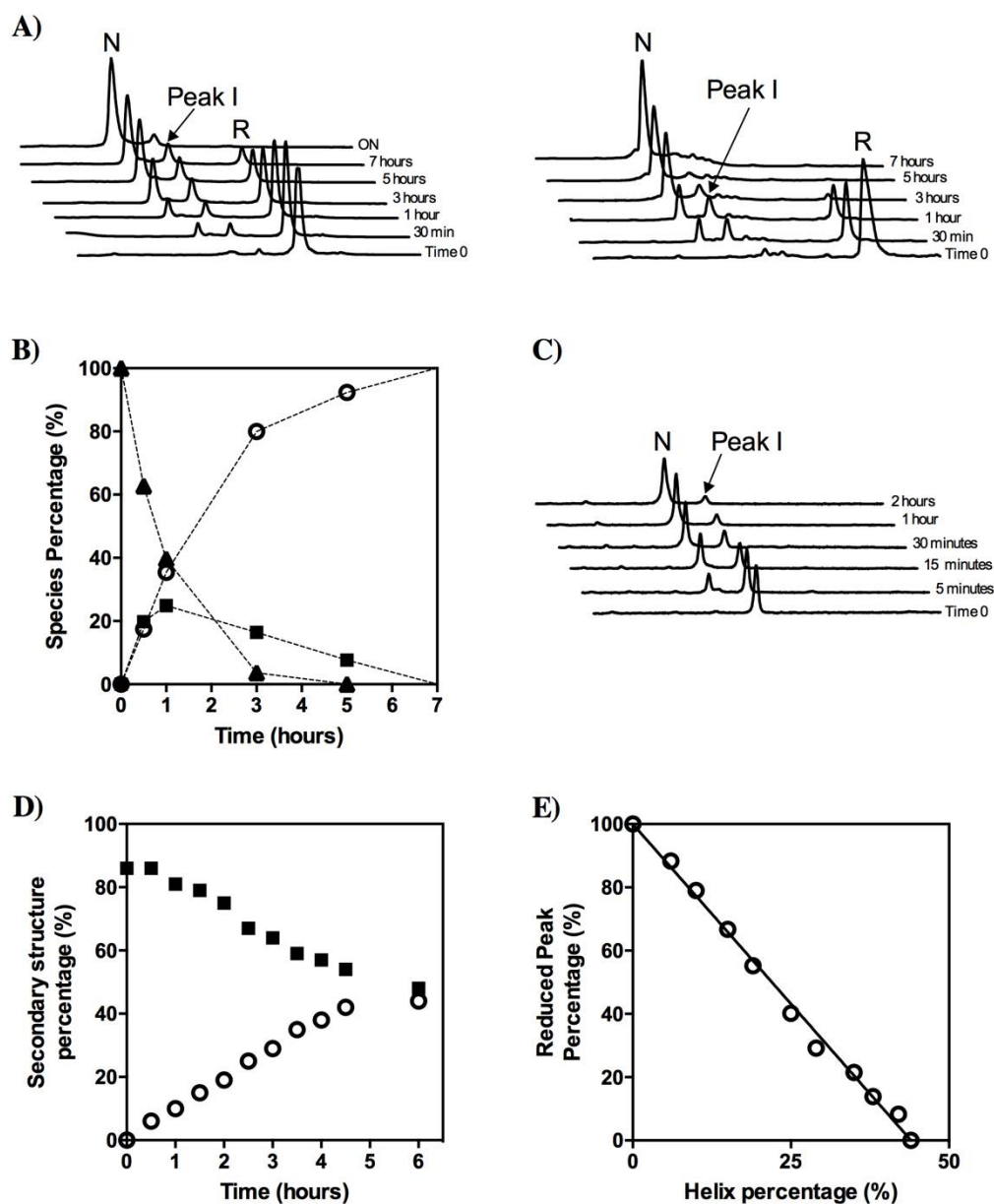
Figure 4. Inner disulfide bond species represents a folding intermediate and displays secondary structure features linked with the native conformation. A) hCox17 was allowed to re-oxidize after DTT removal at pH 7 (left) and pH 8.4 (right). At defined time points, the reaction was quenched and analyzed by HPLC. B) RP-HPLC peak areas from the refolding reaction at pH 8.4, are plotted as function of time. Reduced (solid triangles), intermediate (solid squares) and native (open circles) species. C) Acid-trapped intermediate I was purified by RP-HPLC, freeze-dried, and dissolved in Tris-HCl (pH 8.4) to reinitiate the folding reaction. Folding reaction was subsequently quenched with acid at different time points and re-analyzed by RP-HPLC. D) Reduced hCox17 was allowed to air oxidize at pH 8.4 inside a sealed FTIR spectrometer cuvette and the spectra recorded at specific refolding times. The relative contributions of random coil (solid squares) and □-helix (open circles) conformations to the global amide I spectra are

plotted as a function of time. E) Duplicates from the FTIR assay at pH 8.4 were acid-quenched and the different species present in the reaction analyzed by RP-HPLC as indicated above. The area of the reduced specie peak is plotted versus the □-helix content as deduced by FTIR.

Consistent with CD data, the analysis of the folding reaction using FTIR indicates an increase in α-helical structure as hCox17 folds. This technique allowed us to confirm that this increase in structure comes at expenses of a decrease in disorder (Fig. 4D). Analyzing the FTIR samples using RP-HPLC we find a complete correlation between the gain in secondary structure and the disappearance of the reduced protein (r2 = 0.995) (Fig. 4E), which confirms its disordered nature and indicates that the inner disulfide intermediate already contains a significant degree of α-helical structure.

Formation of the inner disulfide bond involves a transition between a disordered and a compact native-like state in hCox17

We proceeded testing whether, as suggested by SEC, hCox17 folding resulted in significant changes in protein shape. SAXS can be used for the characterization at low-resolution of protein structure and dynamics[32]. In the context of hCox17, SAXS is useful to assess the time-dependent compaction along the folding pathway. Kratky representations of the measured curves present a continuous increase of the $I(s)s^2$ with the momentum transfer indicating that reduced hCox17 remains highly disordered (Fig. 5A). Interestingly, the increase in the intensity of the maximum suggests an enhanced compaction with time. This compaction is better monitored by the P(r) function that presents more noticeable changes as folding proceeds. Both the systematic decrease of the Dmax value and the shift of the maximum towards smaller distance values demonstrate the compaction process induced by the folding of hCox17 (Fig. 5B). Figure 5C displays the Rg derived from the SAXS curves measured along the oxidative folding of hCox17 at 3 mg/mL and 2 mg/mL at pH 8.4, to check whether protein concentration can increase or disturb global compaction and shape. At initial state, hCox17 exhibits a Rg of 24.2 +/- 0.3 Å, which is in very good agreement with a fully disordered protein of 67 residues, 22.8 Å[33]. During the following two hours hCox17 experiences a fast compaction process to form a particle with an Rg of 21 Å that remains unchanged during the following four hours. Finally, Rg displays a slight decrease of 1 Å after six hours, that remains stable until the end of the assay. A RP-HPLC analysis of the folding reaction

under the same conditions than those used in the SAXS experiment (Fig. S5) evidences the presence of around 10% of intermediate at 6h, suggesting that this last decrease in Rg might reflect its conversion into native hCox17.
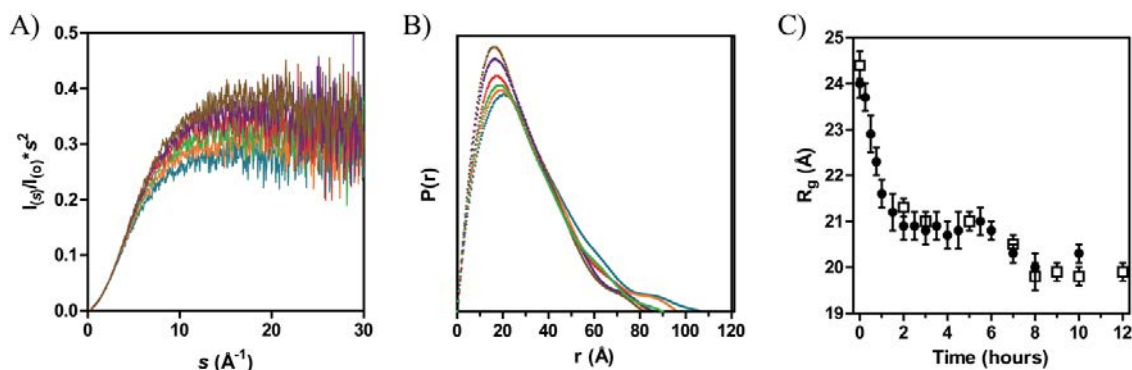

Figure 5. Small-angle X-ray scattering analysis of hCox17 compaction upon folding. A) Kratky representation at selected folding time-points: 0 hours (blue), 30 minutes (orange), 1 hour (green), 2 hours (red), 4 hours (purple) and 8 hours (brown). B) P(r) functions representing the time-points colored as in A. C) Radius of Gyration plot versus time of folding. Both 2 mg/mL (spheres) and 3 mg/mL (squares) samples are represented with the error bars. The folding reaction was performed at pH 8.4.

Comparison of RP-HPLC and SAXS kinetics suggest that formation of the first disulfide and compaction of hCox17 state are connected. To confirm this hypothesis, we used mass spectrometry (MS) under non-denaturing conditions (native MS), which allows monitoring compaction and oxidation distinctly and simultaneously. While accurate mass determination reports on the oxidation state, charge-state distributions (CSDs) allow detection of coexisting conformers[34]. Starting from fully reduced hCox17, folding was initiated at pH 6.7 and samples analyzed by ESI-MS as the reaction proceeds. The fully reduced protein (Fig. 6A) exhibits a broad CSD centered on the 7+ ion and an average mass of 7406.6 (± 0.1) Da, (expected 7406.6 Da). An average charge state of 7.2 is in good agreement with the value expected for a completely unstructured protein of this size[35], confirming the identification of reduced hCox17 as an IDP. Under the conditions employed here, the species containing one disulfide bond attains its maximal accumulation after 8 h (Fig. 6B), as indicated by a mass of 7404.8 (± 0.3) Da for the 6+ ion. The mass of the other peaks (7405.0-7406.3 Da) are consistent with mixtures of 1 S-S and fully reduced protein, which indeed still constitutes the major hCox17 species at this time point in these conditions. Gaussian fitting of the CSD highlights the presence of two conformational components, a narrow one centered on the 6+ ion, corresponding to

the 1 S-S intermediate, and a broad one centered on the 7+ ion, corresponding to the fully reduced protein. This result indicates that the formation of the first disulfide bridge is linked to a significant chain compaction, accompanied by a dramatic narrowing of the CSD and a shift in average charge state from 7.2 to 5.9. After 17 h of incubation (Fig. 6C), the native 2 S-S form accumulates, as indicated by a further mass shift to 7402.6 (± 0.1) Da. However, the prevalent charge state remains 6+, suggesting that the overall protein compactness is not affected significantly by the formation of the second disulfide bridge.
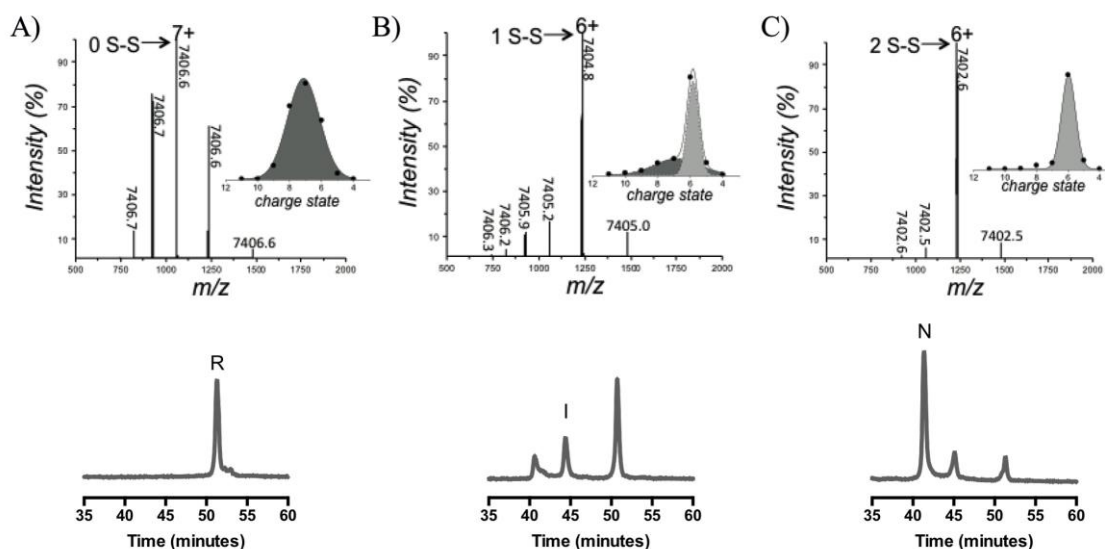
Figure 6. Analysis of the conformation and oxidative state of hCox17 species by native MS. Nano-ESI-MS spectra of 10 μM hCox17 refolding reaction at pH 6.7. hCox17 species are labeled according to its charge state and number of disulfide bridges. Each peak is labeled by the isotopically-averaged mass of the protein. Insets show the Gaussian fitting of the CSD. (A) Fully reduced hCox17 in the presence of 50 mM DTT. (B) Partially oxidized hCox17, 8 h after the removal of DTT. (C) Completely oxidized hCox17, 17 h after the removal of DTT. RP-HPLC analysis is plotted below each time-point, where native (N), intermediate (I) and reduced (R) are labelled.

NMR relaxation data of the amide group can also be used to estimate the rotational diffusion tensor which, in turn, can be reinterpreted in terms of a single correlation time (τ_χ) to describe the overall molecular tumbling[36]. Taking advantage of the slow timescale for the folding event, three NMR subsets of signals were independently analyzed to obtain the correlation times for the different species detectable in the NMR spectrum (see

below). The extended conformation of the unfolded state results in a τ_χ of 5.9 ± 0.9 ns, while an intermediate form (associated to the formation of the Cys36-Cys45 disulfide bond) shows increased compaction (τ_χ of 6.6 ± 1.3 ns), but this species is still more flexible than the native form (τ_χ of 7.1 ± 0.9 ns).

Overall SAXS, MS and NMR data converge to indicate that the formation of the inner disulfide bond is concomitant with compaction of hCox17. Glycerol is a small polyol know to promote protein compaction, leading to protein conformations with reduced solvent accessible surface area[37]. As it can be seen in Fig. 7A and 7B, we observed a concentration-dependent increase of folding rates in the presence of glycerol, which further supports compaction as an important event in hCox17 folding.
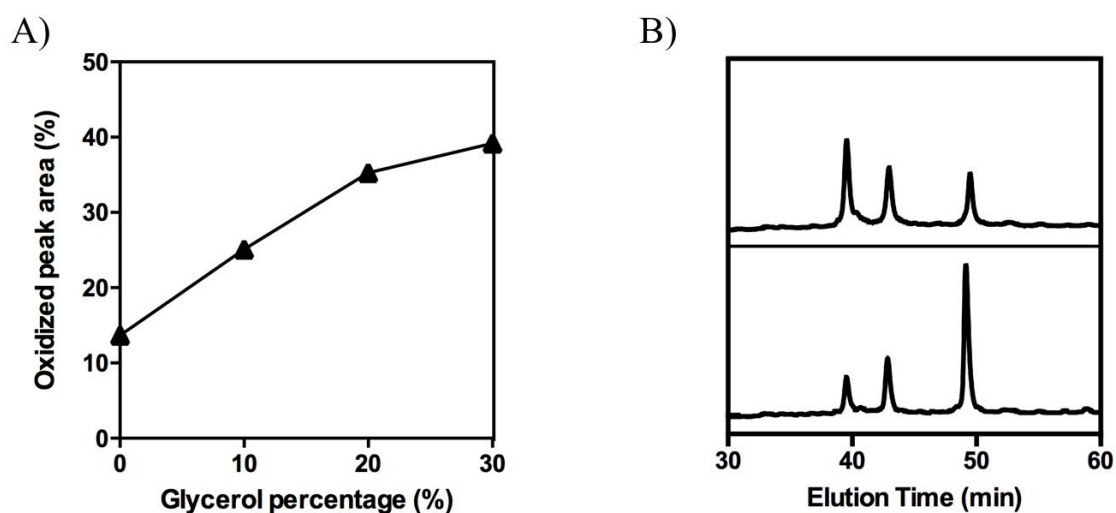
Figure 7. Impact of glycerol on hCox17 folding rates. A) hCox17 was allowed to refold in presence of increasing glycerol concentrations at pH 8.4. After 1 hour of reaction, the mixtures were acid-quenched and analyzed by RP-HPLC. The native peak area is plotted as a function of the glycerol percentage (v/v). B) Representative chromatograms after 1 h of refolding in the absence (bottom) and in the presence (up) of 30% of glycerol. Peaks from left to right: native, intermediate and reduced hCox17.

Adoption of native structure around the hCox17 inner disulfide precedes outer disulfide formation

In order to collect structural information of the hCox17 folding pathway at the residue level, the reaction was monitored using NMR spectroscopy. For this purpose, 15N labelled hCox17 was first reduced in the presence of 100 mM DTT and, following buffer exchange, its oxidation tracked in real time at pH 6.5. The folding reaction progression

was monitored by the chemical shift changes in a ^1H-^{15}N-HSQC over a total period of 36 h. In Fig. S6 representative spectra at 0 h, 8 h and 30 h are shown. It can be observed how some signals from the unfolded/reduced state decrease, how signals from the native state arise and how some intermediate signals appears and disappears in the final folded state; thus suggesting that R, I and N are in slow exchange.
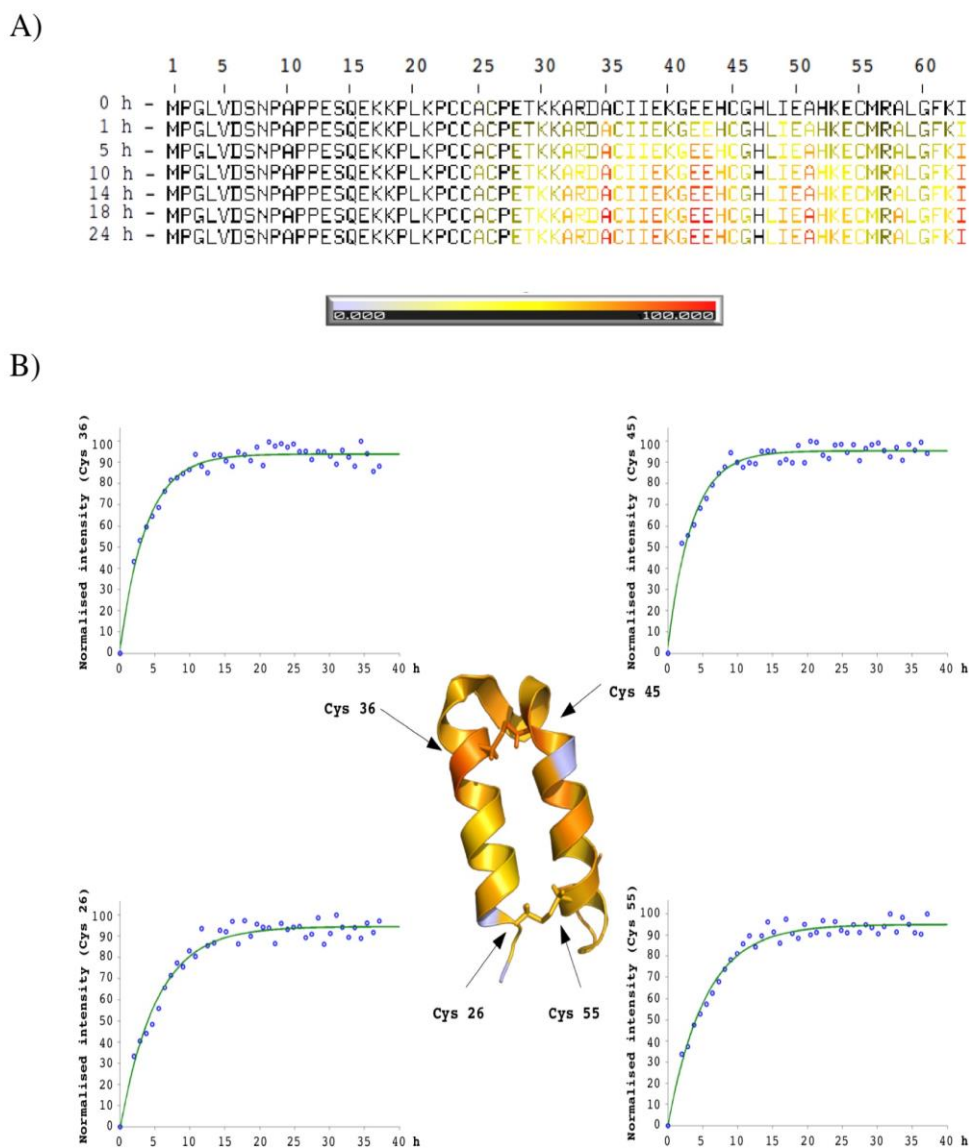
Figure 8. hCox17 oxidative folding followed by real-time solution NMR. Reduced hCox17 was allowed to refold and HSQCs were collected along the reaction. Residues were assigned and their intensities measured. A) Normalized chemical shift intensities depicted according to the color bar, where in white is 0 % formation and in red 100 % formation. Residues 1 to 25 remained disordered and are not taken into account, His 47 was not assigned. B) Normalized chemical shift intensities after 22 h of reaction shown on top of hCox17 solution NMR structure (PDB file: 2RN9), color code as in A. Folding kinetics from Cys

26, Cys 36, Cys 45 and Cys 55 amino acids are represented as an example. In the x axis time is represented as hours, and in y axis normalized chemical shift intensities.

Fig. 8A shows the time evolution represented by a color scale with residue resolution. Under the refolding conditions most of the peaks adopted near final conformation before 22 h. The transition from the disordered to the folded state is nucleated by the formation of the inner disulfide bond (Cys36-Cys45), in agreement with previous results. Quite striking is the relative rapid folding of the residues involved in the stabilization of the ☐☐hairpin fold, including a small hydrophobic core around the covalent link and Arg33 and Glu42 involved in an electrostatic interaction that connects helix I and the 310 helix located between the two ☐☐helical segments[17]. Residues around the second disulfide bond only display native like chemical shifts later on the reaction, whereas the N-terminal tail and Cys23 and Cys24 remain disordered along all the experiment. Figures 8B and S7 show the exponential changes in intensity, reflecting the differences in the progression of the folding reaction between residues.

Mia40 accelerates the transition of hCox17 to the folded state acting on the inner disulfide bond

We proceeded analyzing the effect of Mia40 on the evolution of the different species identified in the hCox17 folding process using the approach we previously reported for Cox19[14]. Indeed, although the Mia40 pathway consists of at least two more components, Erv1 and cytochrome c, Mia40 has been shown to fully catalyze substrate oxidation when added in excess[12,38]. Mia40 is incorporated to the reaction as a fusion with GST which allows to dissect its impact on the hCox17 folding reaction by HPLC-RP since it doesn't elute within the used organic solvent gradient. When an excess of Mia40 was mixed with reduced hCox17 an overall acceleration of the folding reaction was observed (Fig. 9A). As it occurred for Cox19[14], the folding pathway in the presence of Mia40 involves the same species as in non-catalyzed folding, the inner disulfide intermediate being formed rapidly in the first minutes, followed by the formation of native hCox17. Our results also suggest that the reaction can be separated in two independent steps and, accordingly, that Mia40 does not release the fully oxidized substrate, but the same semi-oxidized and compact inner disulfide bonded species that was observed during non-catalyzed folding. We tested this hypothesis using a simple kinetic model that considers the transition from

41

unfolded (U) to a single intermediate (I) with a rate constant K1 and the latter one to the native protein (N) with a rate constant K2 and fitting the areas from the RP-HPLC curves of the species in the absence and presence of Mia40 (Fig. 9B). The r2 is > 0.95 for all the species in both conditions. As we expected, K1 suffers a large 23-fold increase, from 0.3 to 6.9 h-1 in the presence of Mia40, whereas the change in K2 is much more moderate, a 2.5-fold increase, from 0.7 to 1.8 h-1. These data indicate that the formation of the inner disulfide is the rate-limiting step during non-catalyzed folding, whereas, in the presence of Mia40, it is the other way around. In addition, when there is no more reduced species available, the complete depletion of the intermediate requires the same time ($\Box$2 hours) both in the absence and presence of Mia40 (Fig. 9C), which supports that, at least in our experimental setup, Mia40 promotes the transition of hCox17 towards the folded state mainly by accelerating the formation of the inner disulfide bond.

The redox environment impacts both hCox17 folding rate and pathway

To test the influence of the redox environment in hCox17 folding we used oxidized glutathione (GSSG) as a thiol catalyst. GSSG strongly accelerates the conversion of reduced hCox17 towards the native species, with a rate that approaches that of Mia40 catalysed reaction (Fig. 9D). Despite GSSG can promote the formation of both native and non-native disulfide bonds in an unspecific manner, peak I continues being the major 1-SS species, which reflects a sequence encoded preference to form the inner disulfide bond. The acceleration of the reaction comes, however, at the cost of the accumulation of highly compact off-pathway isomers, according to their elution in RP-HPLC. These forms cannot evolve towards native hCox17, likely because the presence of GSSG impedes disulfide reshuffling and proofreading.
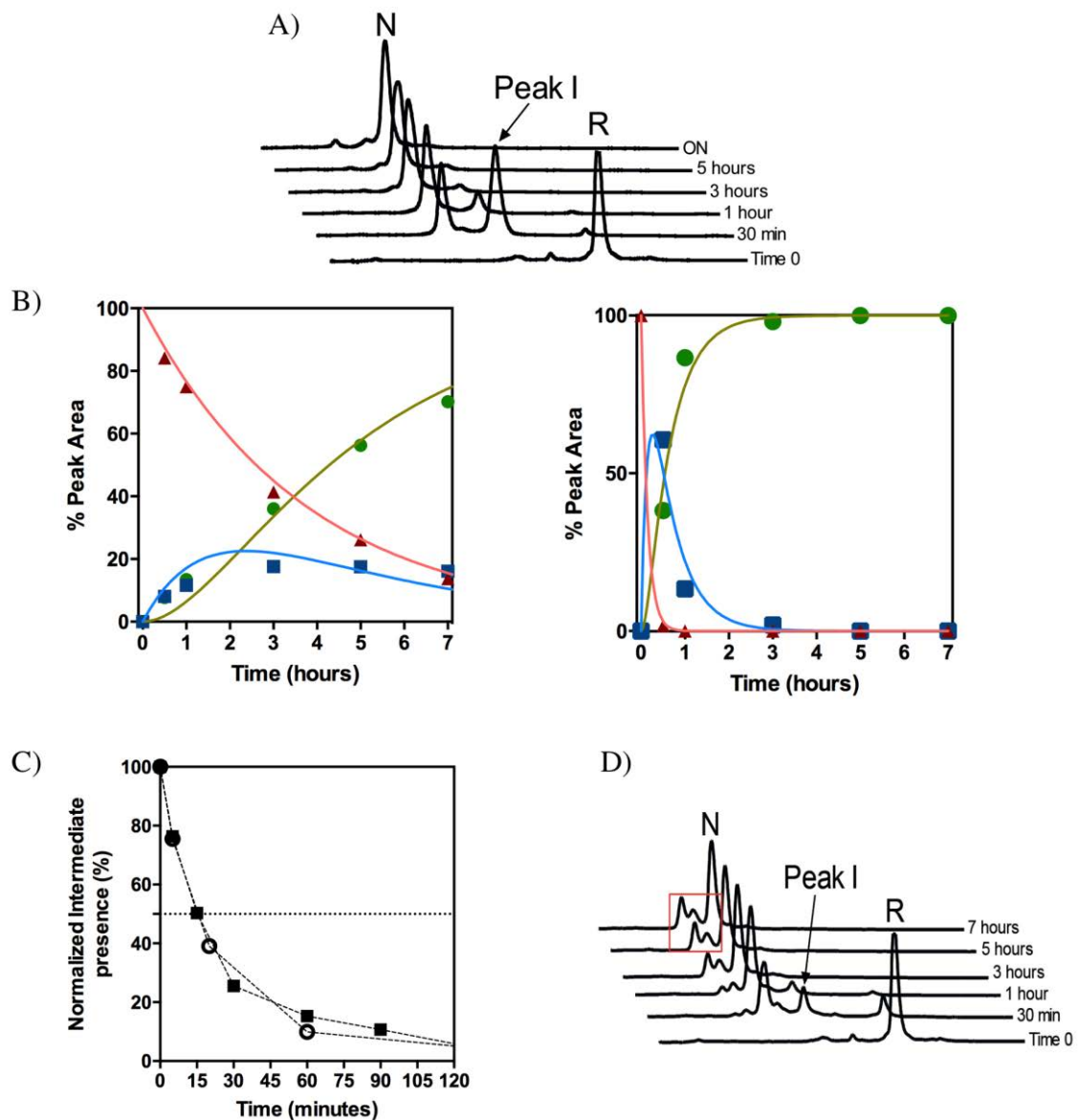
Figure 9. Impact of Mia40 and GSSG in the oxidative folding reaction of hCox17. A) Reduced hCox17 was allowed to refold in the presence of Mia40 (molar ratio 1:2) and the reaction analyzed by RP-HPLC. B) The normalized peak area of reduced (red triangles), intermediate (blue squares) and native (green circles) species are represented as a function of time in the presence (left) and absence (right) of Mia40. The fitting to a U-I-N transition is displayed in solid lines. C) hCox17 folding intermediate disappearance as a function of time in the presence of Mia40 (open circles) and in stop and go assays (solid squares). A dashed line at 50% is also represented to identify the half-life of the specie. D) RP-HPLC chromatograms of hCox17 folding reaction in the presence of 1 mM GSSG. Red boxes highlight off-pathway intermediates.

Inducing secondary structure formation in the reduced state does not increase hCox17 folding rate

We tested whether induction of helical formation in the reduced state of hCox17 could accelerate the conformational transition towards the native state as we observed previously for Cox19[14]. The fluorinated alcohol 2,2,2-trifluoroethanol (TFE) promotes a concentration dependent coil-to-helix transition in peptides and disordered proteins[39]. Accordingly, addition of increasing concentrations of TFE, from 5 to 40%, while keeping the pH at 8.4, to reduced hCox17 resulted in an increase in its helical content, as deduced from the CD spectra (Fig. 10A). However, when we assessed the impact of TFE on hCox17 folding reaction we could not observe any increase in folding rates. Indeed, the folding rate is reduced as the helical content increases (Fig. 10B), a result that could respond to an inhibitory effect of pre-formed helical structure or alternatively TFE-induced formation of non-native helical conformations[40].

The helix/coil transition algorithm AGADIR predicts very low propensities for both hCox17 helices, 2.27 % for helix I and 0.47 % for helix II (Table S4). We redesigned individually helices I and II to test whether an increase in their local □-helix propensities had any impact on the kinetics of hCox17 folding. To this aim we introduced residues with higher intrinsic helical propensity and novel i+3, i+4 contacts between side chains (Table S4), a strategy that has been used to drastically stabilize and increase the folding rates of helical proteins[41]. The expected increase in helical content upon mutation was 6-fold for helix I and 10-fold for helix II (Table S4). The two mutants were purified to homogeneity and their folding rates compared with WT hCox17. As shown in Fig. 10C, helix I mutant folds at the same rate that WT. helix II mutant folds faster than WT, but the impact is small, only 1.5-fold.
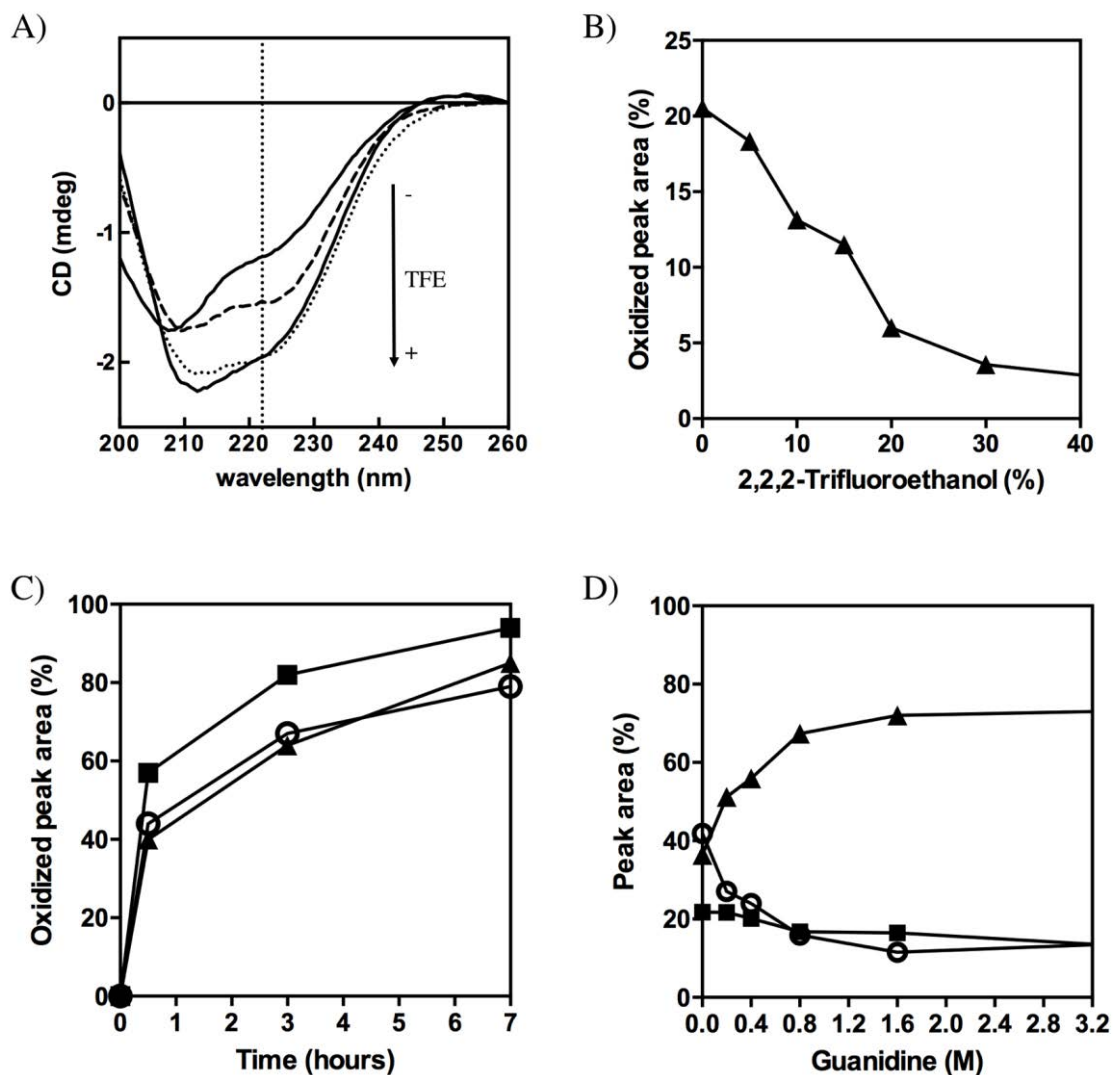
Figure 10. Influence of secondary structure in the oxidative folding of hCox17. A) Far-UV CD spectra of reduced hCox17 in the absence (upper solid line) and in the presence of 10% (dashed line), 20% (dotted line) and 40% (lower solid line) of TFE. The characteristic □-helix 222 nm signal is indicated as a vertical dotted line. B) Reduced hCox17 was allowed to refold in the presence of increasing concentrations of TFE. After 1 h of reaction, the samples were acid-quenched and analyzed by RP-HPLC. The native peak area was integrated and plotted as a function of TFE concentration. C) Reduced hCox17 mutant H1 (circles), H2 (squares) and wild type (solid triangles) proteins were allowed to refold and the normalized native RP-HPLC peak areas are plotted as a function of the refolding time. D) hCox17 refolding in the presence of increasing concentrations of Guanidine-HCl was quenched with TFA after 2 hours of reaction. The amount of native (solid triangles), intermediate (solid squares) and reduced (open circles) species are plotted as a function of denaturant concentration.

To further discard that residual helical propensity in the reduced state could play a relevant role in the folding of hCox17, we tested if the reaction could be completed in the presence of GdnHCl, a classic denaturant agent that will difficult the formation of secondary structure at the early stages of folding. At low concentrations (< 3 M) the presence of GdnHCl does not preclude hCox17 folding, but instead accelerates the reaction in a concentration dependent manner, achieving maximum rates at 1.6 M concentration (Fig. 10D), which indeed suggests that transient interactions in the reduced state can play a certain inhibitory effect on hCox17 folding.

Discussion

From as early as Anfinsen seminal work we know that the structure of globular proteins is imprinted in its genetic sequence[42]. It is also true that the lack of specific regular structure, characteristic of IDPs, is coded in the protein sequence[43,44]. The question that arises is: How can a sequence encode at the same time for ordered and disordered conformations when both states are biologically relevant? This is the case of hCox17, for which the initially disordered state is a requirement for its diffusion across the mitochondrial outer membrane, whereas the transition to a folded state in the IMS is necessary to transfer copper to cytochrome c oxidase. This places hCox17 in the new of class of conditionally disordered proteins, defined as polypeptides that under physiological conditions can exist in at least two states, one that shows a high degree of flexibility and disorder and a second state that shows a higher degree of order[21]. This mechanism allows to fine tune the function of these proteins according to the environmental context with a low energy cost. The slow folding reaction of the CHCH hCox17 domain and the possibility to trap and identify the major intermediates that populate the pathway has allowed us to integrate a set of techniques reporting in real time on covalent linkage, compactness, secondary structure and protein shape as it folds towards its functional conformation.

Disulfide bond formation is error-prone, particularly at the early stages of packing, often resulting in the accumulation of mispaired intermediates, which slows down folding reactions, since reshuffling is required to attain the native disulfide connectivity[45]. Despite being potentially able to form a large number of covalently-linked intermediates, hCox17 folds using sequential native disulfide formation, starting with the inner disulfide followed by the formation of the second outer disulfide, which results in a highly funneled pathway where a single native cross-linked on-pathway intermediate is accumulated, both in the absence and presence of Mia40. In the view of the simplicity of hCox17 folding pathway one would expect it to proceed fast. In contrast, spontaneous folding of hCox17 is exceedingly slow. The same behaviour was observed for Cox19, yet another substrate of Mia40 belonging to the CX_9C family[14]. This parsimony allows enough residence in the reduced and disordered state for the proteins to be translocated from the cytosol to the IMS, since folding is a strong kinetic competitor of mitochondrial import[11].

The succession of events leading to hCox17 folding is neither governed by the local vicinity of Cys residues, nor by the reduction in entropy of the system upon disulfide crosslinking, since the closure of the outer disulfide bond, or a non-native Cys23-Cys55 connection would close a much longer loop that the inner bond. This points at differential structural protection upon disulfide formation as the most probable factor responsible for the observed hCox17 folding hierarchy. It is clear from our data that the formation of the first disulfide bond is associated with the transition of hCox17 from an initially disordered state to a compact conformation. Structural collapse, disulfide formation and local secondary structure formation occur in a coordinated way, according to NMR, SAXS, FTIR and MS data. In excellent agreement with our results, Banci and coworkers, showed that a mimic of the inner disulfide bond intermediate, in which the outer Cys were mutated to Ser, exhibited stable inter-helical interactions and displayed a structure similar to that of native hCox17 containing two disulfides[47]. NMR data indicate that the first hCox17 regions to attain native chemical shifts are those around the inner disulfide where a small hydrophobic core and an electrostatic interaction between the carboxylate group of Glu42 and the guanidine group of Arg33 exist. These interhelical interactions surround the inner disulfide bond and would bring the two helices close to each other into the detected compact conformer. This species is likely a stable disulfide secure intermediate in which the disulfide bond is significantly protected from SH/SS reshuffling and reduction, and thereby is able to form the native protein directly, *i.e*., by oxidation of their exposed and reactive thiols. These properties suggest that this specie would not diffuse freely backward to the cytosol. The second disulfide can thus form more slowly in the IMS, without affecting much the equilibrium between unfolded and folded species.

It is generally assumed that protein folding drives subsequent disulfide formation, however our data suggest that conditionally disordered proteins like hCox17 deviate from this general rule, all the data converging to indicate that the formation of the inner disulfide is itself the molecular event triggering the transition of hCox17 between the initially disordered state and the folded functional conformation. This would promote immediate native-like compactness by forcing and stabilizing the interactions that envelop this covalent link. Indeed, NMR data indicate that all the residues between and adjacent to Cys36 and Cys45 arrive to their native positions with a similar fast speed, independently if they are polar, charged or hydrophobic. This occurs before the two α-helices have attained their definitive relative positions. In fact, NMR and SAXS data

indicate that despite the inner disulfide intermediate has already attained a compactness equivalent to that of the native state, is still flexible. This might explain the accelerating role of glycerol, known to reduce protein flexibility[48], which by decreasing the motion of the yet metastable helices might favor the formation of the outer disulfide bond.

In a way, the transition between the disordered and ordered states of hCox17 can be considered as a conformational switch controlled by a single selective post-translational modification: inner disulfide oxidation. Spontaneous formation of helical structure and hydrophobic collapse are usually fast events that cannot act as the drivers for the folding of a protein that should stay long in a disordered state waiting for translocation. Disulfide oxidation in the cytosol is a much slower reaction that might provide an effective time window for protein import. Another point to consider in hCox17 translocation to the IMS is aggregation. Indeed, the unfolded or partially unfolded states of globular proteins usually have an exacerbated propensity to aggregate and are recognized by chaperones to mediate their degradation[49]. How does hCox17 avoid these side reactions in the cytosol? An analysis of its sequence with the TANGO aggregation prediction algorithm indicates that it is absolutely devoid of aggregation-prone sequences (not shown), a property that shares with a number of bona fide IDPs[50]. This explains why the protein can remain soluble for long time in the unfolded state and why it lacks a defined nucleus that can trigger its folding. The disordered state of hCox17 should be viewed as an essential component of hCox17 function and therefore its sequence has evolved to optimize its properties as well as those of the folded state.

Many disordered proteins refold when they bind their partners and, indeed, this is the case of hCox17 *in vivo*, which folds in the IMS upon binding to Mia40. We show here that this binding event acts a catalyser of a disorder-to-order transition that is already delineated in the primary sequence of this conditionally disordered protein, since the detected folding intermediates in catalysed and non-catalysed reactions are identical. Indeed, unspecific oxidants like GSSG suffice to attain folding speeds equivalents to those obtained in the presence of the chaperone. However, this acceleration comes at the cost of the formation of off-pathway non-native linked intermediates, which indicates that Mia40 allows preserving the protein intrinsic propensities during accelerated folding, in good agreement with previous results obtained by Banci and co-workers, using Cys to Ser mutants of hCox17 and solution NMR[51]. The formation of the inner disulfide is the rate-limiting step in the folding reaction of hCox17 and therefore it prevents that hCox17

would fold prematurely. As expected for a canonical foldase, Mia40 catalyses this rate-limiting molecular event, having limited impact on the formation of the second native disulfide, at least in our experimental conditions.

There are two main models[52,53] explaining how a disordered protein regains structure by binding to its partners: 1) The conformational selection model, which assumes that a small proportion of the intrinsically disordered polypeptide population is in appropriate metastable configuration, capable to interact with its target. 2) The folding upon binding model, which assumes that intrinsically disordered regions first bind to the partner to subsequently fold. The recent observation that in the disordered ensemble of yCox17, helix II seems to be transiently populated has suggested that Mia40 substrates from the CX_9C family would fold according to a conformational selection mechanism, which would result in the initial formation of the second helix, oxidation of the inner disulfide and subsequent formation of the first helix and the outer disulfide[54]. The sequence of helix II in yCox17 has much higher intrinsic helical propensity (18.78 %) than any of the two helices in hCox17 and our data does not support such conformational selection driving the folding of hCox17. Indeed, it has been just proposed that Mia40 acts as a holding trap, trapping incoming disordered substrates via hydrophobic binding, a reaction that is essential and sufficient for translocation, immediately followed by catalyzed oxidative folding[55], a pathway that is consistent with a folding upon binding mechanism. The oxidoreductase and holdase activities of Mia40 seem to reside in different regions of the protein and are both essential. However, despite an oxidase-deficient Mia40 mutant is inviable, it can be partially rescued by the addition of exogenous chemical oxidants[55], an observation that suggests that many Mia40 substrates encode for its spontaneous transition between a disordered translocation-competent state and its functional conformation, as we show here for hCox17.

Materials and Methods

Analysis of protein disorder

Intrinsic disorder in hCox17 was predicted using the web-based algorithms FoldIndex[24], IUPRED[26], ESpritz[25], PONDR-FIT[27] and RONN[28] with default settings. An 11-residue sliding window was used in all cases.

Protein purification

hCox17-GST cDNA, gently sent by L. Banci and coworkers, was transformed to *E. coli* BL21 DE3, and *E. coli* Origami2 cells in order to express *wild type* (WT) and mutant proteins. As the protein, did not have any Trp or Tyr residue in its sequence, a point S7W mutation was introduced in the disordered N-tail. This mutation did not affect the kinetics or the species population during the oxidative folding and allowed us proper quantification in the different assays. The protein was expressed and purified as previously described[17].

Infrared spectroscopy

FTIR spectra of human hCox17 were measured at 6 mg/mL in 50 mM Tris pH 8.4, 100 mM NaCl in D_2O. Measurements were performed on a FT600 Bio-Rad spectrometer equipped with a MCT detector, using a demountable liquid cell with calcium fluoride windows and 50 μm spacers. Typically, 500 scans for each background and sample were collected and the spectra were obtained with a nominal resolution of 2 cm^{-1} at 22 ºC. The spectra were acquired in slow kinetics mode and the measurements were performed continuously for 6 h. The buffer contribution to the spectra was subtracted. Data treatment and band decomposition of the original amide I' band was performed as previously described with GRAMS 9.0 software[57]. For each component, four parameters were considered: band position, band height, band width, and band shape. In decomposing the amide I' band, gaussian components were used. The number and position of the component bands were obtained through Fourier self-deconvolution using Lorentzian lineshape and the parameters recommended for protein analysis, 13 cm^{-1} for full-width at half-height (FWHH) and 2.4 for the resolution enhancement factor [57]. Initial heights were set at 90% of those of the original spectrum for the bands in the wings and

for the most intense component and at 70% of the original intensity for the other bands[57]. The curve-fitting procedure was accomplished in two steps: i) The band position was fixed, allowing width and heights to approach final values, and ii) band positions were left to change. The quality of the fitting was evaluated visually by overlapping the reconstituted overall curve on the original spectrum and by examining the residual obtained by subtracting the fitting from the original curve. Correlation coefficients between the original and the fitted spectra were ≥ 0.999. The mathematical solution of the decomposition may not be unique, but if restrictions are imposed, such as maintenance of the initial band positions in an interval of ± 1 cm^{-1}, preservation of the bandwidth within the expected limits, or agreement with theoretical boundaries or predictions, the result becomes, in practice, unique.

Size Exclusion Chromatography

Oxidized and reduced (overnight incubation with 200 mM DTT) hCox17 were injected using a 500 μL loop into a Superdex 75 100/30G GL column (GE Healthcare) coupled to an AKTA - Fast Protein Liquid Chromatography device (GE Healtcare) equilibrated with 50 mM Tris pH 8.4, 100 mM NaCl buffer. hCox17 elution was monitored by following absorbance at 280 and 214 nm. Initial protein concentration was 50 mM.

CD spectroscopy

Far UV CD spectra were acquired on a Jasco-710 spectropolarimeter continuously purged with nitrogen and thermostated at 25 °C and quartz cuvettes of 2 mm width. Spectra were recorded from 260 to 200 nm setting the bandwidth at 1 nm, scan speed at 200 nm/min and data pitch at intervals of 0.5 nm. For folding experiments, fully reduced hCox17 with 200 mM DTT overnight was allowed to oxidize in 50 mM NaCl, 20 mM phosphate buffer at pH 7.0 after removing the reducing agent with a PD-10 desalting column (GE Healthcare, REF. 17-0851-01). Scans with a data pitch of 0.5 nm from 200 to 260 nm were collected every 30 minutes for 20 hours. For thermal denaturation studies the molar ellipticity at 222 nm was recorded each 0.2 ºC with an increasing heat rate of 1 ºC/min in the 25 to 95 ºC range. For TFE experiments, reduced hCox17 was mixed with TFE at 10, 20 and 40% final concentration in 20 mM phosphate buffer 50 mM NaCl and the pH adjusted to 7.0.

High-Performance Liquid Chromatography

hCox17 at 1 mg/mL was reduced in 50 mM Tris pH 8.4, NaCl 100 mM and 200 mM DTT over night at room temperature. To initiate folding, samples were passed through a PD-10 desalting column previously equilibrated with 50 mM Tris pH 8.4, NaCl 100 mM. Selected concentrations of GSH, Glycerol, TFE and Mia40 were added to the collected fraction to modulate or enhance the folding kinetics. To follow the oxidative reaction, removed aliquots at selected times were quenched with 2% TFA and loaded in a *Waters 2690 HPLC* system with a C4 column *(Phenomenex) 250 x 4.6 5 mm*. A linear gradient of 20%-35% of 0.1% TFA in acetonitrile was applied for 60 minutes at a flow-rate of 0.75 ml/min, Eluted species were monitored by a coupled *Waters 2487* detector at 214 nm and 280 nm wavelength. In the same way, a preparative *VYDAC protein C4* was used to purify the intermediate species of the reaction. An increased 3 mL/min flow was applied to the same linear gradient of acetonitrile as above for 120 minutes. The purified species were identified by mass spectrometry and the reaction intermediate were quantified for the stop/go experiments in 50 mM Tris pH 8.4 NaCl 100 mM at 0.5 mg/mL[14,22].

MALDI-TOF and TOF/TOF analysis

The assays were performed on an Autoflex Speed mass spectrometer (Bruker, Bremen). The free Cys residues of Peak I folding intermediate species were alkylated by adding the samples, dissolved in 0.1% trifluoroacetic acid (30 μL), to an excess of 75 mM iodoacetamide in 50 mM ammonium bicarbonate buffer, pH=8, (70 μL), enough to rise the pH of the reaction mixture, to minimize the risk of disulfide reshuffling. The reaction was allowed to proceed for 30 min at room temperature.

Samples before and after alkylation were then analyzed by MALDI-TOF mass spectrometry using sinapinic acid as matrix. A mixture of protein standards was used for internal calibration.

For fingerprint analysis, the alkylated proteins were first purified using ZipTip (Millipore) C18 micro columns, evaporated, redissolved in 1 M Urea, 25 mM ammonium bicarbonate, subjected or not to reduction by treatment with 10 mM DTT for 1 hour at r.t., and digested with trypsin (3h at 37 °C). The digests were then purified using ZipTip C18 and analyzed by MALDI TOF/TOF mass spectrometry using HCCA as matrix.

Electrospray-ionization mass spectrometry (ESI-MS)

Lyophilized hCox17 (0.5 mg) was dissolved in 3 mL of 10 mM ammonium acetate, pH 8.0, 50 mM DTT, and incubated statically overnight at room temperature, in order to reduce Cys residues. The sample was buffer-exchanged by reversed-phase chromatography (C-18 Zip Tip, Merck-Millipore, Darmstadt, Germany), eluting by 50% acetonitrile, 0.5% formic acid, and dried in a Speed Vac centrifugal system (Analitica De Mori, Milano, Italy). The pellet was ressuspended in 3 mL of 10 mM ammonium acetate, pH 6.7, to initiate oxidative refolding and incubated statically at room temperature for variable time. Aliquots were injected into a hybrid quadrupole-time of flight (TOF) mass spectrometer (QSTAR Elite, AB Sciex, Forster City, CA, USA), employing metal-coated borosilicate capillaries with emitter tip of ~1 μm internal diameter (ThermoFisher Scientific, Waltham, MA, USA). The main instrumental parameters were: ion-spray voltage, 1.1 kV; curtain gas, 20 PSI; declustering potential, 80 V. The sample source and the instrument interface were kept at room temperature. Spectra were averaged over 1-minute acquisition.

Small-angle X-ray scattering

SAXS measurements were carried out on ESRF beamline BM29 at $\lambda = 0.992$ Å [58]. Scattering curves were recorded in the momentum transfer range $0.033 < s < 0.498$ Å-1 ($s = 4\pi \sin (\theta)/\lambda$, 2θ is the scattering angle) using a Pilatus 1 M detector (Dectris Ltd., Baden, Switzerland). Two samples of hCox17 at 6 mg/mL and 4 mg/mL were reduced with DTT, then the reducing agent was removed as described for oxidative folding assays in HPLC to give rise to final concentrations of 3 mg/mL and 2 mg/mL in a 50 mM Tris pH 8.4 NaCl 100 mM buffer. Both solutions were allowed to air oxidize and small volumes (40 μL) were taken periodically during 10 (2 mg/ml) and 12 (3 mg/mL) hours and submitted to a SAXS measurement. Buffer samples were measured before and after protein solutions. Data reduction, correction for radiation damage, and buffer subtraction, were performed by the beam line automated procedure. Both concentrations provided equivalent curves with to indication of interparticle interactions and were treated independently. The radius of gyration, R_g, for each of the curves was derived using Guinier's approximation with the ATSAS software[59]. The maximum particle size (D_{max}) was determined from the pair-wise distance distribution function, $P(r)$, computed with PRIMUS[60]

Nuclear Magnetic Resonance

^{15}N-labelled hCox17 was reduced in 50 mM Tris pH 8.0, 100 mM NaCl and 100 mM DTT overnight at room temperature. To initiate the folding process, the sample buffer was exchanged to 50 mM Tris pH 6.5, 100 mM NaCl and 1 mM DTT using disposable PD-10 desalting columns. Subsequently, the protein was concentrated up to 150 μM. Refolding process was monitored by collecting ^{1}H-^{15}N HSQC spectra every 52 minutes for 36 hours (starting at 2 h) in an 800 MHz Bruker Avance III spectrometer at 298 K. Resonances were assigned based on Banci's previous work (BMRB: 11019 & 11020)[17]. Peak intensities were measured using NMRPipe[61]. Normalized folding transition percentage for each residue in hCox17 was calculated using in-house built MatLab® scripts.

^{15}N-^{13}C-labelled hCox17 was reduced in 50 mM Tris pH, 100 mM NaCl and 100 mM DTT. ^{1}H-^{15}N-HSQC and HNCACB spectra were collected in an 800 MHz Bruker Avance III spectrometer at 298 K. Backbone chemical shifts were assigned for Ala, Trp and Ser and their neighbouring residues as they have a clear Cβ chemical shifts. Protein flexibility and secondary structure was predicted by comparing the Random Coil index values from the backbone chemical shifts (HN, N, Cα and Cβ)[31].

Backbone amide ^{15}N relaxation data (R_1, $R_{1\rho}$ and heteronuclear ^{15}N-^{1}H-NOE were measured for hCox17 at 600 MHz by published methods[62]. For $R_{1\rho}$ experiments, the employed spin-lock field was 3 kHz. Sampled relaxation time points ranged from 0.1 to 1.2 s for R_1, and 50 to 250 ms for $R_{1\rho}$. For the heteronuclear ^{15}N-^{1}H-NOE experiment, steady-state HN saturation was achieved with a 4 s train of square 120° pulses, and a 12 s interscan delay was used in the reference (non-saturated) spectrum. The backbone ^{15}N relaxation data was analyzed with the Lipari-Szabo formalism[63] using in-house MatLab® scripts.

References

1.   Creighton, T. E. The protein folding problem. *Science* **240**, 267, 344 (1988).

2.   Tanford, C. Protein denaturation. *Adv Protein Chem* **23**, 121–282 (1968).

3.   Gelman, H. & Gruebele, M. Fast protein folding kinetics. *Q Rev Biophys* **47**, 95–142 (2014).

4.   Schonbrun, J. & Dill, K. A. Fast protein folding kinetics. *Proc Natl Acad Sci U S A* **100**, 12678–12682 (2003).

5.   Eaton, W. A. *et al.* Fast kinetics and mechanisms in protein folding. *Annu Rev Biophys Biomol Struct* **29**, 327–359 (2000).

6.   Goldbeck, R. A., Thomas, Y. G., Chen, E., Esquerra, R. M. & Kliger, D. S. Multiple pathways on a protein-folding energy landscape: kinetic evidence. *Proc Natl Acad Sci U S A* **96**, 2782–2787 (1999).

7.   Sekhar, A. *et al.* Probing the free energy landscapes of ALS disease mutants of SOD1 by NMR spectroscopy. *Proc Natl Acad Sci U S A* **113**, E6939–E6945 (2016).

8.   Callender, R. H., Dyer, R. B., Gilmanshin, R. & Woodruff, W. H. Fast events in protein folding: the time evolution of primary processes. *Annu Rev Phys Chem* **49**, 173–202 (1998).

9.   Akiyama, S. *et al.* Conformational landscape of cytochrome c folding studied by microsecond-resolved small-angle x-ray scattering. *Proc Natl Acad Sci U S A* **99**, 1329–1334 (2002).

10.   Jaswal, S. S. Biological insights from hydrogen exchange mass spectrometry. *Biochim Biophys Acta* **1834**, 1188–1201 (2013).

11.   Mesecke, N. *et al.* A Disulfide Relay System in the Intermembrane Space of Mitochondria that Mediates Protein Import. *Cell* **121**, 1059–1069 (2005).

12.   Banci, L. *et al.* MIA40 is an oxidoreductase that catalyzes oxidative protein folding in mitochondria. *Nat Struct Mol Biol* **16**, 198–206 (2009).

13.   Banci, L., Barbieri, L., Luchinat, E. & Secci, E. Visualization of Redox-Controlled Protein Fold in Living Cells. *Chemistry & Biology* **20**, 747–752 (2013).

14.   Fraga, H. *et al.* The mitochondrial intermembrane space oxireductase Mia40 funnels the oxidative folding pathway of the cytochrome c oxidase assembly protein Cox19. *J Biol Chem* **289**, 9852–9864 (2014).

15.   Fraga, H. & Ventura, S. Oxidative Folding in the Mitochondrial Intermembrane Space in Human Health and Disease. *IJMS* **14**, 2916–2927 (2013).

16.   Arnesano, F., Balatri, E., Banci, L., Bertini, I. & Winge, D. R. Folding Studies of Cox17 Reveal an Important Interplay of Cysteine Oxidation and Copper Binding. *Structure* **13**, 713–722 (2005).

17.   Banci, L. *et al.* A Structural-Dynamical Characterization of Human Cox17. *J Biol Chem* **283**, 7912–7920 (2008).

18.   Chojnacka, M., Gornicka, A., Oeljeklaus, S., Warscheid, B. & Chacinska, A. Cox17 Protein Is an Auxiliary Factor Involved in the Control of the Mitochondrial Contact Site and Cristae Organizing System. *J Biol Chem* **290**, 15304–15312 (2015).

19.   Remacle, C. *et al.* Knock-down of the COX3 and COX17 gene expression of cytochrome c oxidase in the unicellular green alga Chlamydomonas reinhardtii. *Plant Mol Biol* **74**, 223–233 (2010).

20.   Reichmann, D. & Jakob, U. The roles of conditional disorder in redox proteins. *Curr Opin Struct Biol* **23**, 436–442 (2013).

21.   Jakob, U., Kriwacki, R. & Uversky, V. N. Conditionally and transiently disordered proteins: awakening cryptic disorder to regulate protein function. *Chem Rev* **114**, 6779–6805 (2014).

22.   Fraga, H., Graña-Montes, R., Illa, R., Covaleda, G. & Ventura, S. Association between foldability and aggregation propensity in small disulfide-rich proteins. *Antioxid Redox Signal* **21**, 368–383 (2014).

23.   Darby, N. & Creighton, T. Dissecting the disulphide-coupled folding pathway of bovine pancreatic trypsin inhibitor. *J Mol Biol* **232**, 873–896 (1993).

24.   Prilusky, J. *et al.* FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics* **21**, 3435–3438 (2005).

25.   Walsh, I., Martin, A. J. M., Di Domenico, T. & Tosatto, S. C. E. ESpritz: accurate and fast prediction of protein disorder. *Bioinformatics* **28**, 503–509 (2012).

26.   Dosztányi, Z., Csizmok, V., Tompa, P. & Simon, I. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* **21**, 3433–3434 (2005).

27.   Xue, B., Dunbrack, R. L., Williams, R. W., Dunker, A. K. & Uversky, V. N. PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. *Biochim Biophys Acta* **1804**, 996–1010 (2010).

28.   Yang, Z. R., Thomson, R., McNeil, P. & Esnouf, R. M. RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics* **21**, 3369–3376 (2005).

29.   Koch, J. R. & Schmid, F. X. Mia40 targets cysteines in a hydrophobic environment to direct oxidative protein folding in the mitochondria. *Nat Commun* **5**, 3041 (2014).

30.   Deponte, M. & Hell, K. Disulphide bond formation in the intermembrane space of mitochondria. *J Biochem* **146**, 599–608 (2009).

31.   Berjanskii, M. & Wishart, D. S. NMR: prediction of protein flexibility. *Nat Protoc* **1**, 683–688 (2006).

32.   Graewert, M. A. & Svergun, D. I. Impact and progress in small and wide angle X-ray scattering (SAXS and WAXS). *Curr Opin Struct Biol* **23**, 748–754 (2013).

33.   Bernadó, P. & Blackledge, M. A self-consistent description of the conformational behavior of chemically denatured proteins from NMR and small angle scattering. *Biophys J* **97**, 2839–2845 (2009).

34.   Kaltashov, I. A., Bobst, C. E. & Abzalimov, R. R. Mass spectrometry-based methods to study protein architecture and dynamics. *Protein Sci* **22**, 530–544 (2013).

35.   Testa, L., Brocca, S. & Grandori, R. Charge-surface correlation in electrospray ionization of folded and unfolded proteins. *Anal Chem* **83**, 6459–6463 (2011).

36.   Ortega, G., Castaño, D., Diercks, T. & Millet, O. Carbohydrate affinity for the glucose-galactose binding protein is regulated by allosteric domain motions. *J Am Chem Soc* **134**, 19869–19876 (2012).

37.   Vagenende, V., Yap, M. G. S. & Trout, B. L. Mechanisms of protein stabilization and prevention of protein aggregation by glycerol. *Biochemistry* **48**, 11084–11096 (2009).

38.   Bien, M. *et al.* Mitochondrial disulfide bond formation is driven by intersubunit electron transfer in Erv1 and proofread by glutathione. *Mol Cell* **37**, 516–528 (2010).

39.   Chiti, F. *et al.* Acceleration of the folding of acylphosphatase by stabilization of local secondary structure. *Nat Struct Biol* **6**, 380–387 (1999).

40.   Sanders, C. R. & Sönnichsen, F. Solution NMR of membrane proteins: practice and challenges. *Magn Reson Chem* **44 Spec No**, S24-40 (2006).

41.   Viguera, A. R., Villegas, V., Avilés, F. X. & Serrano, L. Favourable native-like helical local interactions can accelerate protein folding. *Fold Des* **2**, 23–33 (1997).

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

43.   Uversky, V. N., Gillespie, J. R. & Fink, A. L. Why are 'natively unfolded' proteins unstructured under physiologic conditions? *Proteins* **41**, 415–427 (2000).

44.   Romero, P. *et al.* Sequence complexity of disordered protein. *Proteins* **42**, 38–48 (2001).

45.   Arolas, J. L., Aviles, F. X., Chang, J.-Y. & Ventura, S. Folding of small disulfide-rich proteins: clarifying the puzzle. *Trends Biochem Sci* **31**, 292–301 (2006).

Supplementary Information for:

Disulfide driven folding for a conditionally disordered protein

Hugo Fraga[1,2,3‡], Jordi Pujols[1,2‡], Marcos Gil-Garcia[1,2], Alicia Roque[2], Ganeko Bernardo-Seisdedos[4], Carlo Santambogrio[5], Joan-Josep Bech-Serra[6], Francesc Canals[6], Pau Bernadó[7], Rita Grandori[5], Oscar Millet[4] and Salvador Ventura [1,2]*

[1] Institut de Biotecnologia i Biomedicina. Universitat Autònoma de Barcelona, 08193-Bellaterra, Spain.

[2] Departament de Bioquímica i Biologia Molecular. Universitat Autònoma de Barcelona, 08193-Bellaterra, Spain.

[3] Departamento de Bioquimica, Faculdade de Medicina da Universidade do Porto, Portugal

[4] Protein Stability and Inherited Diseases Laboratory, CIC bioGUNE, 48160-Derio, Spain.

[5] Department of Biotechnology and Biosciences, University of Milano-Bicocca, Italy

[6] Vall d'Hebron Institute of Oncology (VHIO), Barcelona, Spain

[7] Centre de Biochimie Structurale, INSERM-U1054, CNRS UMR-5048,Université de Montpellier. 29, rue de Navacelles, 34090-Montpellier, France

Table S1. Reference chemical shifts. Chemical shifts for Ala, Trp and Ser and their neighbouring residues. Same nuclei chemical shifts are compared with BMRB 11019.

Num.	Res.	COX17_Red N	H	CA	CB	COX17_2S-S (BMRB 11019) N	H	CA	CB
-1	F			58,14	40,14				
0	T	117,71	8,36	61,91	70,19	118,31	8,32	60,2	72,47
1	M	124,95	8,59	53,67	32,95	125,37	8,53	51,88	34,59
9	P			63,14	32,34			61,49	34,1
10	A	126,11	8,49	50,58	18,46	126,75	8,56	48,93	19,95
13	E			57,26	30,38			55,52	32
14	S	117,1	8,61	64,2	58,7	117,43	8,53	56,87	65,84
15	Q	122,98	8,68	56,25	29,76	123,15	8,55	54,36	31,34
24	C			57,53	30,97			56,56	30,62
25	A	124,95	8,46	53,38	19,2	126,03	7,96	50,66	22,29
26	C	118,87	8,59	57,24	30,31	116,96	9,38	57,22	45,55
28	E			57,31	30,54			58,83	30,77
29	T	116,43	8,45	62,52	70,06	113,77	8,9	63,09	68,82
30	K	124,95	8,59	53,67	32,95	128,05	8,25	59,2	35,09
31	K			55,48	33,27			57,87	34,4
32	A	126,25	8,63	53,02	19,63	120,87	7,75	53,36	20,32
33	R	120,97	8,62	56,88	27,98	119,54	8,11	58,18	32,1
34	D			54,9	41,48			56,14	41,04
35	A	124,52	8,45	53,16	19,59	121,73	8,24	53,41	20,12
36	C	118,69	8,57	59,15	28,18	118,58	7,61	59,17	41,46
50	E			58,59	28,67			57,54	30,95
51	A	127,22	8,7	53,04	19,72	121,67	8,09	53,67	19,78
52	H	126,12	8,59	56,67	33,43	119,04	7,89	58,16	33,37
57	R			57,54	20,31			57,86	32,24
58	A	124,9	8,54	52,66	19,31	123,38	8,1	53,16	19,76
59	L	120,83	8,29	55,72	42,74	117,6	7,45	53,42	45,1

Table S2. FTIR contributions and assignments. Secondary structure assignments for selected time-points of a hCox17 refolding reaction at pH 8.4.

Assignment	0h		2.5h		6h	
	Band (cm^{-1})	%	Band (cm^{-1})	%	Band (cm^{-1})	%
Turns	1670.2	9	1671.3	7	1671	8
Turns	1660.2	5				
α-helix			1653.46	25	1653.46	44
Random coil/flexible regions	1642.8	86	1639.5	68	1640	48

Table S3. Disulfide pairing of hCox17 folding intermediate. Acid-trapped intermediate I was purified by RP-HPLC and alkylated by the addition of an excess of iodoacetamide solution in 50 mM ammonium bicarbonate buffer, pH 8. The alkylated intermediate was digested with trypsin with (+DTT) or without (-DTT) a previous reduction with DTT. MALDI-TOF mass spectrometry was used to analyse the tryptic digestion and identify each single resulting peptide; however, only those peptides providing evidence for the disulfide pairing assignment, are indicated. The identity of the identified peptides was further confirmed by TOF/TOF fragmentation analysis. While the treatment with iodoacetamide allowed the detection of free cysteines, the incubation with DTT allowed the detection of disulfide bridges. In the lower panel, we show a sequence map of the peptides used for disulfide identification in Peak I. In red cysteines that reacted with iodoacetamide, not forming disulfides in the intermediate, and in green cysteines in disulfide bonds (identified only after incubation with DTT). The only possibility is the presence of a Cys36-Cy45 disulfide bond.

Peptide Sequence	$[M+H]^+$ Obs.	$[M+H]^+$ Calc.	-DTT	+DTT	Cys in S-S bonds	Free Cysteines (CAM reactive)	Cys-Cys
[18-30] KPLKPC(CAM)C(CAM)AC(CAM)PETK	1588.81	1588.76	Y	N		23,24,26	
[18-30] KPLKPC(CAM)C(CAM)AC(CAM)PETK	1588.81	1588.76	N	Y		23,24,26	
[41-57] GEEHCGHLIEAHKEC(CAM)MR	2035.99	2035.88	N	Y	45	55	
[34-57] DACIIEKGEEHCGHLIEAHKEC(CAM)MR	2808.43	2808.26	N	Y	36,45	55	
[34-57] DACIIEKGEEHCGHLIEAHKEC(CAM)MR	2806.39	2806.25	Y	N		55	36-45
[34-53] DACIIEKGEEHCGHLIEAHK	2232.18	2232.05	N	Y	36,45		
[34-53] DACIIEKGEEHCGHLIEAHK	2230.13	2230.03	Y	N			36-45

Peptide Sequence	MPGLVDWNPAPPESQEKKPLKPCCACPETKKARDACIIEKGEEHCGHLIEAHKECMRALGFKI
[18-30]	KPLKPCCACPETK
[41-57]	GEEHCGHLIEAHKECMR
[34-57]	DACIIEKGEEHCGHLIEAHKECMR
[34-53]	DACIIEKGEEHCGHLIEAHK

Table S4. Sequence of the helix-designed mutants and their specific values of helical propensity according to AGADIR.

Helix 1	Sequence	helix propensity (%)
Wild Type	ACPETKKARDACIIEKGEE	2.27
Mutant H1	ACPE**E**KK**K**RDACI**L**K**K**GEE	13.63
Helix 2	**Sequence**	**helix propensity (%)**
Wild Type	HCGHLIEAHKECMRALG	0.47
Mutant H2	HCG**E**LI**QKY**KECMRALG	4.80
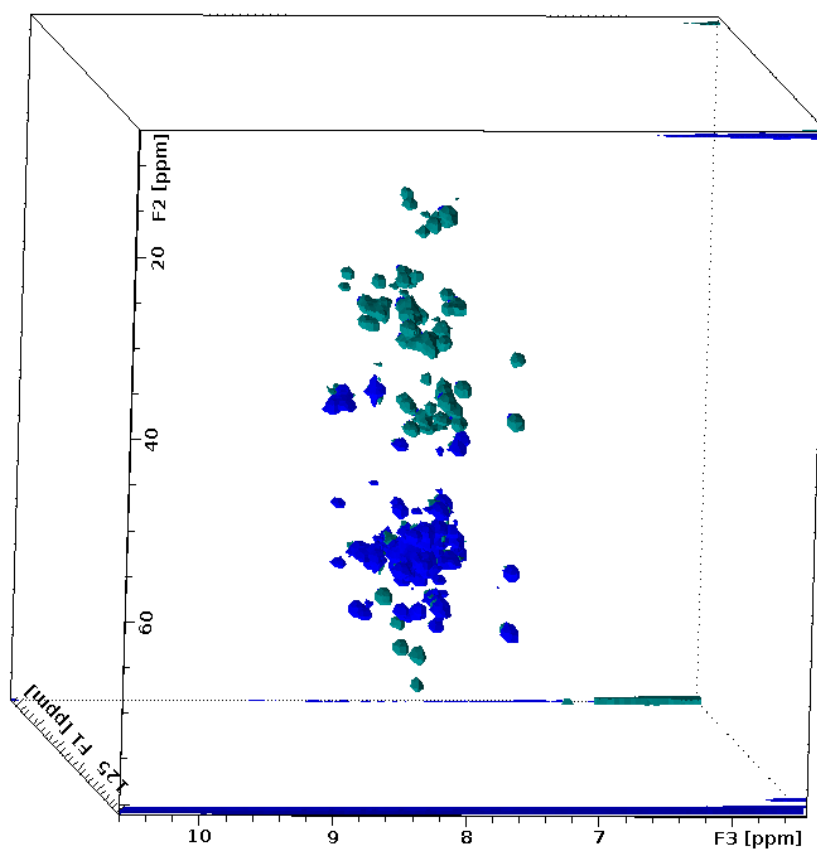
Figure S1. 3D cube representation of the HNCACB spectrum. Cα chemical shifts are represented in blue and Cβ are represented in green. Ala (15-25 ppm), Ser and Thr (55-75 ppm) are easily distinguished as shown in the image.
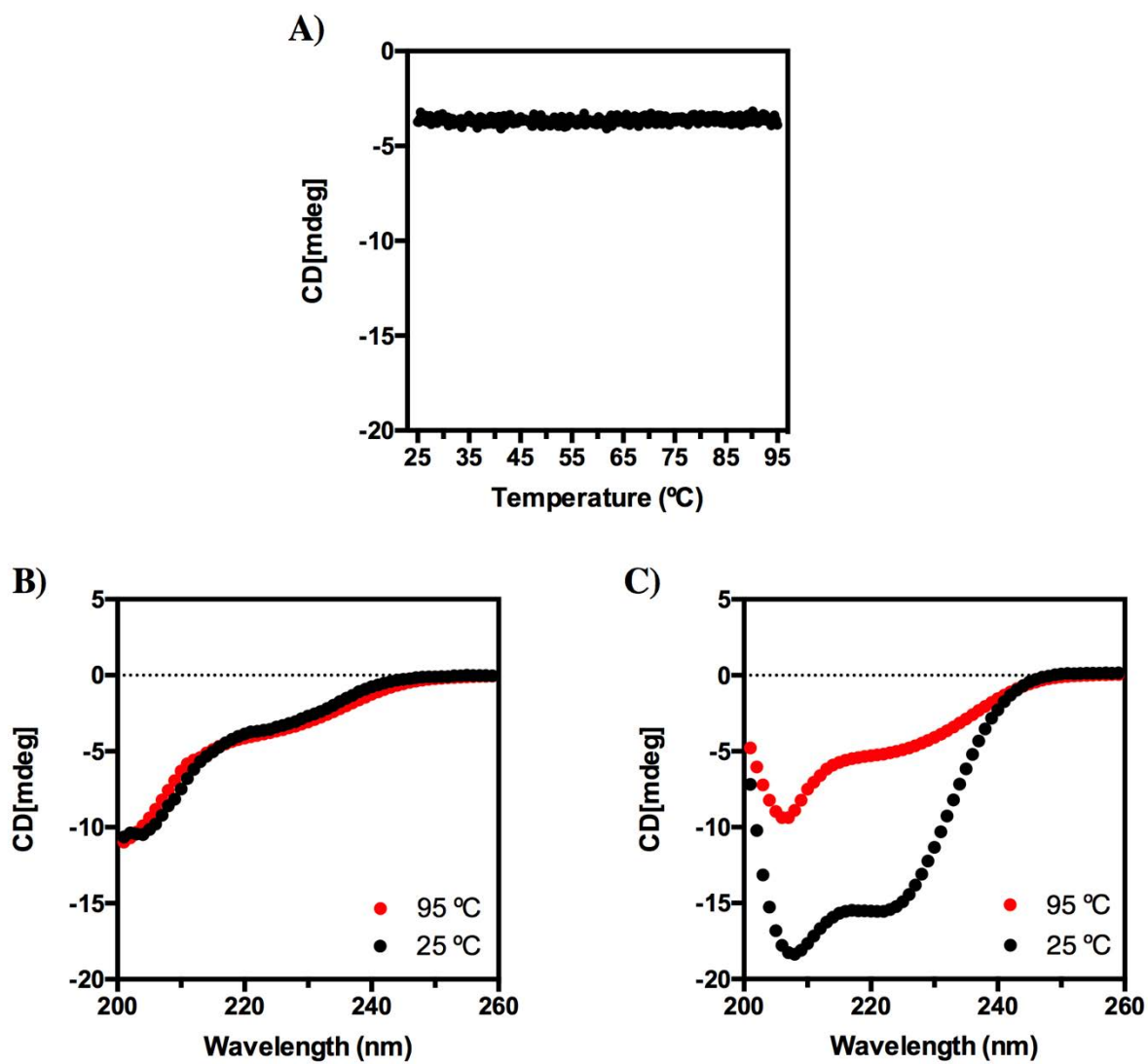
Figure S2. Thermal denaturation of hCox17. A) Thermal denaturation of reduced hCox17 followed by far-UV CD at 222nm. B) and C) Far-UV CD representative scans of reduced (B) and native (C) hCox17 incubated at 25ºC and 95ºC.
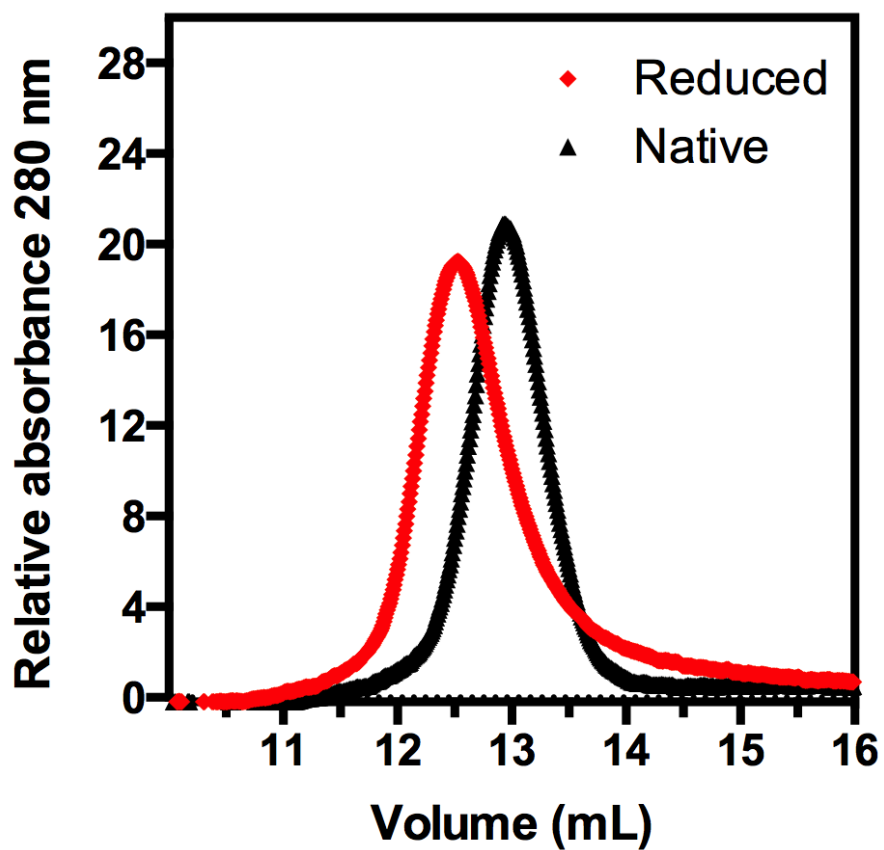
Figure S3. Hydrodynamic properties of hCox17. Reduced (red) and native (black) hCox17 were analysed by size exclusion chromatography (SEC) on a Superdex 75 column. Protein elution was monitored measuring absorbance at 280 nm.
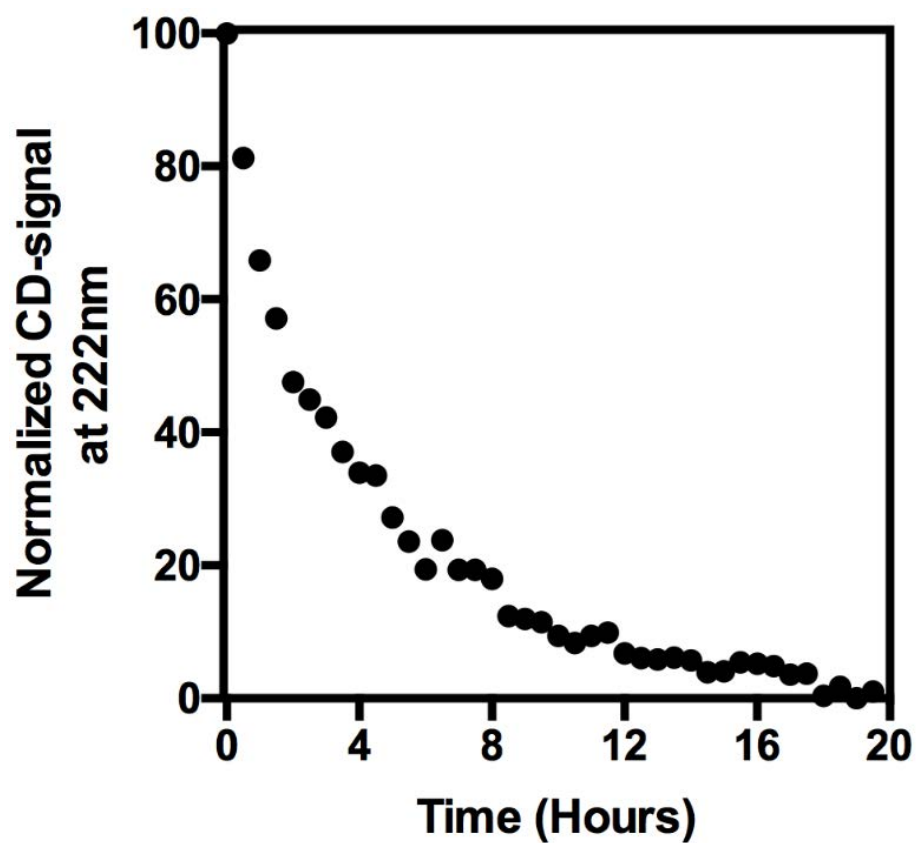
Figure S4. Refolding reaction of hCox17 in far-UV CD. Reduced hCox17 was allowed to refold at pH 7.0 inside a sealed CD cuvette and secondary structure formation was monitored using far-UV CD. Normalized signal at 222 nm is plotted as a function of time.
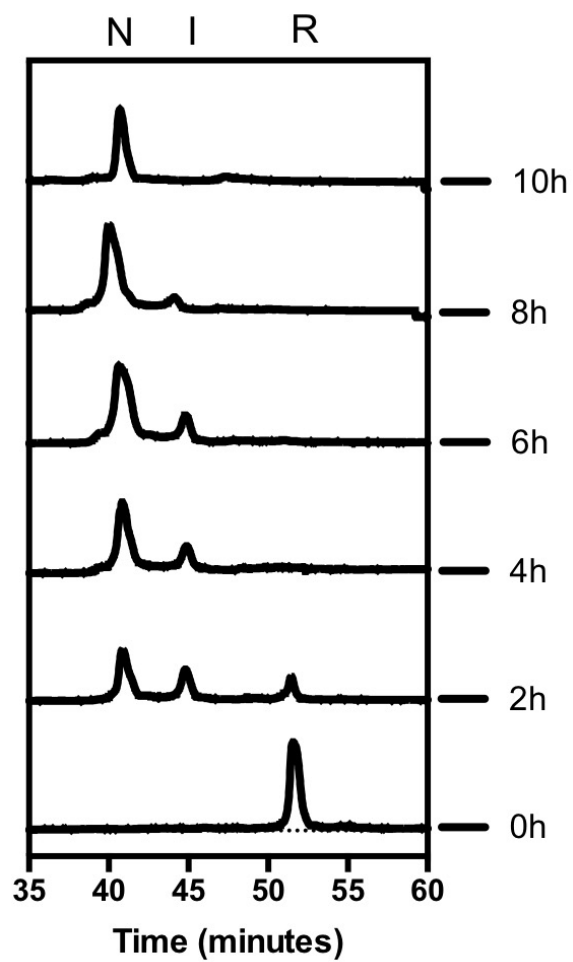
Figure S5. RP-HPLC analysis of a hCox17 refolding reaction at SAXS refolding conditions. Several RP-HPLC chromatograms corresponding to specific reaction times of hCox17 oxidative refolding at 3 mg/mL in Tris buffer pH 8.4.
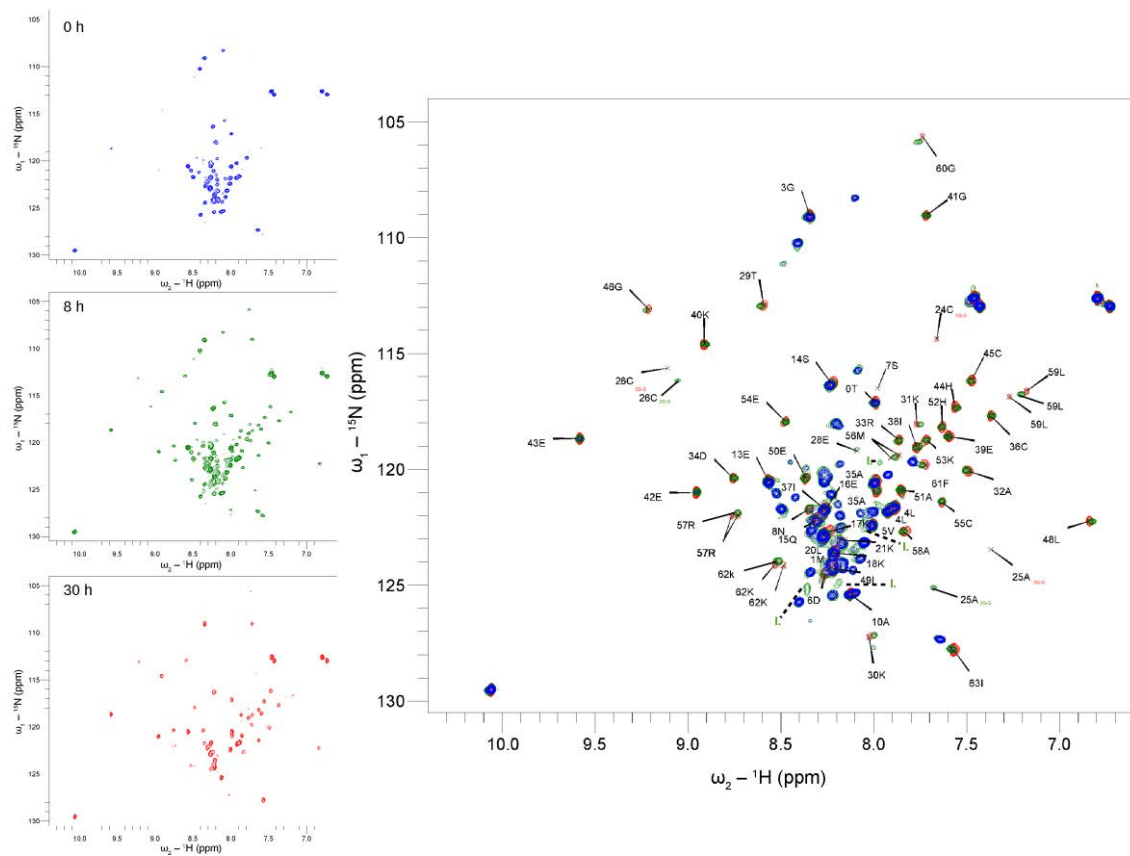
Figure S6. 1H-15N HSQC spectra of hCox17 folding process at different times. On the left different spectrum are shown, at 0 h (blue) at 8 h (green) and at 30 h (red). The central panel is an overlay of the three spectra. Chemical shifts are assigned for the N state of hCox17 protein. Some intermediate signals are indicated by green dashed lines (I).
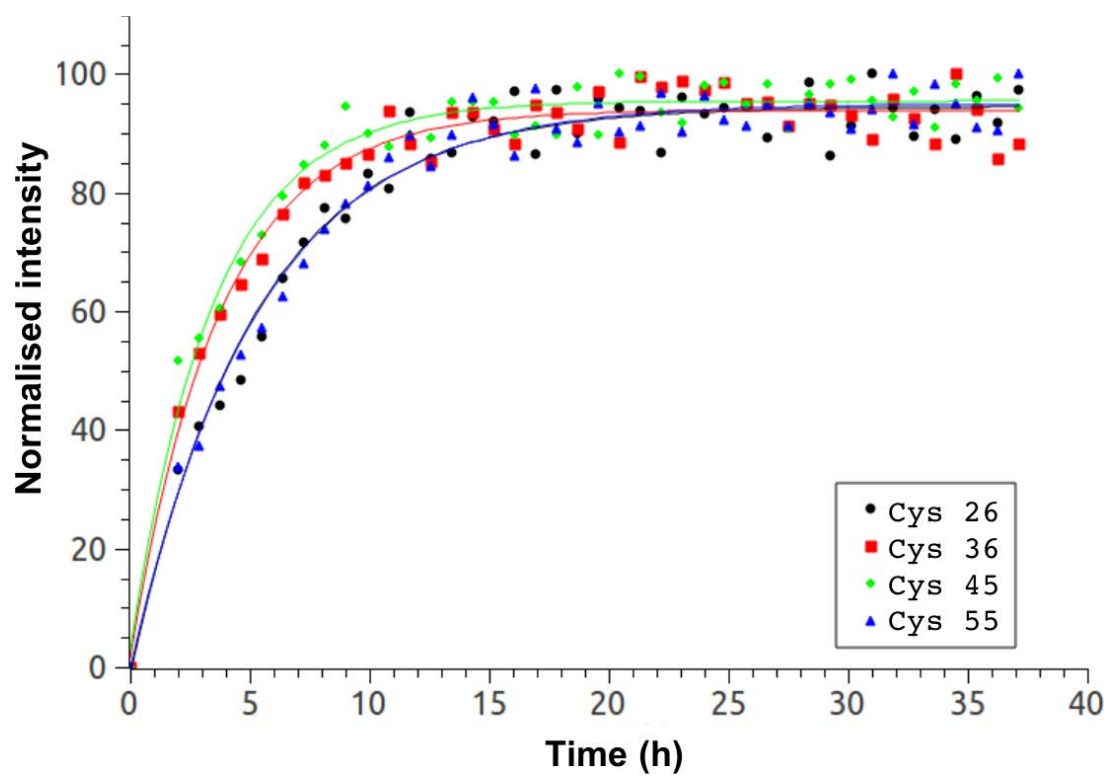
Figure S7. Reduced hCox17 was allowed to refold and HSQCs were collected along the reaction. Normalized chemical shift intensities of cysteines involved in hCox17 structural disulfides was plotted as a function of time.

Work 2

MIA40 funnels the oxidative folding of a

kinetically trapped molten globule

MIA40 bypasses the folding constraints

imposed by TRIAP1 function

Jordi Pujols[1,2], Marcos Gil-Garcia[1,2], Sebastián Esperante[3*] and Salvador Ventura [1,2*]

[1] Institut de Biotecnologia i Biomedicina. Universitat Autònoma de Barcelona, 08193-Bellaterra, Spain.

[2] Departament de Bioquímica i Biologia Molecular. Universitat Autònoma de Barcelona, 08193-Bellaterra, Spain.

[3] Fundacion INFANT. Gavilan 94, C1406. CABA, Buenos Aires. Argentina

[*] Correspondence:

sebandre@gmail.com;

salvador.ventura@uab.es; Tel.: (+34) 93 586 8956

Abstract

The MIA40 relay system is a dedicated mechanism that mediates small cysteine-rich proteins' import into the intermembrane mitochondrial space (IMS). MIA40 promotes the oxidation of critical native disulfide in its substrates to facilitate their folding. In this process, folding is simultaneously coupled to protein translocation and functional regulation. Mia40 substrates are initially synthesized in a reduced state in the cytosol and, despite sharing little or no homology, they have all been assumed to adopt a disordered conformation in this compartment, which should facilitate their translocation to the mitochondria. Here we characterize the oxidative folding of TRIAP1, an IMS protein that participates in the trafficking of phospholipids between mitochondrial membranes. Despite bearing a canonic twin CX9C motif, TRIAP1 exhibits a folding pathway that differs significantly from typical MIA40 substrates. TRIAP1 rapidly populates a molten globule in the reduced state, a feature that biases the folding reaction and makes it exceedingly slow. The folding of TRIAP1 is strongly constrained by the structural properties required to function in phospholipids traffic; however, the accumulation of a compact conformation in the cytosol could be functional on itself and connected with the p53-dependent cell survival pathway. The role of MIA40 consists in resolving the molten globule into a helix-loop-helix fold in the mitochondria, preventing the persistence of high kinetic barriers and therefore making available active TRIAP1 immediately upon its import into this organelle.

Introduction

Since the initial description of the oxidative folding pathways of ribonuclease A[1] and BPTI[2], many small disulfide-rich proteins have been investigated to determine the molecular determinants that guide protein folding and native disulfide bond formation. The slow kinetics of disulfide bond formation, together with the particular thiol chemistry, allowed the isolation and characterization of the major intermediates that populate the oxidative folding pathway of various proteins[3,4]. The spontaneous formation of native disulfide bonds is slow because it depends on adequate cysteines spatial positioning and requires an efficient final electron acceptor. Therefore, a dedicated oxidative folding machinery has been evolved to assist this process in-vivo. In eukaryotic cells, disulfide bond formation takes place mainly in the endoplasmic reticulum, catalyzed by the protein disulfide isomerase family (PDI) and Ero1p[5–7], and in the mitochondrial intermembrane space (IMS) by oxidoreductase chaperones, such as MIA40[8].

TP53-regulated inhibitor of apoptosis (TRIAP1) is a small disulfide-rich protein that targets the IMS and participates in the trafficking of phospholipids between the mitochondrial membranes[9]. Its native structure consists of an antiparallel helix-loop-helix connected by two intramolecular disulfides, referred to as the inner disulfide (Cys18-Cys37) and the outer disulfide (Cys8-Cys47), followed by a disordered C-terminal region[10]. Both α-helices are remarkably amphipathic and, as a consequence of the disulfide bridging, their hydrophobic faces become juxtaposed and partially exposed to solvent, which endorses the protein surface with an unusual non-polar character. In the IMS, TRIAP1 binds to distinct PRELI domain-containing proteins, termed PRELID1, PRELID2, SLMO1, and SLMO2, to form heterodimeric complexes that extract and transfer specific lipids from the outer membrane of the mitochondria (OMM) to the inner membrane of the mitochondria (IMM)[11]. Binding to TRIAP1 stabilizes PRELI-like proteins, which are unstable and rapidly degraded in their free state. TRIAP1 exposed hydrophobic cleft delineates the interacting surface with PRELID-like proteins[10]. PRELI-TRIAP1 is responsible for the biogenesis and accumulation of cardiolipin (CL) and phosphatidylethanolamine (PE) in IMM[12]. Of note, the CL and PE's asymmetric composition of mitochondrial membranes determines the organelle's morphology and

function. Perturbing lipid homeostasis and unbalancing CL levels result in mitochondria dysfunction and the activation of apoptotic signaling pathways[13,14].

In the cytosol, TRIAP1 has been associated with the p53-dependent cell survival pathway by inhibiting the formation of the apoptosome complex of Apaf-1, cytochrome c, and caspase-9 proteins[15]. In low DNA-damage cellular conditions, p53 upregulates the expression of TRIAP1 that, in complex with HSP70, hijacks Apaf-1, thus acting as an anti-apoptotic agent. On the contrary, the uncontrolled and p53-independent overexpression of TRIAP1 correlates with the onset and malignancy of some types of cancers[16,17]. In this later case, TRIAP1 behaves as an oncogene, becoming an active molecular agent of pathogenesis.

TRIAP1, like almost all mitochondrial proteins, is synthesized in the cytosolic ribosomes. Therefore, it must be imported to the IMS in order to carry out its function in lipid homeostasis. The translocation of disulfide-rich proteins to the mitochondria relies on the MIA40 pathway, which couples import with oxidation and folding[18]. TRIAP1 cysteines are encompassed in a highly conserved CX9C twin motif common to many MIA40 substrates. To be imported, disulfide-rich proteins must be in their reduced states. As they cross the mitochondria's outer membrane, MIA40 recognizes its substrates through hydrophobic interactions and introduces a mixed disulfide that pulls the polypeptide inside the IMS[19,20]. Therefore, MIA40 substrates are mostly reduced and disordered in the cytosol[21], and their premature oxidation, before arriving at the IMS, competes with their efficient import[22]. In this line of evidence, we recently reported that reduction and disorder are correlated for two different MIA40 substrates, Cox17 and Cox19[23,24]. This connection is not exclusive of MIA40 substrates, but rather a generic feature of small disulfide-rich proteins, which, as a trend, resemble bona fide intrinsically disordered proteins (IDPs) in their reduced states[25]. In this context, we hypothesized that TRIAP1 dual function in the cytosol and mitochondria stems from the population of two different redox and conformational states in these compartments.

In this work, we exploit the slow and disulfide-limited folding reactions of small disulfide-rich proteins to characterize the reduced state of TRIAP1. We also study its spontaneous and catalyzed oxidative folding reaction in the absence and presence of Mia40, respectively. We report that TRIAP1 behavior differs significantly from the one

of previously studied MIA40 substrates, and specifically from that of COX17 and COX19[23,24]. Instead of being disordered, the TRIAP1 reduced state is collapsed and contains residual secondary structure elements, resembling a molten globule. The population of this metastable conformation impacts the oxidative folding of TRIAP1, by promoting the accumulation of a kinetically trapped off-pathway intermediate and slowing down the reaction. Remarkably, MIA40 suffices to funnel the oxidative folding and promote an efficient and fast reaction. The present analysis provides new experimental evidence to explain the compartment-dependent function/malfunction of TRIAP1.
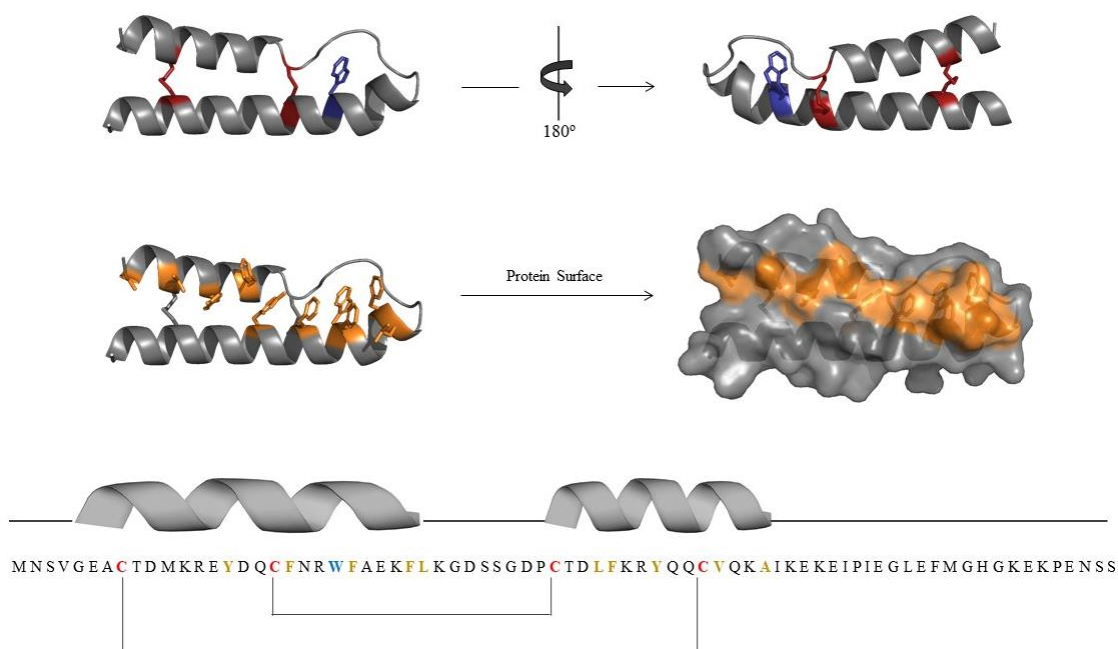

Figure 1. Schematic view of the three-dimensional structure of TRIAP1. a. Ribbon representation of TRIAP1. The cysteine residues and disulfide bonds (in red) and the tryptophan at position 22 (in blue) are indicated. b. Surface representation of residues Y15, F19, W22, F23, F27, F41 and Y44 with its aromatic side chains forming a hydrophobic solvent exposed surface c. Schematic representation of the primary structure of TRIAP1 with its secondary structure elements and disulfide pairing. The Protein Data Bank accession number for the structure of TRIAP1 is 4XZV. All figures were generated with PyMOL.

Results

TRIAP1 sequence intrinsic properties differ from that of canonic MIA40 substrates.

We have recently reported that COX17 and COX19, two CX9C substrates of MIA40, are fully disordered in their reduced states[23,24]. Upon oxidation, these proteins undergo a major disorder to order transition. Since protein disorder is imprinted in protein sequences, we first predicted the degree of disorder of TRIAP1 with the redox-sensitive algorithm IUPRED 2.0 and compared it with COX17 (Fig 2A). In good agreement, with their experimentally determined folded structures, the N- and C-terminus of COX17 and TRIAP1, respectively, are predicted as disordered in the oxidized state. In line with our previous data[23], the reduced state of COX17 is predicted to be fully disordered. In contrast, the reduced form of TRIAP1, does not possesses new extensive continuous disordered regions, with a much less dramatic difference in the predicted disorder for the reduced and oxidiced species than in the case of COX17. Conversely, TRIAP1 sequence has a much higher helical propensity than COX17, as predicted by the AGADIR algorithm. When estimated from the average values of the helices in the respective helix-loop-helix domains, the helix propensity of TRIAP1 is three-fold higher for helix 1 and twenty-four-fold higher for helix 2 (10.5 and 28.5) than in COX17 (3.5 and 1.2) (Fig. 2A).

One common feature of MIA40 substrates is that they display an overall moderate, hydrophobicity. Usually, one of the helices is significantly more hydrophobic since it incorporates the MISS/ITS signal that facilitates MIA40 recognition[26,27]. This helix is amphipathic, with the hydrophobic side becoming mostly protected from the solvent at the interface of the helix-loop-helix in the folded state. The other helix is less hydrophobic and has a low amphipathic character, as illustrated by COX17's first helix (Fig. 2B). Remarkably, both helices of TRIAP1 are highly amphipathic, with a particular enrichment in aromatic residues. Besides, this hydrophobicity remains significantly exposed to solvent in the folded state, resulting in a total area of 783.2 Å2, a 6-fold increase if compared to the 124.0 Å2 of COX17. Since TRIAP1 helices are longer and comprise ten extra residues, the corrected value of hydrophobic exposed area per residue is 19.1 Å2 for TRIAP and 4.0 for COX17 structures, still accounting for a > 4-fold increase.

The low propensity towards disorder, together with a significant ellipticity and hydrophobicity, indicates that TRIAP1 sequential intrinsic properties deviate from those of the canonic MIA40 substrates studied to date.
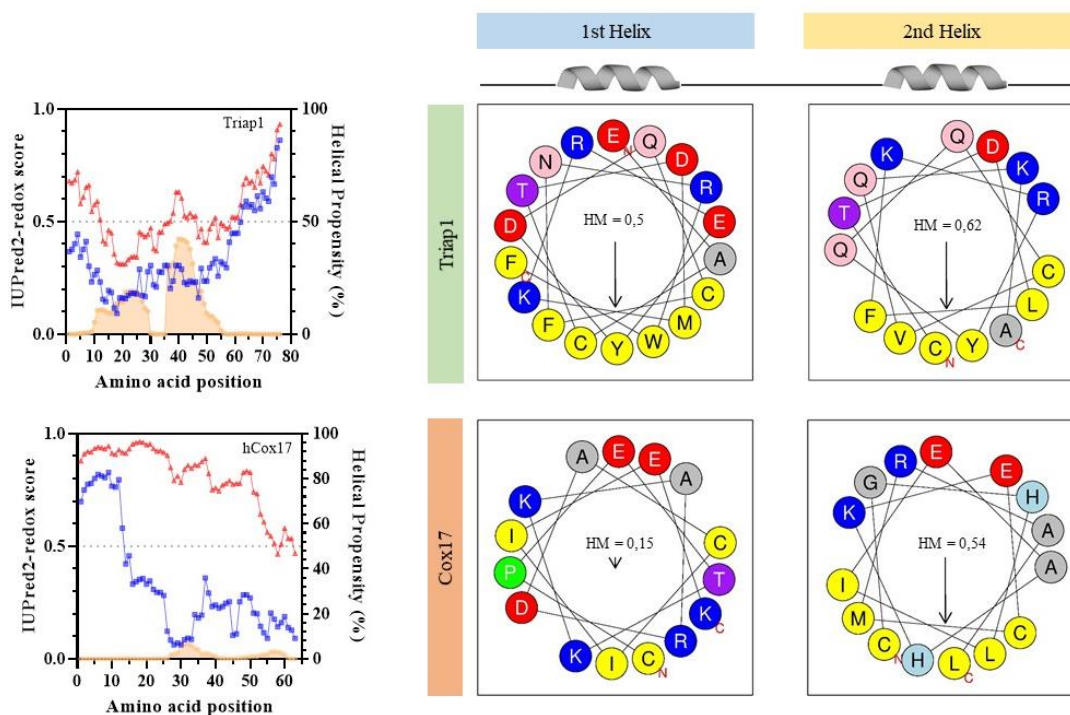

Figure 2. Predictions on TRIAP1 disorder and ellipticity. A) Disorder prediction of TRIAP1 and COX17 with the redox-sensitive algorithm IUPRED 2.0. Blue lines correspond to the wild type sequence, and red lines to the Cys to Ser versions that mimic the reduced state. Predictions of protein ellipticity provided by the AGADIR are plotted as orange areas. B) schematic representation of the amino acid spatial positioning for the first and second helices of TRIAP1 and COX17. Figures generated by the HeliQuest tool (PMID: 18662927). The Hydrophobic Moment (HM) is indicated for each helix.

Reduced TRIAP1 is partially folded and resembles a molten globule

Disulfide bonds stabilize the secondary and tertiary structure of small disulfide-rich proteins, and cysteine reduction is often coupled to unfolding, a process known as reductive unfolding[28]. The reductive unfolding reaction of TRIAP1 was studied by incubating the native protein with reductive reagents at pHs: 7.4, 8.4, and 5.0. The concentration of reductive reagents was selected to obtain the fully reduced protein within 2 h of reaction. The unfolding reaction was trapped in a time-course manner by acidification and analyzed by RP-HPLC (Fig. 3). At pHs 8.4 and 7.4, the unfolding

reactions were almost complete after 1 h of incubation, with a low accumulation of a single intermediate species (i2) eluted at time 33 min. The reductive reaction at pH 5.0 with TCEP minimizes disulfide reshuffling, and in this condition, intermediate i2 accumulates to a greater extend (Fig. 3). Alkylation of i2 with iodoacetamide and vinyl pyridine indicated that i2 corresponds to a 2-cysteine free intermediate.

Reductive unfolding was also followed by monitoring the changes in tertiary structure by tryptophan fluorescence and in helical secondary structure by far UV-CD at 222 nm. Disulfide reduction led to an increase in tryptophan fluorescence and a decrease in the α-helix content. The secondary and tertiary structural changes were completely linked at pHs 8.0 and 7.4, whereas both signals were somehow uncoupled at pH 5.0, with the tertiary structure changes slightly preceding those in secondary structure (Fig. 3). As predicted, fully reduced TRIAP1 still contained a significant amount of residual secondary structure, suggesting that the α-helices were not fully unfolded even when the protein was fully-reduced, according to RP-HPLC data (Fig. 2).
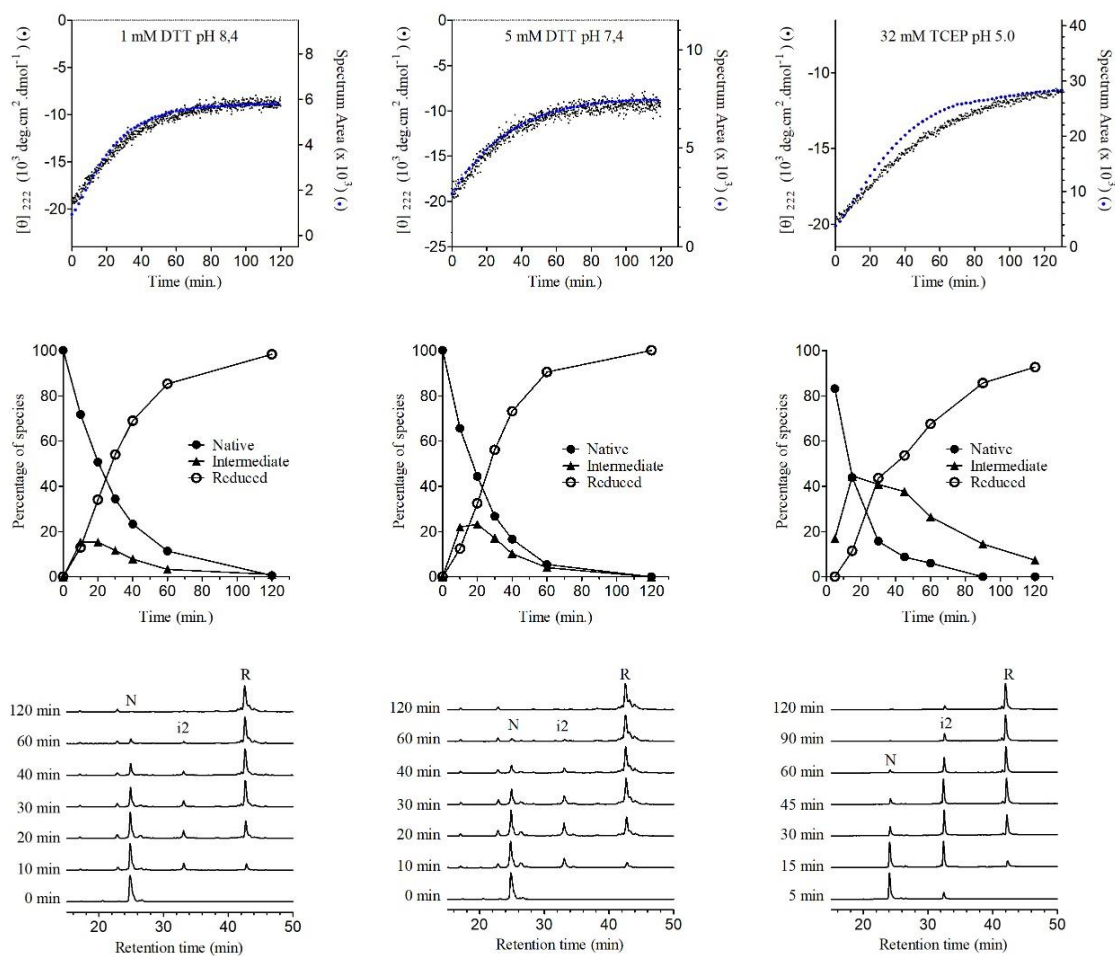
Figure 3. Conformational changes upon disulfide reduction of TRIAP1. Native TRIAP1 at 15 □M was treated with: (a) 1 mM DTT in 0.1 M Tris-HCl pH 8.4; (b) 5 mM DTT in 0.1 M sodium phosphate pH 7.4 and (c) 32 mM TCEP in 0.1 M sodium acetate pH 5.0. Upper panel: conformational changes upon disulfide reduction were followed by the changes in molar ellipticity at 222 nm (left Y-axis) and changes in the tryptophan spectrum area (right Y-axis) over time. Lower panel: After the addition of reducing agents, time course intermediates were trapped by acidification and analyzed by RP-HPLC. Native (N), intermediate (i2), and reduced (R) TRIAP1 peaks are indicated. Middle panel: The normalized peak area of species populating the reductive reaction is represented as a function of reduction time; reduced (○), intermediate (□), and native (●) species.

TRIAP1 harbors a single tryptophan at position 22 and high α-helical content in the native state[10]. To further characterize the fully reduced and the fully reduced/unfolded protein, native TRIAP1 samples were incubated for 16 h at pH 7.4 under non-denaturing conditions or in the presence of 6.0 M urea or 6.0 M Gdm.HCl. The samples were also incubated with 2 mM DTT to induce disulfide reduction or without reductive reagent. The tryptophan fluorescence emission spectrum of TRIAP1 in the absence of denaturant showed that its single tryptophan was partially exposed to the solvent with a maximum emission wavelength of 336 nm (Fig. 4A). When the protein was reduced, the fluorescence intensity increased by 10-fold, suggesting that the disulfide bonds were quenching the tryptophan fluorescence, as was described for other small disulfide-rich proteins (22-24). The fully reduced TRIAP1 exhibited a maximum emission wavelength of 340 nm, indicating a similar tryptophan environment compared to the native protein. The far UV-CD spectrum of the fully reduced protein showed that, as described above, it contained a significant amount of residual secondary structure content (molar ellipticity at 222 nm = -10.400 deg.cm2.mol-1) (Fig. 4B). The disulfide cross-linked TRIAP1 incubated in 6.0 M urea showed secondary structure and Tryptophan emission spectrum close to the native species, while incubation in 6.0 M Gdm·HCl, partially unfolded TRIAP1, albeit still displaying a substantial amount of α-helical content (222 nm = -10.000 deg.cm2.mol-1). The addition of 2 mM DTT in the presence of 6.0 M urea or 6.0 M Gdm·HCl led to the complete loss of the secondary structure, whereas the tryptophan maximum emission wavelength shifted to 356 nm, indicating a complete exposure of the tryptophan to the solvent. These results confirmed that reduced TRIAP1, in the absence of denaturant, retains a significant amount of secondary and tertiary structure. After far-

UV CD and fluorescence measurements, the samples were acid-quenched and subjected to RP-HPLC. This allowed the confirmation that the protein samples incubated in different conditions were either fully oxidized or fully reduced, as planed (Fig. S1).

To study if the TRIAP1 conformers population in the different conditions correspond to collapsed forms where hydrophobic residues are packed in structural clusters and accessible to solvent, we monitored their bis-ANS binding (Fig. 4C). Under denaturing conditions, the protein did not induce any significant bis-ANS fluorescence, both in the presence and absence of DTT. Thus, no packing of hydrophobic residues occurs in these conditions. Remarkably, the reduced but not denatured TRIAP1 showed increased fluorescence if compared to the oxidized and properly folded conformation, which as expected from its structure, already exhibited significant binding. These results indicate that the partially folded species of TRIAP1 corresponds to a metastable structure where hydrophobic residues pack to each other and are less protected from the solvent than in the native structure.
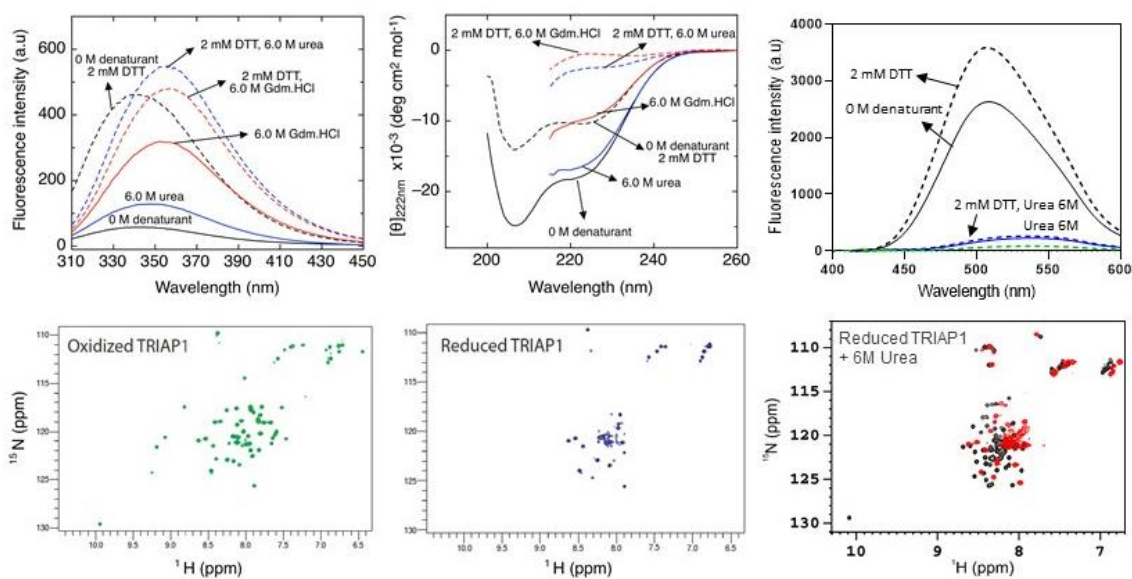

Figure 4. Conformational properties of the fully reduced and fully-reduced/unfolded TRIAP1. 15 μM of TRIAP1 samples were incubated at RT for 20 h in: 50 mM sodium phosphate pH 7.4 with: 0 M denaturant; 6.0 M urea or 6.0 M Gdm.HCl in the absence or presence of 2 mM DTT. After incubation, samples were analyzed by tryptophan fluorescence (a), far-UV CD (b) and binding to bis-ANS (c), green as a control of free bis-ANS. The spectrum of each sample is indicated in the graph. The CD spectra are represented as molar ellipticity as a function of wavelength. After measurements, the samples were acid-quenched and subjected to RP-HPLC to confirm the fully reduced or oxidized state (Fig. S1). (c) 15 N

HSQC NMR spectra of the oxidized and reduced TRIAP1 samples at 0.3 mM in 20 mM sodium phosphate pH 7.0, 10% D20.

To gain further insights into oxidized and reduced TRIAP1 conformations, we labeled the protein and recorded their 15N-HSQC NMR spectra. In the oxidized state, amide proton chemical shifts are well dispersed, indicating that the protein is folded (Fig. 4D). The addition of DTT and the subsequent reduction of disulfides cysteines provoked the disappearance of several peaks that can be attributed to native contacts absent in this specie. In the spectra of reduced TRIAP, the detected peaks are primarily clustered between 8 and 8.5 ppm, usually assigned to disordered conformers (Fig. 4E). However, we could not detect the appearance of new peaks associated with those amino acids in a new chemical environment in the reduced state, which suggests that the polypeptide chain is not entirely disordered. Instead, the reduced protein resembles an intermediate state where certain protein regions are in conformational exchange on the millisecond time scale and give rise to extreme line broadening and peak merging in the NMR spectrum. This assumption was further confirmed when TRIAP1 15N-HSQC NMR spectra were recorded in the presence of DTT and urea 6 M (Fig. 4F). The combination of disulfide reduction and chemical denaturation results in a clean NMR spectrum with the appearance of well-defined peaks between 8 and 8.5 ppm that correspond to a more disordered conformation (black) if compared to reduced but not denatured TRIAP1 spectra (red).

Finally, we performed thermal denaturation experiments following changes in molar ellipticity at 222 nm. Reduced TRIAP1 gradually lost its secondary structure upon heating displaying a low but still significant cooperativity (Fig. S2), confirming that reduced TRIAP1 is at least partially packed. As expected, when native disulfide cross-linked TRIAP1 was subjected to thermal denaturation analysis, it showed little changes in the secondary structure content as its fold is strongly stabilized by the covalent links (Fig. S2). Thermal denaturation reactions of native and reduced TRIAP1 were fully reversible, and the spectrum of the reduced sample at 85ºC showed that the minimum at 208 nm, characteristic of α-helical conformation, shifted to lower wavelengths, indicative of an order to disorder transition upon thermal denaturation (Fig. S2).

Overall, the results in this section indicate that, in contrast to previously characterized MIA40 substrates, the reduced state of TRIAP1 resembles a molten globule

conformation; delineated by the presence of secondary elements in a collapsed, but dynamic state.

TRIAP1 air-oxidizes to spontaneously attain its native folded state

We have previously demonstrated that MIA40 substrates can fold in the absence of oxidoreductases by air-oxidization[23,24]. Purified TRIAP1 was fully reduced and unfolded by incubation with 200 mM DTT and 6.0 M Gdm.HCl. Then, the reduced and unfolded protein was allowed to refold at pH 8.4 in the presence of different redox agents, and folding intermediates were acid-trapped with TFA 1% at selected time points and subjected to RP-HPLC. The heterogeneity and chromatographic behavior of folding intermediates are shown in Figures 5 and Supplementary figure 3. In the absence of redox reagents (Control -), there was a low accumulation of two folding intermediates, the first one at 37.5 min (i1), accumulated to a greater extent and showed a retention time closer to the reduced species (42 min), suggesting a high exposure of hydrophobic residues. The second intermediate, with a retention time of 33 min (i2) corresponds to the only intermediate populated in the reductive unfolding. It showed a low accumulation along with a slow refolding reaction, which was completed after 24 h incubation. At 36 h and 48 h incubation, native TRIAP1 was in equilibrium with 3 other minor species that were not populated previously along the refolding reaction. These species were not present when either freshly purified native protein or TRIAP1 incubated 24 h in 50 mM sodium phosphate pH 7.4 were monitored by RP-HPLCE, indicating that they correspond to off-pathway trapped species. The presence of 0.25 mM 2-mercaptoethanol (Control +), which promotes disulfide rearrangement in oxidized species, neither affected the overall chromatographic pattern nor accelerated the folding reaction, suggesting that disulfide isomerization was not a critical step in the formation of the two native disulfide bonds.
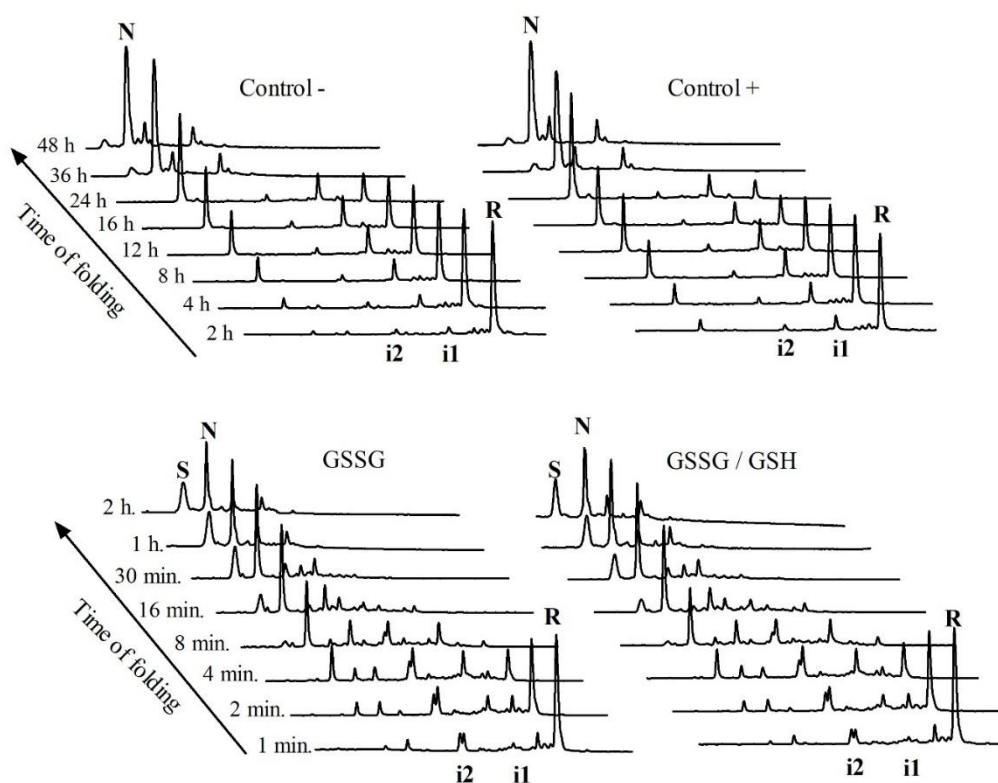
Figure 5. RP-HPLC analysis of TRIAP1 refolding intermediates. In-vitro refolding was carried out in Tris.Cl buffer (pH 8.4) in the absence (Control -) and presence of selected redox agents: 0.25 mM 2-mercaptoethanol (Control +), 0.5 mM GSSG or a mixture of 0.5 mM GSSG and 1 mM GSH. Intermediates were acid-trapped at the noted times and analyzed by RP-HPLC. The native (N), fully reduced/unfolded (R), scrambled species (S), as well as the two major intermediates (i1 and i2) are indicated. A detailed comparison of RP-HPLC chromatographic profiles is shown in Figure S1.

Oxidative folding of TRIAP1 was subsequently performed in the presence of 0.5 mM oxidized glutathione (GSSG) and a mixture of reduced and oxidized glutathione (GSH:GSSG, 1.0:0.5 mM, respectively). The presence of GSSG accelerated the folding reaction that was almost complete after 15 min. The RP-HPLC chromatographic patterns obtained in both redox conditions were similar and the folding reaction progressed at similar rates. Intermediates i1 and i2 were also present, together with the population of other minor species and, in contrast to the folding reaction in the absence of redox reagents (Control -), intermediate i2 accumulated to a greater extent. Noteworthy, intermediate peak i2 displayed a heterogeneous profile containing at least two species with similar retention times (Fig. S1). After 15 min, the native peak was in equilibrium with other species, as was shown above. Moreover, after 1 and 2 h of refolding, a species

with a lower retention time (22 min) than the native peak accumulated, we speculate that it corresponds to a compact scrambled species that exposes less hydrophobic surfaces than the natively cross-linked form (See figure 5 and S1). Overall, the acceleration promoted by GSSG seems to stem from a greater extend of stochastic oxidation with the population of species that are not necessarily productive.

The secondary and tertiary structural changes of TRIAP1 along the folding process (Control -) were monitored by far-UV circular dichroism (CD) and tryptophan fluorescence emission spectra over time (Figure S2). The changes in both spectroscopic signals overlapped along time indicating that, in TRIAP1, secondary and tertiary structure formation are coupled (Fig. S4).

TRIAP1 molten globule impacts the folding pathway of TRIAP1 and promotes the accumulation of a kinetically trapped intermediate

To test whether the population of the TRIAP1 molten globule impacts the folding reaction, TRIAP1 was reduced with 200 mM DTT or reduced and unfolded with 200 mM DTT and 6.0 M Gdm.HCl. Then, the samples were buffer exchanged and allowed to refold at pH 8.4, in the absence or presence of 0.25 mM β-mercaptoethanol. Surprisingly, the reduced only sample attained its native disulfide bonds after 16 h incubation, whereas for the reduced and unfolded sample it took more than 24 h (Fig. 6). Thus, the molten globule compactness facilitates subsequent oxidative folding.

In order to understand the role of the molten globule in guiding TRIAP1's folding, we performed refolding experiments of fully reduced and unfolded TRIAP1 under denaturing conditions. First, we assessed the conformational stability of TRIAP1 using the technique of disulfide scrambling. In this method, native disulfide-rich proteins are incubated with denaturants in the presence of a thiol initiator. As a consequence of their partial unfolding, they might reshuffle their native disulfide bonds and form disulfide-scrambled species (25,26). Native TRIAP1 was incubated in the presence of 0.25 mM β-mercaptoethanol as a thiol initiator and increasing concentrations of urea, Gdm.HCl or guanidine isothiocyanate (Gdm.SCN) at pH 8.4. The protein samples were allowed to reach equilibrium for 20 h, trapped by acidification, and analyzed by RP-HPLC (Fig. S5). As observed previously for other small disulfide-rich proteins (27), denaturation was dependent on denaturant strength. Urea at 8.0 M was unable to promote TRIAP1

unfolding, while above 6.0 M Gdm.HCl or 4.0 M Gdm.SCN, a major scrambled species with a higher retention time than the native protein accumulated.
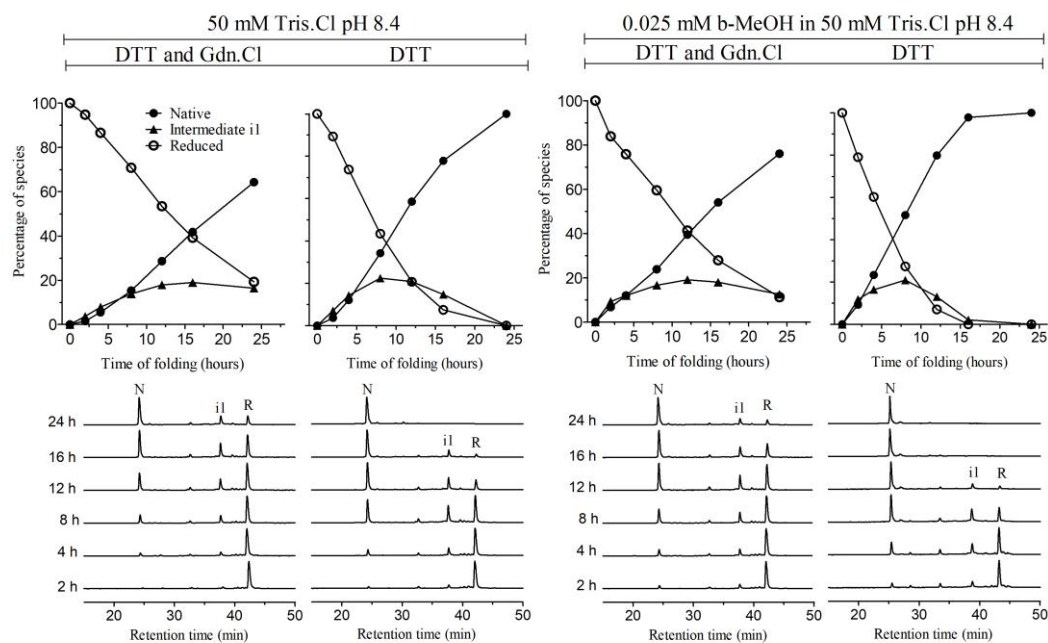

Figure 6. Oxidative folding progression analysis by RP-HPLC of reduced, and reduced-unfolded TRIAP1. TRIAP1 was reduced with 200 mM DTT or fully reduced and unfolded with 200 mM DTT and 6.0 M Gdn.Cl. In-vitro refolding of the samples was carried out in 0.1 M Tris.Cl buffer (pH 8.4) in the absence (Control -) and presence of 0.25 mM 2-mercaptoethanol (Control +). The reactions were acid-trapped at the noted times and analyzed by RP-HPLC. The native (N), reduced (R), and the major intermediate (i1) peaks are indicated. The normalized peak area of species populating the reaction is represented as a function of refolding time; reduced (○), intermediate i1 (ê), and native (●) species.

We selected Urea 6M to perform the refolding experiment in the presence of denaturant; since Triap1 native state was still the predominant species and, accordingly, the more stable conformation in that condition. Besides, as shown in figure 4, 6M Urea impedes the population of the molten globule in reduced Triap1'. Surprisingly, TRIAP1 refolding under denaturing conditions proceeded 2-times faster than in the absence of Urea 6M (Fig. 7). The refolding reaction was almost completed after 12 h incubation, and the identity of the intermediates populated along the reaction changed. Interestingly, the intermediate i1, which was the limiting species under non-denaturing conditions, was completely absent in the presence of 6.0 M urea, indicating that conformational bias favors the accumulation of this disulfide species in the absence of urea, likely determined

by the molten globule properties. In contrast, intermediate i2 accumulated to a greater extent than in non-denaturing conditions, together with some other minor intermediate species. When refolding was performed in 6.0 M urea in the presence of 0.5 mM GSSG, the reaction was drastically accelerated and completed after 16 min. Intermediate i2 was the main intermediate and a peak with a retention time close to the native two-disulfide peak accumulated along with the reaction and disappeared after 30 min incubation. This peak probably consisted of a scrambled species with compactness close to the native protein that isomerizes its disulfides to attain the energetically favorable two native disulfide species. These results suggest that the molten globule of reduced Triap1 facilitates the formation of i1, which is both an off-pathway intermediate and a kinetic trap.
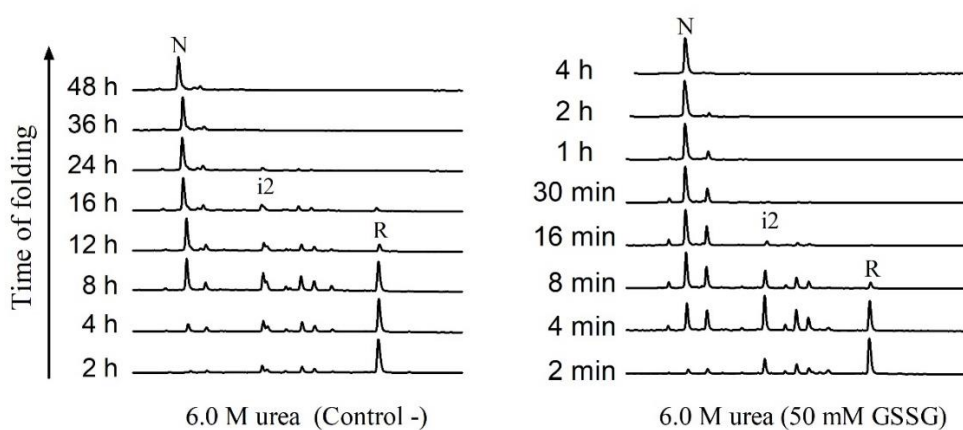

Figure 7 *Oxidative folding of TRIAP1 in the presence of 6.0 M urea.* The fully-reduced/unfolded protein was allowed to refold in 0.1 M Tris-HCl pH 8.4 in the presence of 6.0 M urea in the absence (left panel) or presence of 1 mM GSSG (right panel). Intermediates were acid-trapped at the noted times and analyzed by RP-HPLC. The elution positions of native (N), reduced (R), intermediate (i2), and scambled species (S) are indicated.

MIA40 funnels the oxidative folding of TRIAP1, bypassing the kinetic trap

MIA40 catalyzes the formation of native disulfide bonds to newly imported polypeptides in the IMS (8). To gain insights into the process that occurs in living cells, we next analyzed the effect of MIA40 on the oxidative folding of TRIAP1. Fully reduced and unfolded TRIAP1 was allowed to refold in the presence of two-fold molar excess of

Mia40 at two different pHs (pH 7.0 and pH 8.4). The samples were acid-trapped at selected time points and subjected to RP-HPLC to characterize the folding intermediates (Fig. 8). Mia40 catalyzed the rapid formation of intermediate i2, suggesting that i2 contains a native disulfide bond, likely the inner one (Cys18-Cys37). This major intermediate species represented up to 50% of the species populating the reaction in the first minute of folding at pH 8.4. At pH 7.0, the folding reaction progressed slightly slower, and intermediate i2 reached a maximum amount within 5-8 min of folding. Interestingly, at the end of the reaction, a major scrambled species accumulated, mainly at pH 7.0, with a retention time longer than native TRIAP1. Remarkably, i1 was not accumulated, confirming that it corresponds to an off-pathway intermediate. The rapid and specific formation of the productive intermediate i2 would preclude the formation of non-native disulfides, bypassing the conformational bias and kinetic impediment imposed by the propensity of TRIAP1 to form a molten globule.
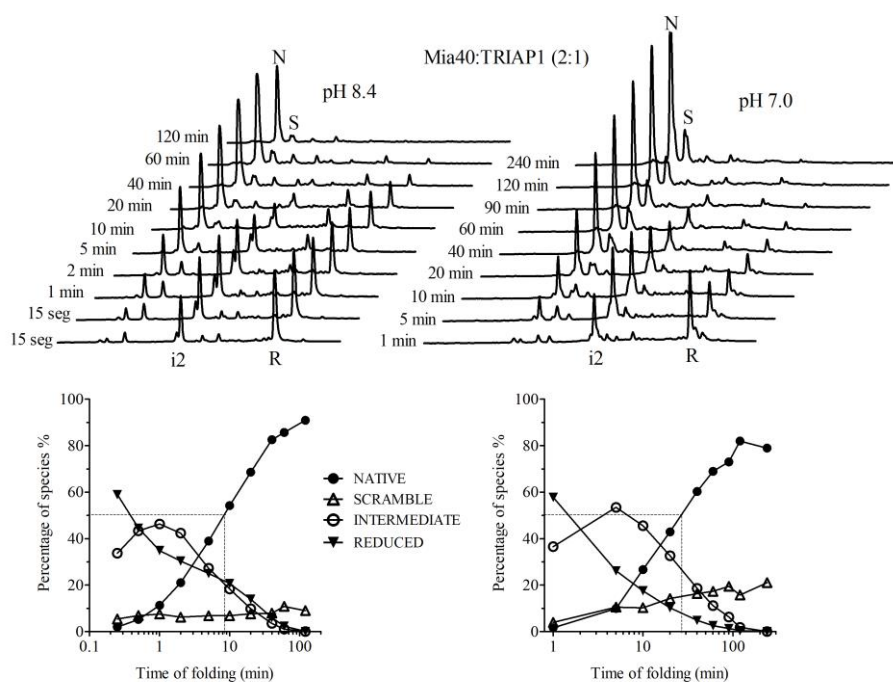

Figure 7. *In-vitro oxidative folding of TRIAP1 catalyzed by Mia40.* GST-tagged Mia40 was added to TRIAP1 at a 2-fold molar excess in 50 mM Tris.Cl, 0.1 M NaCl pHs: 8.4 (left panel) and 7.0 (right panel). The samples were acid-quenched at the indicated time points and subjected to RP-HPLC. The retention time of peak i2 coincides either with that obtained from spontaneus oxidative folding of TRIAP1 or with the intermediate peak of disulfide reduction at pH 5.0. The major peaks were integrated and plotted as a function of folding time (lower panel). The peaks integrated are: native (N), reduced (R), intermediate i2 (i2) and scramble (s).

Discussion

Since the seminal work of C. Anfinsen with RNase, the study of disulfide-rich proteins has provided data to advance on our knowledge on the molecular principles that govern protein folding[1]. They have been useful models to overcome methodologic limitations, such as the inaccessibility to unfolded states in native buffering conditions or the transient nature of intermediates in folding reactions[29]. The appearance of new concepts like protein disorder, conditional folding, and functional moonlighting, suggested that apart from the well-stablished role of disulfide bonding in defining and stabilizing the native structure, they may impact physiologic processes acting as inducible and reversible PTMs[30,31]. Several elegant works point towards this direction, with disulfide-rich proteins shifting between active and inactive conformations in a redox controlled-manner[32,33].

The IMS's disulfide-rich proteins illustrate well this mechanism since they populate two different and equally relevant conformations in the cell[18]. A reduced import-competent state and an oxidized folded and functional structure. For these proteins, synthesized in the cytosol, oxidation is compartment-specific and promoted by the specific binding and catalysis of MIA40. Since this dual character is imprinted in the primary sequence, it can be studied *in vitro*. Accordingly, we previously characterized the disorder-to-order transition of two MIA40 substrates, COX17 and COX19[23,24].

In this work, we characterize the oxidative folding of TRIAP1. We report that TRIAP1 is far from behaving as a typical CX_9C MIA40 substrate. TRIAP1 has an unusual hydrophobic load if compared with other MIA40 substrates. Reduced TRIAP1 collapses into a compact conformation with a significant content of secondary structure elements and hydrophobic patches exposed to solvent. NMR data showed that this metastable conformation is also dynamic. Altogether, these properties indicate a collapsed early-stage intermediate, classically referred to as molten globule. Folding intermediates cannot be studied by standard biophysical methodologies at physiological pH, essentially because they are formed transiently in this condition. However, TRIAP1 molten globule is stable and endurable in the presence of DTT, which will allow its further detailed structural characterization.

We demonstrate that the population of this molten globule impacts the folding pathway of TRIAP1. MIA40 substrates usually follow simple folding pathways, dominated by a single intermediate that contains the inner native disulfide bond. Instead, TRIAP1

populates an additional intermediate (i1) that in the RP-HPLC elutes close to the reduced species. Although i1 requires in-depth biophysical analysis, plus the identification of the cys involved in the disulfide bond, we hypothesize that i1 corresponds to a non-native intermediate with one disulfide bond. Since i1 accumulates first in the reaction it likely arises from the conformational bias imposed by the reduced molten globule state. We allowed TRIAP1 to fold under denaturing conditions, in which the molten globule cannot longer accumulate. In urea 6M, TRIAP1 folds at much higher rate, without i1 accumulation. Therefore, the molten globule promotes the formation of a non-native intermediate that acts as a kinetic trap of the spontaneous reaction in native conditions.

TRIAP1 can fold into its native state by simple air-oxidation, although in an extremely slow reaction. The addition of MIA40 funnels the oxidative folding and increases the folding rate of TRIAP by 30-fold. Similar to the urea 6M condition, MIA40 abolishes the formation of the i1 kinetic trap, favoring the direct formation of the productive i2 species. MIA40 catalyzes the oxidation of a large pool of different proteins. Even though some of them share a common helix-loop-helix fold, they execute different cellular functions. Thus, they are shaped by different evolutionary constraints. TRIAP1 requires the presence of an exposed hydrophobic cleft to stabilize its partners, the PRELI-like family. Phospholipid distribution is not restricted to the IMS but is common in other membranes and organelles[34]. PRRELI-like proteins share a striking structural homology with another member of the StARkin family, the phosphoinositol transfer protein α (PTIP-α)[10]. The structure of the PRELI-like family lacks the stabilizer and regulatory C-terminal α-helix of PTIP-α, which allows the structure to accommodate the phospholipid cargo by conformational changes. The TRIAP1 interaction substitutes the missing helix and binds, with the equivalent topology, the exposed region of PRRELI-like proteins. It would be interesting to decipher, whereas the PRELI-TRIAP complex responds to structural convergence events or responds to an ancestral sequence's C-terminal a-helix domain deletion.

In this scenario, the hydrophobic load of TRIAP1 is evolutionary conserved and responds to a physiologic requirement. In biologic systems, the need for exposed hydrophobicity to establish intra- or inter-molecular interactions comes at the expenses of an inherent propensity to aggregate since the same physicochemical processes govern both processes. TRIAP1 should have evolved to avoid spurious inter-molecular contacts that might ultimately drive to aggregation. Since TRIAP1 remains exposed in the cytosol before its import and oxidation, we hypothesize that its prominent ellipticity and the formation of

the molten globule might, somehow protect TRIAP1 from aggregation. Indeed, the formation of helical conformations constitutes an evolutionary strategy of proteins to cope with b-sheet aggregation and amyloidogenicity. In addition, the disordered C-terminal tail, of unknown function, would act as an entropic bristle, increasing the solubility of the polypeptide since 46 % of its residues are charged and has a predicted high solubility (GRAVY = -1.33). Indeed, this tail is more soluble that the disordered N-terminal region of Cox17 (GRAVY = -0.74) or the N- (GRAVY = -0.67) and C- (GRAVY = -0.81) terminal extensions of Cox19, suggesting and adaptative role to protect TRIAP1 against aggregation as it comes out from the ribosomes in its reduced form in the cytosol. Indeed, in contrast to most disordered regions, the sequence of the C-tail of TRIAP1 is well conserved among vertebrates.

Cellular protective systems such as the quality control machinery and chaperoning stand as alternative strategies prevent aggregation, refold misfolded species or promote their degradation. Among them, the heat-shock protein 70 (HSP70) recognizes partially folded species by binding to abnormally exposed hydrophobic regions. Since the cytosolic function of TRIAP1 results from an interaction with HSP70, it is logic to assume that the reduced molten globule is actively involved in the p53-dependent cell survival pathway[15], although this should be confirmed by studies *in vivo*.

Last but not least, it is important to note that the packed conformation of TRIAP1 might have implications regarding the disulfide relay system itself. Canonic CX_3C and CX_9C MIA40 substrates require fully disordered states so they can cross the outer membrane and interact with MIA40, which promotes a folding upon binding to its substrates. Recent evidences highlight that MIA40 is a versatile oxidoreductase that accommodates multiple topologies and promotes the disulfide bridging of several non-canonic substrates[35]. TRIAP1 seems to be a comprehensive model system to understand the functioning of the disulfide relay system in these special cases.

Experimental procedures

Protein expression and purification

Human TRIAP1 cDNA was cloned into pET 28b(+) vector as C-terminal 6xHis-tagged fusion protein and the resulting plasmid was sequenced and transformed into E. coli Origam2 cells. Cells were grown in LB broth at 37 °C and protein expression was induced at DO_{600nm}= 0.8 with 1 mM IPTG and the culture was incubated ON at 20 °C. Cells were harvested by centrifugation, the pellet was resuspended in 0.1 M Tris.HCl pH 8.0, 0.5 M NaCl and lysed by sonication. The supernatant was purified using a nickel-affinity chromatography followed by a size exclusion chromatography in a HiLoad 26/60 superdex 75 prep-grade column (GE healthcare) pre-equilibrated in 20 mM Tris.HCl pH 8.0 and 0.1 M NaCl. The purity of the isolated TRIAP1 protein was determined by Tris-Tricine/SDS-PAGE, RP-HPLC and matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry (MS). The concentration of purified TRIAP1 was determined by measuring the absorbance at 280 nm using a molar extinction coefficient of 8730 $M^{-1}cm^{-1}$. The catalytic Mia40 C-terminal domain was purified as an N-terminal GST fusion protein from Escherichia coli cells as was previously described (28).

Oxidative Folding and Reductive Unfolding Experiments

Purified native TRIAP1 (1mg/ml = 101 μM) was reduced and unfolded by incubation in 0.1 M Tris.HCl (pH 8.4) containing 200 mM DTT and 6.0 M Gdm.HCl for 3 h at RT. To initiate refolding, the fully-reduced/unfolded protein samples were loaded onto a HiTrap desalting column (GE Healthcare) connected to an AKTA purifier system, previously equilibrated in 0.1 M Tris.HCl (pH 8.4). The protein was eluted in 1.2 ml of equilibration buffer (protein concentration ~ 0.5 mg/ml) and incubated at RT in the absence (Control -) and in the presence of redox agents: 0.25 mM 2-mercaptoethanol (Control +), 0.5 mM GSSG, or 0.5 mM/1.0 mM GSSG/GSH. Refolding experiments were also performed in 6.0 M urea in the absence or presence of redox reagents. The refolding reaction was

monitored by removing aliquots of the sample at selected time-points and quenching them with 2% aqueous trifluoroacetic acid (TFA). Acid-trapped intermediates were subsequently analyzed by RP-HPLC using a linear 25-40% gradient of acetonitrile with 0.1% TFA over 50 min in a 4.6-mm C4 column (Phenomenex) at a flow rate of 0.75 ml/min. Reductive unfolding experiments were performed by incubating native TRIAP1 (15 μM) at RT upon the addition of: 1 mM DTT in 0.1 M Tris.HCl pH 8.4; 5 mM DTT in 0.1 M sodium phosphate pH 7.4 and 32 mM TCEP in 0.1 M sodium acetate pH 5.0. To monitor the progression of disulfide reduction, time course aliquots of the samples were trapped with 2% TFA acid and analyzed by RP-HPLC in a C4 column in the same conditions described above.

Far-UV Circular Dichroism (CD) and Fluorescence spectroscopy

The conformational changes along oxidative folding or TRIAP1 disulfide reduction were followed by far-UV CD and tryptophan fluorescence spectroscopy. To monitor the oxidative folding reaction, the sample was prepared by dissolving the fully-reduced/unfolded protein to a final concentration of 15 μM in 20 mM Tris-HCl (pH 8.4). CD spectra were collected at 20 min intervals in a spectropolarimeter (Jasco J-710) at 25 °C using an optical path length of 1 mm. Samples for fluorescence spectroscopy were measured in the same buffer on a fluorescence spectrophotometer (Jasco FP-8500) at 25 °C. The spectra were measured every 20 min in the 310-450 nm interval using a 294-nm excitation wavelength (5 nm excitation and emission slits, 0.1 s averaging time). For disulfide reduction, native TRIAP1 samples were prepared at 15 μM in: 50 mM Tris.Cl pH 8.4; 0.1 M sodium phosphate pH 7.4 or 0.1 M sodium acetate pH 5.0. After measuring the native protein spectrum, the reducing agent was added, without diluting the samples, to a final concentration of 5.0 mM DTT (pH 8.4); 1.0 mM DTT (pH 7.4) and 32 mM TCEP (pH 5.0). Tryptophan emission spectra were measured at various time points in the same conditions described for oxidative folding. The secondary structure changes were followed at 222 nm along time, upon the addition of the reducing agent.

Disulfide scrambling

The native TRIAP1 protein (15 μM) was dissolved in 0.1 M Tris.HCl pH 8.4 containing 0.25 mM 2-mercaptoethanol and increasing concentrations of denaturants: urea (0-8 M), Gdm.HCl (0-7 M) or Gdm.SCN (0-5 M). The samples were incubated for 20 h at RT to

reach equilibrium and quenched with 2% TFA acid for further RP-HPLC analysis, as detailed in oxidative folding. The disulfide scrambling experiments were performed in duplicates at RT.

Disulfide-bond content and disulfide pairing analysis of the major intermediates of oxidative folding and reductive unfolding of NvCI.

The acid-trapped intermediates were purified by RP-HPLC and freeze-dried. The free cysteines of the purified intermediates (20-40 µg) were blocked by derivatization with 0.1 M iodoacetamide (IAA) in Tris.HCl (pH 8.4) at RT for 45 min in the dark. An aliquot of the samples was analyzed by MALDI-TOF-MS, to characterize the number of disulfide bonds of the folding intermediates. The derivatized samples were then freed form reagents using Zip-Tip C4 pipette tip (Millipore), evaporated, and the disulfide bonds were reduced by incubation for 1 h at RT of the dried samples resuspended in 10 µl of 0.1 M Tris.HCl (pH 8.4), 10 mM DTT. The free cysteines, corresponding to those initially involved in disulfide bridges, were then subjected to a second alkylation reaction by adding 10 µl of 0.1 M 4-Vinylpyridine (VP)in 0.1 M Tris.pH 8.4. Then, the double-derivatized samples were purified by Zip-Tip C4 and air-dried. The dried samples were resuspended in 10 µl of ammonium bicarbonate pH 8.8 containing 0.15 µg of MS grade trypsin (Trypsin Gold, Promega) and were incubated for 16 h at 37 °C. The resulting derivatized digestion peptides were analyzed by MALDI-TOF-MS.

Stop/Go folding and Mia40 catalyzed reaction of Cys18-Cys37 TRIAP1 intermediate.

Intermediate i2 (Cys18-Cys37 TRIAP1) was isolated as follows: TRIAP1 (1 mg/ml) was reduced by adding to the reaction 16 mM TCEP in 0.1 M sodium acetate pH 5.0. After 30 min incubation at RT, the sample was acid quenched and subjected to RP-HPLC. The intermediate was isolated, lyophilized and quantified by RP-HPLC using a standard calibration curve of TRIAP1. Lyophilized intermediate was allowed to reinitiate folding at 15 µM in 0.1M Tris.Cl (pH 8.4), 0.1 M NaCl at RT (Control -). The Mia40 catalyzed reaction was performed by adding 30 µM of Mia40 to 15 µM of purified intermediate i2 in 0.1M Tris.HCl (pH 8.4), 0.1 M NaCl at RT.

Time-course aliquots of the folding reactions were trapped with 2% TFA acid and analyzed by RP-HPLC in a C4 column using the same gradient described for oxidative folding.

MALDI-TOF analyses

MALDI-TOF mass spectra were recorded on an UltrafleXtreme mass spectrometer (Bruker Daltonics) and the samples were spotted onto a MALDI-TOF-MS ground steel plate using the dried-droplet method. For peptide analyses the samples were mixed with α-cyano-4-hydroxycinnamic acid (hcca) and acquired in the positive ion reflectron mode, with ion acceleration set to 25 kV. All mass spectra were externally calibrated using a standard peptide mixture (Bruker Daltonics). For protein analyses the samples were mixed with sinapic acid and acquired in the positive ion lineal mode, with ion acceleration set to 25 kV. All mass spectra were externally calibrated using a standard protein mixture ranging from 4 kDa to 20 kDa (Bruker Daltonics).

References

1.   Anfinsen, C. B., Haber, E., Sela, M. & White, F. H. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc Natl Acad Sci U S A* **47**, 1309–1314 (1961).

2.   Creighton, T. E. Experimental studies of protein folding and unfolding. *Prog Biophys Mol Biol* **33**, 231–297 (1978).

3.   Arolas, J. L., Aviles, F. X., Chang, J.-Y. & Ventura, S. Folding of small disulfide-rich proteins: clarifying the puzzle. *Trends Biochem Sci* **31**, 292–301 (2006).

4.   Arolas, J. L. & Ventura, S. Protease Inhibitors as Models for the Study of Oxidative Folding. *Antioxid Redox Signal* **14**, 97–112 (2011).

5.   Tu, B. P., Ho-Schleyer, S. C., Travers, K. J. & Weissman, J. S. Biochemical basis of oxidative protein folding in the endoplasmic reticulum. *Science* **290**, 1571–1574 (2000).

6.   Ellgaard, L. & Ruddock, L. W. The human protein disulphide isomerase family: substrate interactions and functional properties. *EMBO Rep* **6**, 28–32 (2005).

7.   Soares Moretti, A. I. & Martins Laurindo, F. R. Protein disulfide isomerases: Redox connections in and out of the endoplasmic reticulum. *Arch Biochem Biophys* **617**, 106–119 (2017).

8.   Mordas, A. & Tokatlidis, K. The MIA pathway: a key regulator of mitochondrial oxidative protein folding and biogenesis. *Acc Chem Res* **48**, 2191–2199 (2015).

9.   Miliara, X. *et al.* Structural determinants of lipid specificity within Ups/PRELI lipid transfer proteins. *Nat Commun* **10**, 1130 (2019).

10.   Miliara, X. *et al.* Structural insight into the TRIAP/PRELI-like domain family of mitochondrial phospholipid transfer complexes. *EMBO Rep* **16**, 824–835 (2015).

11.   Tatsuta, T. & Langer, T. Intramitochondrial phospholipid trafficking. *Biochim Biophys Acta Mol Cell Biol Lipids* **1862**, 81–89 (2017).

12.   Aaltonen, M. J. *et al.* MICOS and phospholipid transfer by Ups2-Mdm35 organize membrane lipid synthesis in mitochondria. *J Cell Biol* **213**, 525–534 (2016).

13.   Claypool, S. M. & Koehler, C. M. The complexity of cardiolipin in health and disease. *Trends Biochem Sci* **37**, 32–41 (2012).

14.   Dudek, J. Role of Cardiolipin in Mitochondrial Signaling Pathways. *Front Cell Dev Biol* **5**, 90 (2017).

15.   Park, W.-R. & Nakamura, Y. p53CSV, a novel p53-inducible gene involved in the p53-dependent cell-survival pathway. *Cancer Res* **65**, 1197–1206 (2005).

16.   Adams, C. *et al.* Apoptosis inhibitor TRIAP1 is a novel effector of drug resistance. *Oncol Rep* **34**, 415–422 (2015).

17.   Zhang, T.-M. TRIAP1 Inhibition Activates the Cytochrome c/Apaf-1/Caspase-9 Signaling Pathway to Enhance Human Ovarian Cancer Sensitivity to Cisplatin. *Chemotherapy* **64**, 119–128 (2019).

18.   Fischer, M. & Riemer, J. The Mitochondrial Disulfide Relay System: Roles in Oxidative Protein Folding and Beyond. *Int J Cell Biol* **2013**, 12 (2013).

19.   Araiso, Y. *et al.* Structure of the mitochondrial import gate reveals distinct preprotein paths. *Nature* **575**, 395–401 (2019).

20.   Backes, S. & Herrmann, J. M. Protein Translocation into the Intermembrane Space and Matrix of Mitochondria: Mechanisms and Driving Forces. *Front Mol Biosci* **4**, 83 (2017).

21.   Banci, L., Barbieri, L., Luchinat, E. & Secci, E. Visualization of Redox-Controlled Protein Fold in Living Cells. *Chemistry & Biology* **20**, 747–752 (2013).

22.   Morgan, B. & Lu, H. Oxidative folding competes with mitochondrial import of the small Tim proteins. *Biochem J* **411**, 115–122 (2008).

23.   Fraga, H. *et al.* Disulfide driven folding for a conditionally disordered protein. *Sci Rep* **7**, 16994 (2017).

24.   Fraga, H. *et al.* The Mitochondrial Intermembrane Space Oxireductase Mia40 Funnels the Oxidative Folding Pathway of the Cytochrome Oxidase Assembly Protein Cox19. *J Biol Chem* **289**, 9852–9864 (2014).

25.   Fraga, H., Graña-Montes, R., Illa, R., Covaleda, G. & Ventura, S. Association Between Foldability and Aggregation Propensity in Small Disulfide-Rich Proteins. *Antioxid Redox Signal* **21**, 368–383 (2014).

26.   Sideris, D. P. *et al.* A novel intermembrane space–targeting signal docks cysteines onto Mia40 during mitochondrial oxidative folding. *J Cell Biol* **187**, 1007–1022 (2009).

27.   Milenkovic, D. *et al.* Identification of the Signal Directing Tim9 and Tim10 into the Intermembrane Space of Mitochondria. *Mol Biol Cell* **20**, 2530–2339 (2009).

28.   Chang, J. Y. A two-stage mechanism for the reductive unfolding of disulfide-containing proteins. *J Biol Chem* **272**, 69–75 (1997).

29.   Mamathambika, B. S. & Bardwell, J. C. Disulfide-Linked Protein Folding Pathways. *Annu Rev Cell Dev Biol* **24**, 211–235 (2008).

30.   Habchi, J., Tompa, P., Longhi, S. & Uversky, V. N. Introducing protein intrinsic disorder. *Chem Rev* **114**, 6561–6588 (2014).

31.   Jakob, U., Kriwacki, R. & Uversky, V. N. Conditionally and transiently disordered proteins: awakening cryptic disorder to regulate protein function. *Chem Rev* **114**, 6779–6805 (2014).

32.   Groitl, B. *et al.* Protein unfolding as a switch from self-recognition to high-affinity client binding. *Nat Commun* **7**, 10357 (2016).

33.   Reichmann, D. *et al.* Order out of disorder: working cycle of an intrinsically unfolded chaperone. *Cell* **148**, 947–957 (2012).

34.   Wong, L. H., Gatta, A. T. & Levine, T. P. Lipid transfer proteins: the lipid commute via shuttles, bridges and tubes. *Nat Rev Mol Cell Biol* **20**, 85–101 (2019).

35.   Finger, Y. & Riemer, J. Protein import by the mitochondrial disulfide relay in higher eukaryotes. *Biol Chem* **401**, 749–763 (2020).

Supplementary Information for:

MIA40 funnels the oxidative folding of a kinetically trapped molten globule

Jordi Pujols[1,2], Marcos Gil-Garcia[1,2], Sebastián Esperante[3*] and Salvador Ventura [1,2*]

[1] Institut de Biotecnologia i Biomedicina. Universitat Autònoma de Barcelona, 08193-Bellaterra, Spain.

[2] Departament de Bioquímica i Biologia Molecular. Universitat Autònoma de Barcelona, 08193-Bellaterra, Spain.

[3] Fundacion INFANT. Gavilan 94, C1406. CABA, Buenos Aires. Argentina

[*] Correspondence:

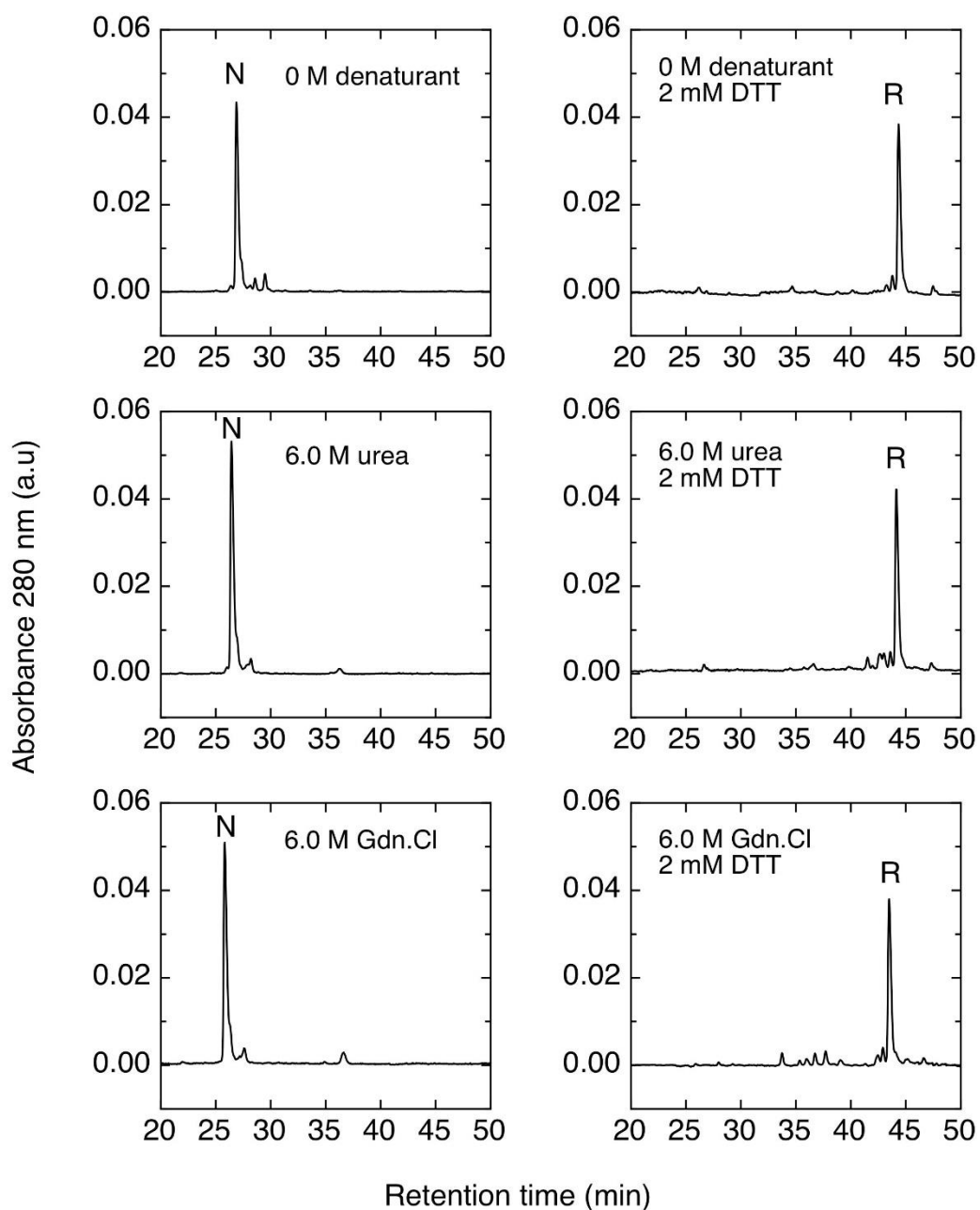sebandre@gmail.com;

salvador.ventura@uab.es; Tel.: (+34) 93 586 8956

Figure S1. RP-HPLC chromatographic profiles of *native, fully-reduced and fully-reduced/unfolded TRIAP1.* 15 μM of TRIAP1 samples were incubated at RT for 20 h in: 50 mM sodium phosphate pH 7.4 with: 0 M denaturant; 6.0 M urea or 6.0 M Gdn.Cl in the absence or presence of 2 mM DTT. The samples were subjected to far-UV CD and tryptophan fluorescence emission measurements followed by acid-quenching and RP-HPLC analysis. The incubation condition of the sample subjected to RP-HPLC is shown in each panel. The native (N) and the fully reduced (R) peaks are indicated.
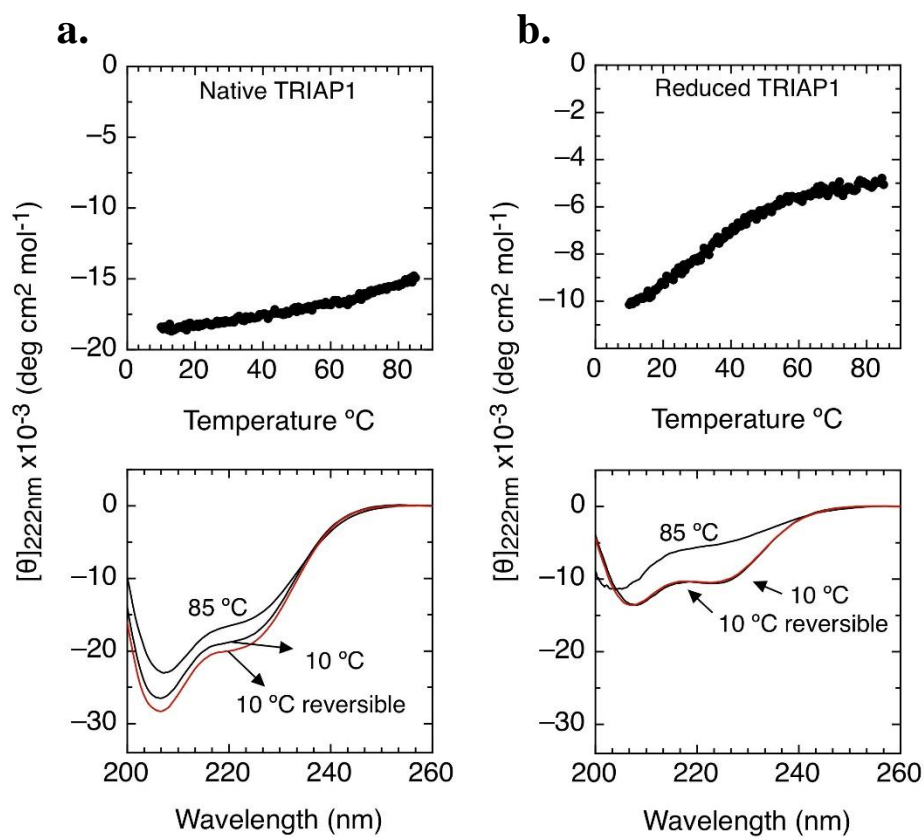
Figure S2. Thermal denaturation of native and fully reduced TRIAP1. The native (a) and fully reduced (b) TRIAP1 samples at 15 μM were dissolved in 50 mM sodium phosphate (pH 7.4) and the temperature wa gradually increased from 10 °C to 85 °C, with a heating rate of 1.0 °C min^{-1}. The changes in the ellipticity at 222 nm were monitored every 0.5 °C. Spectra of the samples were taken at the initial and final temperatures (10 and 85 °C, respectively) and after the thermal denaturation, the samples were equilibrated at 10 °C to confirm the reversibility of the reaction.
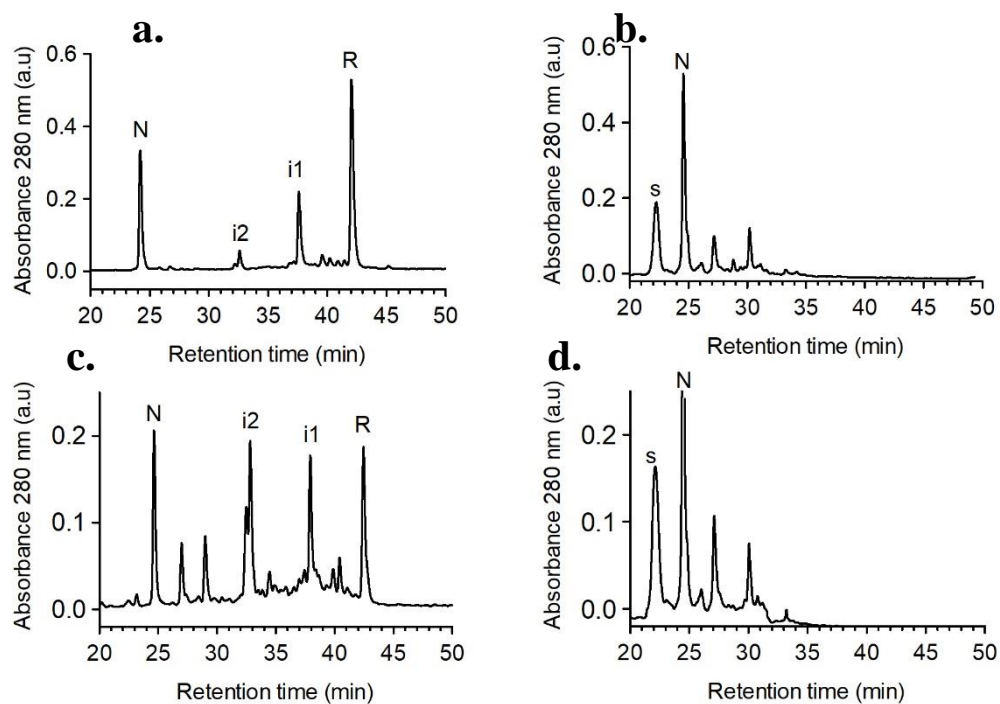
Figure S3. Comparison of RP-HPLC chromatographic profiles of TRIAP1 oxidative folding intermediates. RP-HPLC chromatographic profiles of selected acid-quenched TRIAP1 refolding reactions at different time points and incubating conditions: 8 h, control – (a); 4 min, 1 mM GSSG (b); 1 h (c) and 2 h (d), in 0.5 mM GSH, 1 mM GSSG. The native (N), reduced (R), intermediates (i1 and i2) and scramble (S) peaks are indicated. The heterogeneity of peak i2 is clearly observed in (b). The accumulation of scrambles species upon incubation is clearly represented comparing panel (c) with panel (d).
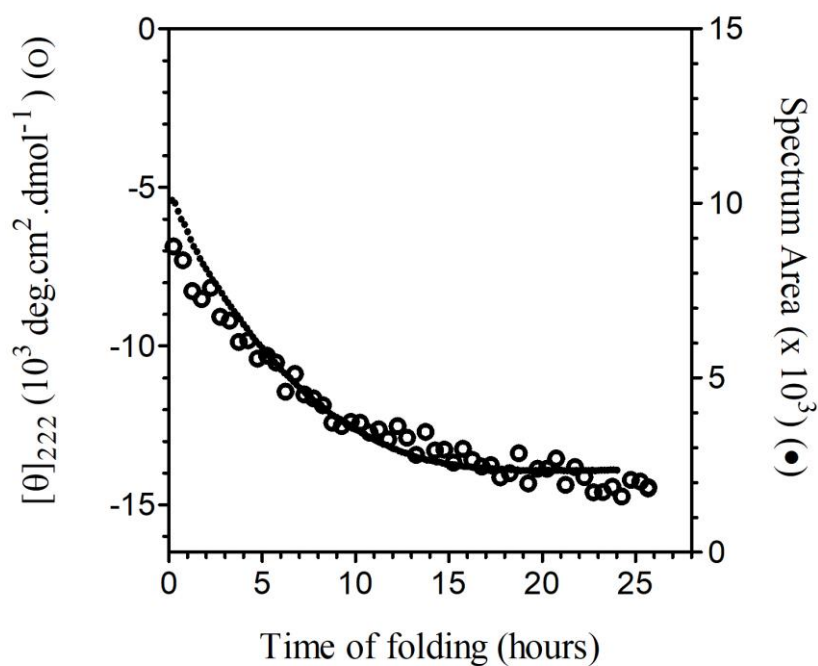
Figure S4. *Secondary and tertiary structural changes upon oxidative folding of TRIAP1.* The secondary and tertiary structure changes along folding time were monitored by far-UV CD and tryptophan fluorescence spectra measurements, respectively. The fully reduced/unfolded protein was allowed to refold in 0.1 M Tris-HCl pH 8.4 at 20 °C. Left Y axis represents the molar ellipticity changes at 222 nm over time while right Y axis shows the changes in the tryptophan spectrum area as a function of folding time. Both signals were coupled along refolding time.
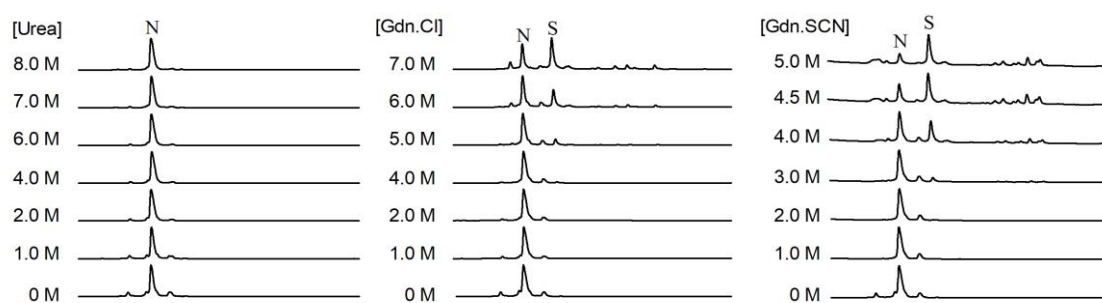
Figure S5. *Disulfide scrambling of TRIAP1 under different concentrations of denaturants.* (a) The native form of TRIAP1 was denatured in Tris-HCl buffer (pH 8.4) containing 0.25 mM 2-mercaptoethanol as thiol initiator and the indicated concentration of denaturants at 20 °C for 20 h. The samples were incubated in increasing concentrations of urea (left panel), Gdn.Cl (middle panel) and GdnSCN (right panel). The denatured samples were quenched with 2% TFA and analyzed by RP-HPLC. The native (N) and the major scrambled species (S) peaks are indicated.

CHAPTER 4

Disorder and amyloidogenicity

Work 3

The disordered C-terminus of yeast Hsf1 contains a cryptic low-complexity amyloidogenic region

The disordered C-terminus of yeast Hsf1 contains a cryptic low-complexity amyloidogenic region

Jordi Pujols[1,2], Jaime Santos[1,2], Irantzu Pallarès[1,2,*] and Salvador Ventura[1,2,*]

[1] Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, E-08193 Bellaterra (Barcelona), Spain.

[2] Departament de Bioquímica i Biologia Molecular, Universitat Autònoma de Barcelona, E-08193 Bellaterra (Barcelona), Spain.

[*] Correspondence:

irantzu.pallares@uab.cat; Tel.: +34-93-581-2154;

salvador.ventura@uab.es; Tel.: +34-93-586-8956

Abstract

Response mechanisms to external stress rely on networks of proteins able to activate specific signaling pathways to ensure the maintenance of cell proteostasis. Many of the proteins mediating this kind of response contain intrinsically disordered regions, which lack a defined structure, but still are able to interact with a wide range of clients that modulate the protein function. Some of these interactions are mediated by specific short sequences embedded in the longer disordered regions. Because the physicochemical properties that promote functional and abnormal interactions are similar, it has been shown that, in globular proteins, aggregation-prone and binding regions tend to overlap. It could be that the same principle applies for disordered protein regions. In this context, we show here that a predicted low-complexity interacting region in the disordered C-terminus of the stress response master regulator Hsf1 protein corresponds to a cryptic amyloid region able to self-assemble into fibrillary structures resembling those found in neurodegenerative disorders.

Keywords: amyloid; Hsf1; protein aggregation; intrinsically disordered regions; Q/N-rich regions; low complexity; molecular recognition features

Introduction

Eukaryotic cells have the capacity to overcome detrimental environmental conditions, such as hyperthermia or oxidative stress. This ability to maintain a balanced proteostasis depends on a restricted group of polypeptides, named Heat Shock Proteins (HSPs) [1, 2]. This set of protective proteins are designed to guarantee protein solubility and impede a massive loss of protein function due to degradation, post-transcriptional modifications or aggregation. In that frame, Heat Shock Factor 1 (Hsf1) has been classically considered as the master orchestrator of this finely tuned response [3]. Hsf1 is an inducible transcription factor that, after activation, leads to the transcription of several HSPs and the subsequent metabolic readjustment [4]. Generically, in order to execute its function, Hsf1 undergoes trimerization, nuclear specific localization and an extensive diversity of post-transcriptional modifications [2]. Indeed, under basal conditions, the protein is subjected to a negative regulation loop in which, intramolecular contacts and specific interactions with other HSPs such as HSP90 preserve the inactive monomeric form of the protein [2, 5, 6].

Hsf1 function is conserved among different species, the DNA Binding Domain (DBD) being the most conserved region in the protein sequence. Despite the DBD corresponds to a folded domain, Hsf1 is thought to be mostly unstructured [7]. In fact, it is well-established that numerous proteins related with DNA-binding and transcriptional regulation lack ordered tertiary structure or contain long disordered segments under physiological conditions [8, 9]. In these proteins, disordered regions contribute to their conformational plasticity and client promiscuity and this is crucial to establish a wide range of specific interactions [10]. In many cases, these interactions seem to depend on the existence of short sequence motifs which may or may/not undergo a disorder to order transition upon binding [11-14]. In addition, it is becoming increasingly evident that the compositional bias of certain disordered low complexity regions is an important determinant of protein-protein interactions in transcription factors [15].

It has been shown that the physicochemical properties behind functional protein-protein interactions overlap significantly with those accounting for the anomalous interactions leading to protein aggregation [16]. As a result, interactions sites in protein complexes [17, 18], as well as the interfaces of homomeric and heteromeric globular proteins display significant amyloidogenic propensity [19, 20]. This implies that exposed aggregation-

prone regions cannot be totally purged out by evolution from the sequences of globular proteins, because they are needed for binding and/or functional assembly [18, 19]. We wondered whether in disordered proteins, where the bulk of the sequence is exposed to solvent, this duality between the establishment of functional and deleterious interactions also occurs. We show here that, in yeast Hsf1, a disordered and low complexity region at the C-terminus of this transcription factor is predicted to act in molecular recognition, at the cost of aggregating spontaneously into ordered amyloid assemblies.

Results

Identification of a cryptic amyloid sequence inside a Hsf1 disordered and low complexity region

We addressed the order/disorder content of yeast Hsf1. We first examined the location of constituent functional domains and unstructured regions in the full-length protein (Fig. 1). PFAM indicates that yeast Hsf1 contains a single annotated functional domain comprising residues 175-275, which corresponds to the HSF_DNA-binding domain [21]. This domain is highly-conserved both structurally and sequentially from yeast to human and binds to specific DNA regulatory regions involved in the modulation of the complex HSF's transcriptional activity [2]. Thus, according to PFAM, less than 15% of Hsf1 residues are implicated in the formation of functional folded moieties in the monomeric state. Indeed, the N-terminal Hsf1 1-167 region has been experimentally shown to lack any regular structure [22]. This is consistent with the predictions of protein disorder provided by two different algorithms, ESpritz-NMR [23] and IUPred [24], which point out that the protein regions at the two sides of the HSF_DNA-binding domain are essentially disordered (Figure 1). The SEG algorithm [25] indicates that Hsf1 disordered regions contain several short low complexity segments (Figure 1). We used FELLS [26] to address if any of these low complexity regions displays compositional bias. Interestingly enough, the region comprising residues 493-527 exhibits a striking compositional bias, with an overall asparagine content of 40% (Figure 1). Notably, this is an N enrichment comparable to that observed in the prion domains of several yeast prions [27, 28]. In yeast, these Q/N-rich domains promote self-assembling interactions, which result in conformational conversion [29] and phase separation [30].
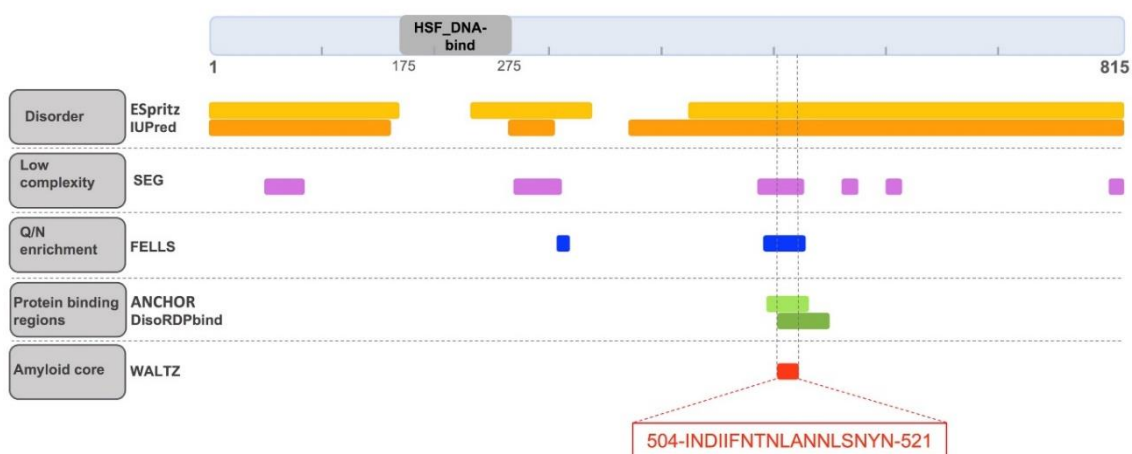
Heat shock factor protein – Hsf1

Figure 1. S. cerevisiae Hsf1 cryptic amyloid core. Hsf1 diagram showing the location of the identified Pfam domain (grey), predicted disordered regions (orange), low complexity segments (violet), Q/N rich regions (blue), protein binding sites (green) and the cryptic amyloid core (red). The sequence of the amyloid core is shown in the red box.

We have recently shown that the disordered, low complexity, Q/N-rich regions of yeast prions and other prion-like proteins might contain cryptic amyloidogenic regions that can contribute significantly to trigger protein assembly [31-33]. These regions differ from the classical amyloid cores found in pathogenic proteins in that they are longer and more polar, in such a way that the amyloid potential is less concentrated, allowing the protein to remain in a soluble state under most physiological conditions, while still keeping a certain amyloid propensity that might facilitate its assembly in certain circumstances [34]. We analyzed if the disordered and N-rich 493-527 Hsf1 region hides an amylogenic stretch using WALTZ, an algorithm that exploits a position-specific scoring matrix to identify potential amyloid sequences [35]. WALTZ, predicted the 504-INDIIFNTNLANNLSNYN-521 sequence as the longest region displaying a significant amyloid potential. It is worth mentioning that, in this segment, 55% of the residues have a polar nature and accordingly, its hydropathicity is much lower than those of the typical amyloid cores found in disease-linked proteins (Table 1). This explains why well-validated aggregation predictors like Aggrescan [36] and Zyggregator [37], which consider sequence hydrophobicity as a major contributor to protein aggregation, did not detected any aggregation-prone region in the 493-527 Hsf1 sequence.

Table 1. *S. cerevisiae* **Hsf1 amyloid core hydropathicity.**

Protein	Amyloid Core	Gravy Score
Hsf1	INDIIFNTNLANNLSNYN	-0.283
α-synuclein	GVLYVG	1.683
	GGAVVTGVTAVAQ	1.238
Aβ42	GAIIGLMVGGVVI	2.462
	QKLVFFAE	0.562

* The GRAVY score (average hydrophobicity and hydrophilicity) was evaluated using the EXPASY ProtParam tool [40]. The GRAVY scores for the amyloid cores of α-synuclein and Aβ42, involved in Parkinson's and Alzheimer's diseases, respectively, are shown for comparison.

We investigated whether our candidate amyloid core could potentially be a region involved in the establishment of intermolecular contacts. To this end, we used ANCHOR [38] and DisoRDPbind [39], two conceptually different algorithms aimed to identify protein binding sequences in disorder proteins or regions. Interestingly enough, both programs successfully identified the sequence stretch around Hfs1 residues 504–521 as a potential protein binding region (Fig. 1).

Overall, the bioinformatics analysis suggest that Hsf1 might contain a functional amyloid core embedded in a disordered and low complexity region. To confirm that this soluble transcription factor bears a sequence with the potential to form amyloid-like assemblies, the 18-residue-long peptide was synthesized and characterized experimentally.

Hsf1 amyloid core assembles into β-sheet enriched aggregates

As a first step, and in order to determine whether the candidate peptide has the ability to aggregate in vitro, Hsf1 amyloid core peptide samples were prepared at 25, 50 and 100 μM and incubated at 25 °C for 120 h. Aggregation was systematically monitored by measuring changes in synchronous light scattering [41]. As shown in Figure 2A, a concentration dependent scattering signal could be observed, reaching a maximum at 100 μM of peptide. These data indicate that despite the candidate amyloid core is clearly less

hydrophobic than those in classic amyloids, it can still establish the initial intermolecular interactions that lead to the buildup of high-order assemblies.
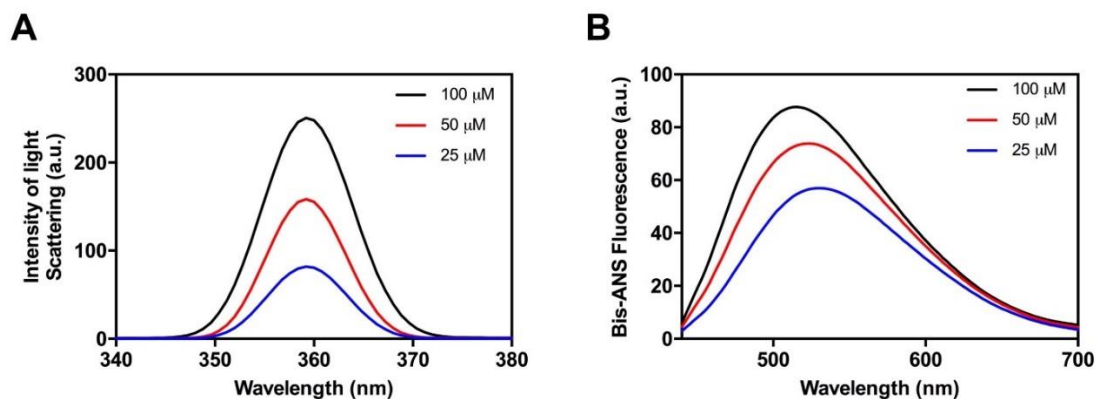

Figure 2. Aggregation of Hsf1 soft amyloid core as a function of the concentration. Aggregation (A) and the formation of hydrophobic surfaces (B) were monitored by light scattering and Bis-ANS fluorescence emission, respectively, at various peptide concentrations.

Next, we explored the presence of exposed hydrophobic clusters in the formed peptide aggregates by measuring their binding to Bis-ANS, a dye that exhibits fluorescence enhancement upon interaction with exposed hydrophobic surfaces as well as a blue shift of the peak maximum [42, 43] (Fig. 2B). We found complete agreement between the Bis-ANS and the light-scattering assays. The Bis-ANS fluorescence presents a concentration dependent intensity increase and the maximum emission shifts from 521 nm, in the presence of 25 μM of peptide, to 505 nm at 100 μM. These spectral changes indicate that the peptide aggregation results in the formation of novel/larger hydrophobic patches, thus suggesting that the interactions between the peptide hydrophobic residues lead to its assembly.

To analyze the secondary structure content of the Hsf1 peptide aggregates we recorded the amide I region of the FTIR spectrum (1700–1600 cm-1) upon its incubation at 100 μM for 120 h. Deconvolution of the obtained spectra allow us to identify a strong band at 1628 cm-1, which is indicative of a predominance of extended intermolecular β –sheets structures. This signal is the largest contributor to the absorbance spectrum accounting for 62% of the total area. The other detected peak at 1667 cm-1 is attributable to the presence of β –turns (Fig. 3). Because no anti-parallel β-sheet band was detected at ~1690 cm-1, it suggested a preferential parallel β-sheet organization for the self-assembled peptide.
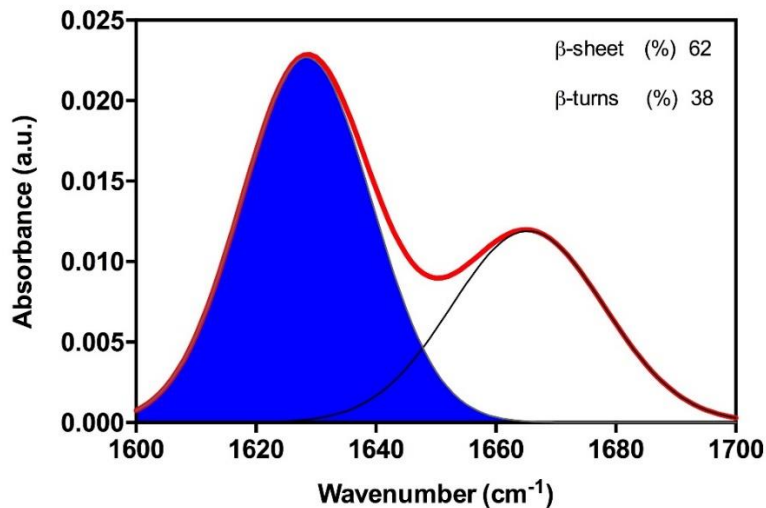
Figure 3. Hsf1 amyloid core secondary structure. Secondary structure was determined from the ATR FT-IR absorbance spectrum in the amide I region. The red line corresponds to the original spectrum; the blue area indicates the contribution of the inter-molecular β-sheet signal to the total area upon Gaussian deconvolution.

Collectively, the data indicate that the Hsf1 predicted amyloid core spontaneously self-assembles into highly ordered supramolecular structures highly enriched in β –sheet content.

Hsf1 amyloid core forms amyloid-like fibrillary structures

Next, to assess whether the detected β –sheets structures in Hsf1 peptide aggregates correspond to amyloid-like assemblies we used the amyloid-specific dyes Congo red (CR), Thioflavin-T (ThT) and Thioflavin-S (ThS) [44]. CR is widely used for the detection of amyloid material [45], when the dye is bound to cross β –sheet structures there is hypercromicity and a red shift of the absorbance maximum. We analyzed CR binding to the aggregated peptide at 25, 50 and 100 μM. As expected, the peptide at 100 μM promoted the maximum shift in the CR spectra from 497 nm, in the absence of peptide, to 515 nm (Fig. 4A).

Similarly, binding of ThT to the Hsf1 amyloid core peptide induced a large enhancement in the intensity of ThT fluorescence emission relative to the free dye, increasing as a function of the peptide concentration (Fig. 4B). Additionally, binding of ThS to 100 μM peptide aggregate was tested by fluorescence microscopy. Areas rich in fibrous material appear stained with ThS, giving a bright green-yellow fluorescence against a dark

background (Fig. 4C). Overall, the results reported from the dye staining experiments strongly suggest an amyloid-like organization of the peptide in the aggregates.
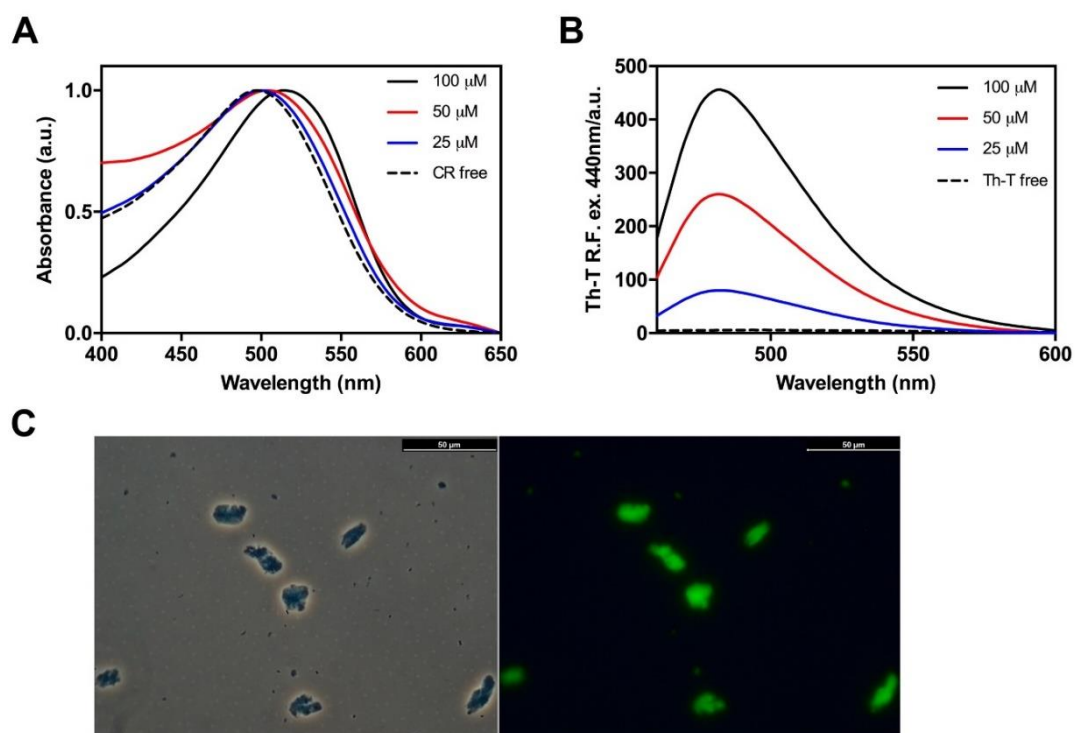

Figure 4. Binding of amyloid dyes to Hsf1 cryptic amyloid core. (A) CR spectral changes in the presence of Hsf1 peptide incubated at various concentrations during 120 h at 25 °C; note the characteristic shift from 497 nm to 515 nm when the dye is bound to amyloid-like aggregates. (B) Changes in the fluorescence emission spectrum of Th-T when excited at 440 nm upon binding to the aggregated peptide at various concentrations after 120 h incubation at 25 °C (C) Thioflavin-S fluorescence of stained Hsf1 aggregated amyloid material at 100 μM in PBS after 120 h incubation at 25 °C. Images were obtained at 40X magnification by phase contrast and fluorescence microscopy displaying the green fluorescence characteristic of amyloid structures.

To check whether the peptide assemblies could have any internal fibrillar structure, we performed a morphological analysis using transmission electron microscopy (TEM) (Fig. 5). In agreement with the above reported results, the 100 μM Hsf1 amyloid core peptide solution self-assembles forming fibrillar arrangements after 120 h incubation at 25 °C. The resultant fibrils are straight, unbranched and exhibit a diameter that varies from 6 to 12 nm, being thus similar to the fibrillar structures formed by amyloidogenic proteins involved in neurodegenerative disorders [46].
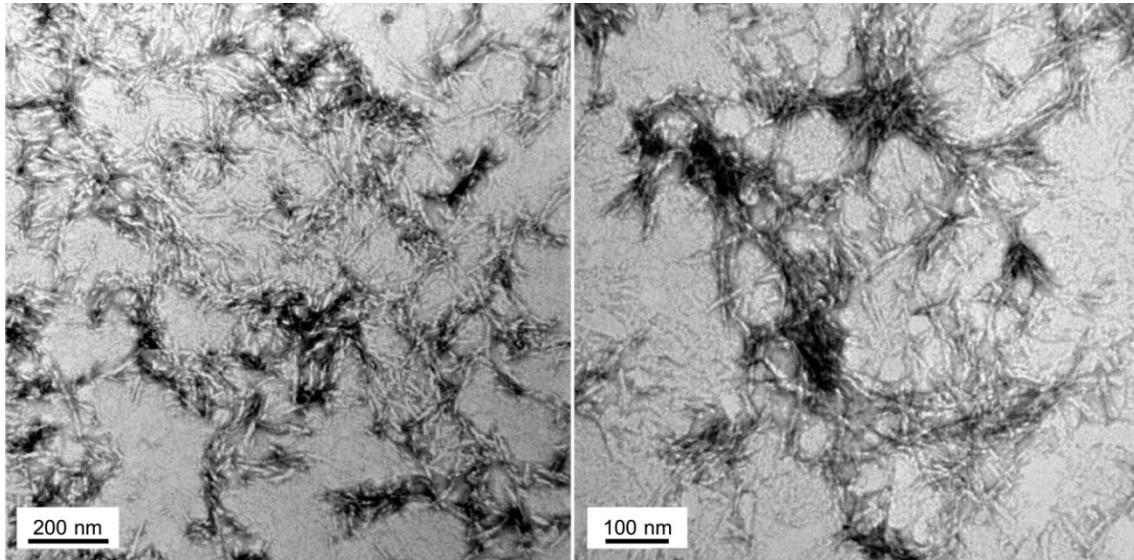
Figure 5. Fibrillar structure of Hsf1 cryptic amyloid core. Representative electron micrographs of Hsf1 aggregated peptide at 100 µM in PBS after 120 h of incubation at 25 °C.

Hsf1 amyloid core self-assembles very fast

Despite classical predictors suggested a low amyloid propensity for the Hsf1 peptide, the above data indicate the contrary. To asses if the aggregation tendency of the peptide is high enough to promote its immediate assembly, we monitored the properties of the solution upon peptide dilution at 100 µM in PBS (time 0 h) and compared them with those of the final aggregates (120 h). Synchronous light scattering indicated an initial level of aggregated material similar to the one observed at the end point of the reaction (Fig. 6A). However, the conformational properties of early and late aggregates are significantly different. Early aggregates display a higher Bis-ANS signal (Fig. 6B); in contrast, they exhibit a much lower ThT binding than late aggregates (Fig. 6C). This observation is consistent with their less ordered architecture as assessed by TEM images, which show the presence of proto-fibrillar assemblies that tend to coalescence in larger aggregates (Fig. 6D). Altogether the data indicate that the Hsf1 amyloid core undergoes a very fast aggregation reaction into proto-fibrils with exposed hydrophobic patches and a low average content of amyloid-like regions. The evolution of these aggregates into the typical amyloid assemblies observed at the end of the reaction necessarily involves the burial of a fraction of previously exposed hydrophobic residues inside the ordered fibrils. These data corroborate the strong aggregation propensity of the identified Hsf1 disordered region.
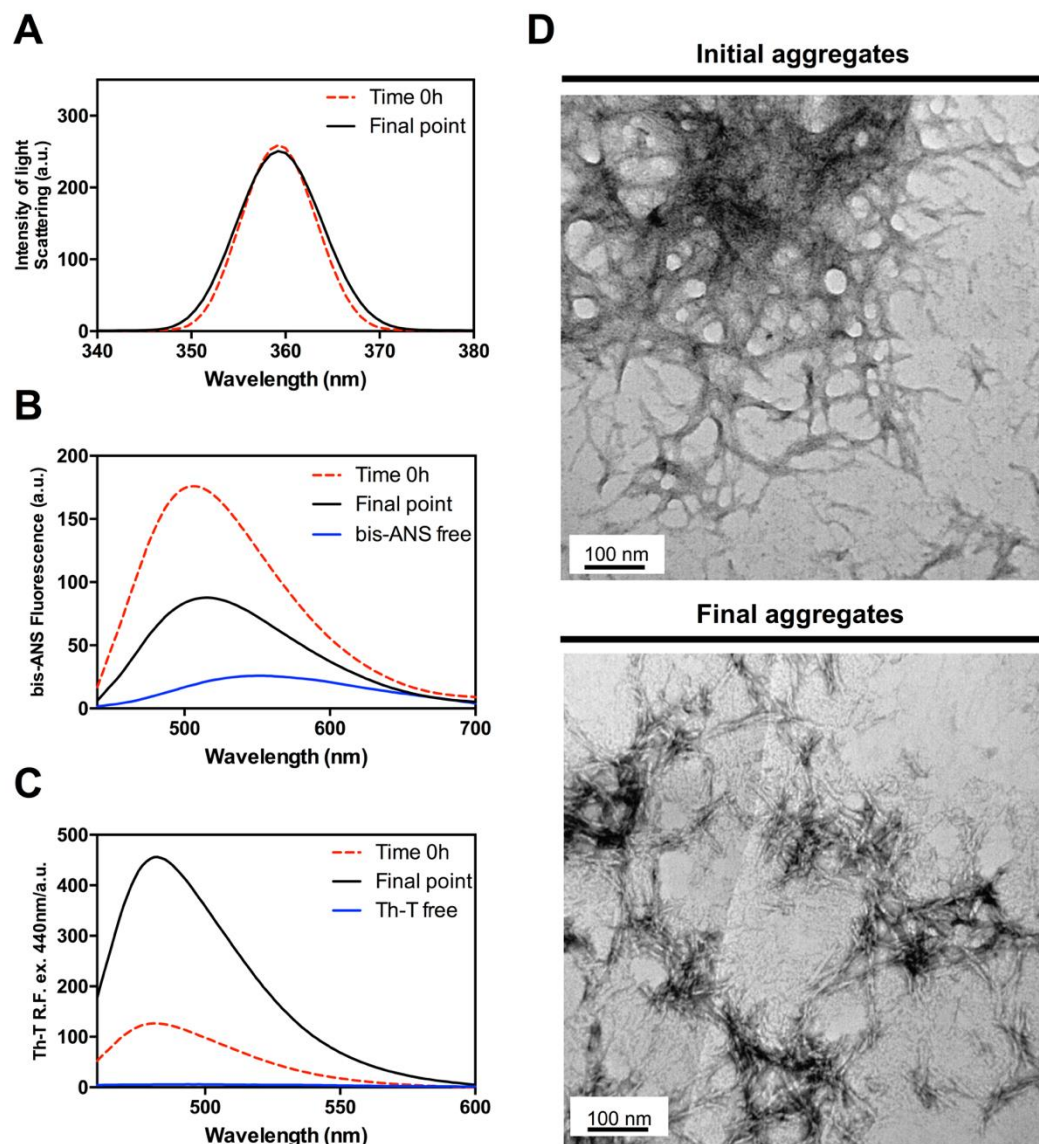
Figure 6. Evolution of Hsf1 amyloid core from initial non-ordered aggregates into mature amyloid fibrils. (A) Light scattering. (B) Bis-ANS fluorescence emission. (C) ThT fluorescence emission. (D) Representative TEM images of initial and final Hsf1 peptide aggregates. The final point correponds to 120 h.

Discussion

The present work exemplifies a strategy to uncover hidden amyloid segments in protein disordered regions. Together, the computational and biophysical analysis we have performed provide evidence for the existence of a cryptic amyloid core embedded in a disordered and low complexity region of yeast Hsf1. A peptide corresponding to this region spontaneously self-assembles into CR and ThT positive β-sheet enriched amyloid fibrils in vitro.

Despite hydrophobic interactions seem to be important for the assembly, the majority of the residues in the amyloid core have a polar nature, which endorses this segment with an overall lower hydrophobicity than the cores in pathogenic amyloids. Indeed, 70% of the polar residues in this region correspond to asparagine, a composition that is reminiscent of the one found in the amyloidogenic regions of prion-like domains, which contribute significantly to their conformational conversion [28, 31]. Even though, in the light of the extensive literature on Hsf1, it is unlikely that, in vivo and in the context of the full-length protein this region would drive Hsf1 amyloid formation; it is now evident that protein evolution tends to purge out aggregation-prone regions if they do not serve functional purposes, especially if they are constantly exposed to solvent, as it occurs in disordered protein regions [47]. In fact, in contrast with classical pathogenic amyloids, when amyloid sequences occur in disordered regions, their aggregation potential tends to be less concentrated, thus precluding uncontrolled aggregation and suggesting they might play instead a functional role [33]. Indeed, ANCHOR [38] and DisoRDPbind [39] predictions converge to suggest that the cryptic amyloid core we identified in Hsf1 might correspond to a molecular recognition element [48]. If this comes to be true, it would imply that both in the context of folded and disordered proteins the determinants for functional and deleterious interactions tend to overlap. Indeed, a number of diseases that have been associated with disorder-containing proteins seem to be caused by undesirable protein interactions [49].

This work contributes new data in favor of the hypothesis that protein aggregation-prone regions might play a role in protein-protein interactions [50]; thus supporting the vision that, despite the medical relevance of pathogenic amyloid formation, this deleterious reaction constitutes only one of the many different aggregation-driven phenomena that might occur in a cell [51].

It is evident that more work is needed to confirm the suggested relationship between binding and amyloid propensities in the context of Hsf1. However, the fact that the properties of the identified amylogenic segment coincide precisely with those of the amyloid core we found previously in the prion-like domain of the bacterial Transcription Terminator Rho [52, 53], which was afterwards confirmed to self-assemble in vivo through this specific region in order to regulate its activity [54], indicates that a deeper characterization of Hsf1 conformational and self-assembling properties will be worth the effort.

Materials and Methods

Computational identification of Hsf1 soft amyloid cores in low complexity regions

The sequence of S. cerevisiae Hsf1 (P10961) was downloaded from Uniprot [55]. It was first scanned for the presence of regions with a compositional bias with at least 30% of Q/Ns inside intrinsically disordered regions using FELLS [26] and IUPred algorithms [24, 35]. The identified region (493-527), was further evaluated in search for soft amyloid cores with WALTZ [35], which resulted in the identification of one positive candidate region. A final analysis in search of protein binding region within this soft amyloid core, was performed using ANCHOR [38] and DisoRDPbind [39], resulting in the identification of the Hsf1 soft amyloid core as a putative protein binding region.

Hsf1 soft amyloid core peptide preparation

An 18 residue peptide with the sequence INDIIFNTNLANNLSNYN, that corresponds to the predicted soft amyloid core of S. cerevisiae Hsf1, was purchased from CASLO ApS (Scion Denmark Technical University). The peptide was initially resuspended in 1,1,1,3,3,3-hexafluoropropanol to remove any residual aggregate and afterwards the solvent was evaporated under vacuum. The peptide stock solution was prepared solubilizing the resulting peptide film at a final concentration of 5 mM in 100% dimethyl sulfoxide and stored at -80 °C. Before each analysis, the sample was diluted to 150 μM in PBS pH 7.4. For aggregation assays the peptide was diluted to 25, 50 and 100 μM and incubated for 120 h at 25 °C.

Synchronous light scattering

Synchronous light scattering was monitored using a Cary Eclipse spectrofluorometer (Varian, Palo Alto, CA, USA). Spectra were recorded as the accumulation of three consecutive scans and the conditions of the acquisition were: excitation wavelength of 360 nm, emission range from 340 to 380 nm, slit widths of 5 nm at 25 °C in PBS pH 7.4 buffer. The samples were sonicated for 10 min in an ultrasonic bath (VWR ultrasonic cleaner) before measurement.

Bis-ANS binding

The relative fluorescence spectra of 4,4-bis-1-anilinonaphthalene-8-sufonate (Bis-ANS, Sigma Aldrich, Merck KGaA, Darmstadt, Germany) were obtained using a Cary Eclipse

spectrofluorometer (Varian, Palo Alto, CA, USA). Spectra were recorded as the accumulation of three consecutive scans and the conditions of the acquisition were: excitation wavelength of 370 nm, emission range from 400 to 700 nm and slit widths of 5 nm, the measurements were performed at 25 °C in PBS pH 7.4 buffer. The samples were sonicated for 10 min in an ultrasonic bath (VWR ultrasonic cleaner) before dye addition. Prior to fluorescence measurements samples were incubated at room temperature for 5 minutes in the presence of 10 µM final Bis-ANS concentration. A Bis-ANS solution without peptide was used as a control.

Attenuated total reflectance (ATR) fourier transform infrared (FTIR) spectroscopy

Attenuated total reflectance Fourier transform infrared (ATR FT-IR) spectroscopy analysis of peptide fibrils was performed using a Bruker Tensor FT-IR Spectrometer (Bruker Optics, Berlin, Germany) with a Golden Gate MKII ATR accessory. Aggregated peptide solution, previously sonicated for 10 min in an ultrasonic bath (VWR ultrasonic cleaner), was dried out under a $N_2(g)$ atmosphere and each spectrum consisted of 16 independent scans, measured at spectral resolution of 2 cm-1 within the 1800-1500 cm-1 range. Spectral data were acquired and normalized using the OPUS MIR Tensor 27 software. The individual components of the spectrum were determined through second derivative analysis of the spectra and deconvoluted afterwards into overlapping Gaussian curves, the amplitude, mass center, bandwidth at half of the maximum amplitude, and area for each Gaussian function were calculated employing the nonlinear peak-fitting program PeakFit v4.12 (Systat Software Inc., USA). The PBS pH 7.4 buffer without peptide was used as a control and subtracted from the absorbance signal before deconvolution.

Binding to amyloid dyes

Congo red binding (CR) to aggregated peptide solutions was determined using a Cary 100 UV/Vis spectrophotometer (Varian, Palo Alto, CA, USA) in the 400-700 nm range using a 1 cm optical length quartz cuvette placed in a thermostated cell holder at 25 °C.

The samples were sonicated for 10 min in an ultrasonic bath (VWR ultrasonic cleaner) before dye addition. Prior to absorbance measurements samples were incubated at room temperature for 10 minutes in the presence of 5 µM final CR concentration. PBS pH 7.4 buffer with 5 µM CR and without peptide was used as a control.

The relative fluorescence spectra of Thiofavin-T (ThT, Sigma Aldrich, Merck KGaA, Darmstadt, Germany) were monitored using a Cary Eclipse spectrofluorometer (Varian, Palo Alto, CA, USA). Spectra were recorded as the accumulation of three consecutive scans and the conditions of the acquisition were: excitation wavelength of 440 nm, emission range from 460 to 600 nm and slit widths of 5 nm, the measurements were performed at 25 °C in PBS pH 7.4 buffer. The samples were sonicated for 10 min in an ultrasonic bath (VWR ultrasonic cleaner) before dye addition. Prior to fluorescence measurements samples were incubated at room temperature for 2 minutes in the presence of 25 μM final ThT concentration to ensure that thermal equilibrium had been achieved. PBS pH 7.4 buffer with 25 μM ThT and without peptide was used as a control.

For the Thioflavin-S (ThS, Sigma Aldrich, Merck KGaA, Darmstadt, Germany) staining by fluorescence microscopy, Hsf1 aggregates were incubated for 1 h at room temperature with continuous agitation at 80 rpm in the presence of ThS 150 μM final concentration in PBS pH 7.4 buffer. Then, the excess of ThS was washed up by centrifugation and resuspension of the precipitated fraction in PBS pH 7.4 buffer for three times. Finally, the precipitated fraction was resuspended in a final volume of 10 μl, placed on a microscope slide and the coverslip sealed. Images of the peptide aggregates bound to ThS were obtained at 40-fold magnification under UV light or using phase contrast in a Leica fluorescence microscope (Leica DMRB, Heidelberg, Germany).

Transmission electron Microscopy (TEM)

An aliquot (5 μl) of 100 μM peptide incubated for 120 h at 25 °C and sonicated for 10 min in an ultrasonic bath (VWR ultrasonic cleaner) was placed on carbon-coated copper grids and allowed to absorb for 5 minutes. Then, the excess of solution was wicked away using small pieces of ashless filter paper, the samples were stained with 5 μl of uranyl acetate (2% w/v) for 2 min for the negative staining. Finally, the excess of uranyl acetate was removed as above, and the grids left to dry before the TEM observation. Amyloid fibrils images were obtained using a JEOL JEM-1400 electron microscope, (JEOL Peabody, MA, USA) operated at 80 kV accelerating voltage.

CHAPTER 5

Chemical chaperones as novel drugs

for Parkinson's Disease

Chemical chaperones as novel drugs for Parkinson's Disease

α-Synuclein: a therapeutic target in Parkinson's disease

Synucleinopathies are a series of heterogeneous neurodegenerative disorders affecting over 10 million people worldwide. Currently, no neuroprotective or neurorestorative therapies are available for synucleinopathies, which include Parkinson's disease (PD), multiple system atrophy (MSA), dementia with Lewy bodies (DLB) and neurodegeneration with brain iron accumulation [110–113]. These diseases all share common neuropathological features, characterized by the accumulation of deposits of the protein α-synuclein (α-syn) in the central and peripheral nervous system, although the cellular and anatomical localization of the aggregates differ among the diseases [114].

PD is the most common synucleinopathy and the second most prevalent neurodegenerative disorder after Alzheimer's disease (AD), affecting 1% of people over 60 years of age [115]. The motor symptoms of PD, including bradykinesia, rigidity, resting tremor, postural instability and gait disturbances, are mainly attributed to a loss of dopaminergic neurons in the substantia nigra pars compacta with a concomitant dopamine deficiency in the basal glia, a key brain region for motor control [116,117]. The affected neurons exhibit α-syn cytoplasmic inclusions in the neuronal body, Lewy bodies (LBs), and/or fibrils deposited in neuronal processes, Lewy neurites (LNs) [118]. Mutations in the gene that encodes for this protein (SNCA) are associated with familial cases of PD [119,120], and duplications and triplications of the SNCA gene lead to early onset of the disease [121,122]. Additionally, transgenic animals expressing *wild type* or mutated α-syn variants exhibit brain amyloid inclusions and develop synucleinopathy-like features [123]. These findings support a connection between α-syn and PD. Accordingly, α-syn-targeting agents are being developed in the search for a first disease-modifying therapy for PD.

α-Synuclein misfolding and aggregation

α-Syn is abundant in neuronal cells and is involved in synaptic transmission, contributing to the release of synaptic vesicles and synaptic membrane recycling [124]. In healthy neurons, α-syn is found either bound to membranes or in a monomeric, soluble and

disordered form in the cytosol, with both species interconverting in a rapid equilibrium. It has been suggested that α-syn can also form helical tetramers in the cytoplasm [125], but this possibility is still under debate [126]. During pathogenesis, α-syn misfolds and self-assembles into insoluble amyloid fibrils displaying a typical cross-β-sheet fold [127].

The conformational plasticity of α-syn is dictated by its unique sequence, which can be subdivided into three distinct domains with extremely different physicochemical properties (Fig. 9A). The N-terminal domain (residues 1-61) is highly conserved and includes five out of the seven eleven-residue repeats in the α-syn sequence. This region is disordered in solution, but the repeats confer an amphipathic character that assists a conformational shift into an α-helix in the presence of phospholipids and membranes [128]. This functional protein-lipid interaction might turn into pathological if the concentration of α-syn locally increases, since at a high protein-lipid ratio, membranes facilitate the nucleation step of α-syn amyloid formation [129]. Importantly, all identified α-syn missense mutations associated with clinical variants of PD and DLB (A30P, E46K, H50Q, G51D, A53E and A53T) map onto this membrane-binding segment. The central region of α-syn (residues 61-95) has a hydrophobic character and is known as the non-amyloid component (NAC) because it was identified as the second major component of brain amyloid plaques in AD [130]. This region is highly amyloidogenic when isolated and has been shown to fold into a Greek-key motif placed at the core of the mature fibrils (Fig. 9B) [131]. In the disordered and soluble state of α-syn, this aggregation-prone region is shielded from the solvent by transient intramolecular contacts. The C-terminal domain (residues 96-140) of α-syn contains 15 acidic residues and thus has a high negative charge density. It has been reported that C-terminal truncation exacerbates α-syn aggregation and cytotoxicity, which suggests that the electrostatic repulsions exerted by this region counteract α-syn aggregation [132]. Interestingly, this domain can be phosphorylated at tyrosines 125, 133, and 136 and at serine 129, impacting the charge distribution and polarity of this region. In addition, C-terminal truncated forms of α-syn account for up to 30% of total α-syn in the LB of PD patients [133]; thus, it is likely that these more aggregation-prone species play a relevant, yet unknown, role in pathogenesis.

As for most amyloidogenic proteins, the formation of α-syn fibrillar structures is preceded by the assembly of the monomeric protein into small and diffusible oligomers and protofibrils (Fig. 10) [44]. Multiple lines of evidence suggest that these metastable species, and not the mature fibrils, might be the main culprits for neuronal degeneration in PD [134].

These oligomeric and protofibrillar α-syn forms can be transmitted from affected to neighboring healthy neurons, where they can seed the aggregation of native α-syn in a prion-like manner (Fig.10) [135,136]. This prion-like transmission provides a plausible mechanistic framework for the Braak's hypothesis, which postulates that PD pathology may indeed originate in synapses of the peripheral nervous system (PNS) and invade the brain via retrograde axonal transport through the vagus [137,138]. This view is consistent with the presence of pathological α-syn aggregates in the PNS of patients long before PD diagnosis [139]. Importantly, the neuron-to-neuron propagated amyloid seeds seem to exhibit a conformational diversity similar to those observed for the strains of the prion protein. This diversity leads to the assembly of the endogenous protein into polymorphic fibrils that, despite sharing a common cross-β architecture, differ in their structural arrangement (Fig. 9B) [140]. As in the case of the prionopathies, the specific conformational features of these strains might result in a selective preference to invade and deposit in different brain cell types. This behavior would explain why the common aggregation of α-syn might result in different pathologies.
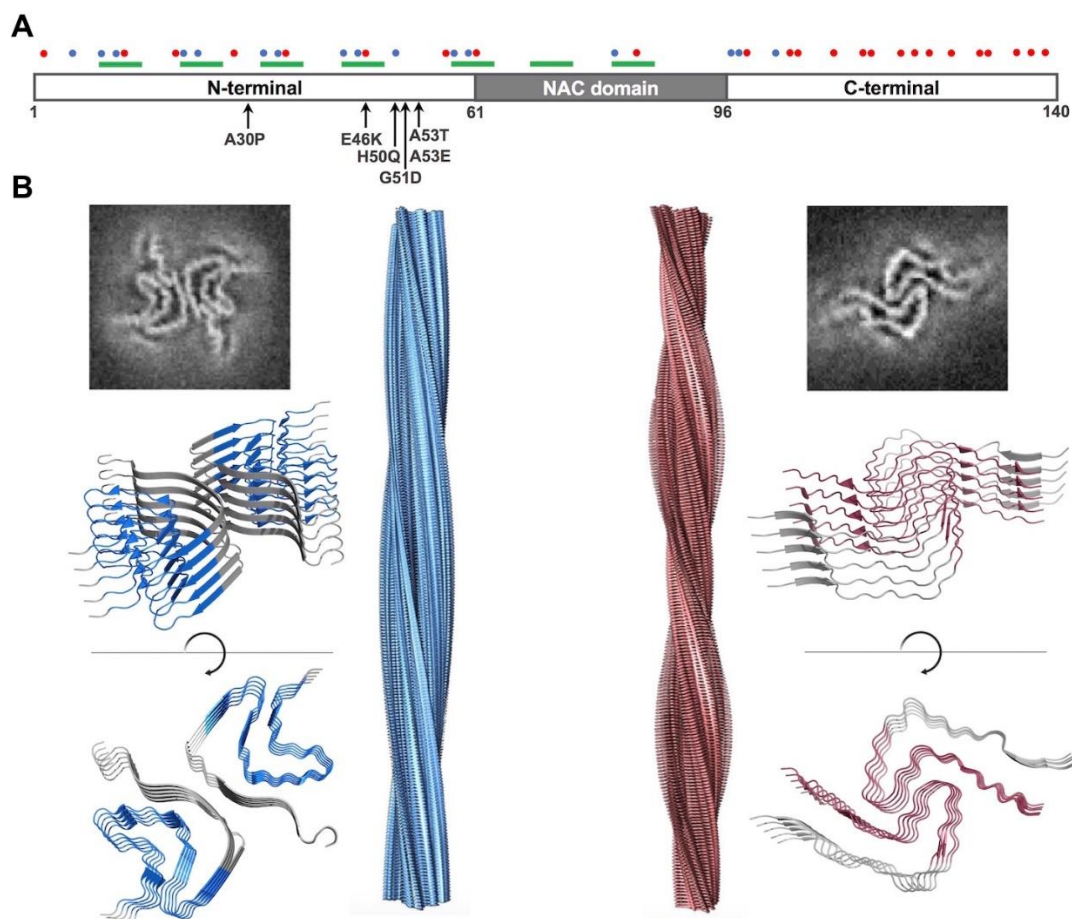
Figure 9. α-syn primary structure, cryo-EM structures and atomic models of α-syn strains. (A) A schematic representation of the α-syn primary structure, domains and specific mutations involved in familial versions of PD. Positively and negatively charged positions are shown as blue and red dots, respectively, and repeats positions are shown with a green line. (B) In the upper panels, the electron cryo-microscopy images of two α-syn fibril strains are shown. In the lower panels, the side and top views of the atomic models of the corresponding strains are shown, with the backbone represented as ribbons and the NAC regions colored. In the center, the macromolecular modeling of the amyloid fibril structure for both strains is shown, adapted from [32].

It is important to state here, that despite the association of LB pathology with clinical dysfunction is well established, roughly 20 % of people over 70 years display α-syn inclusions in their brains, without associated motor or memory problems. In addition, LB also appear in unrelated neurodegenerative disorders, like AD. This has led to hypothesize that LB might be a way to accumulate an excess of α-syn in an inert compartment and not the cause of PD [141]. This should be kept in mind, in case the α-syn centric strategies we describe in the next sections fail to demonstrate clinical efficacy.
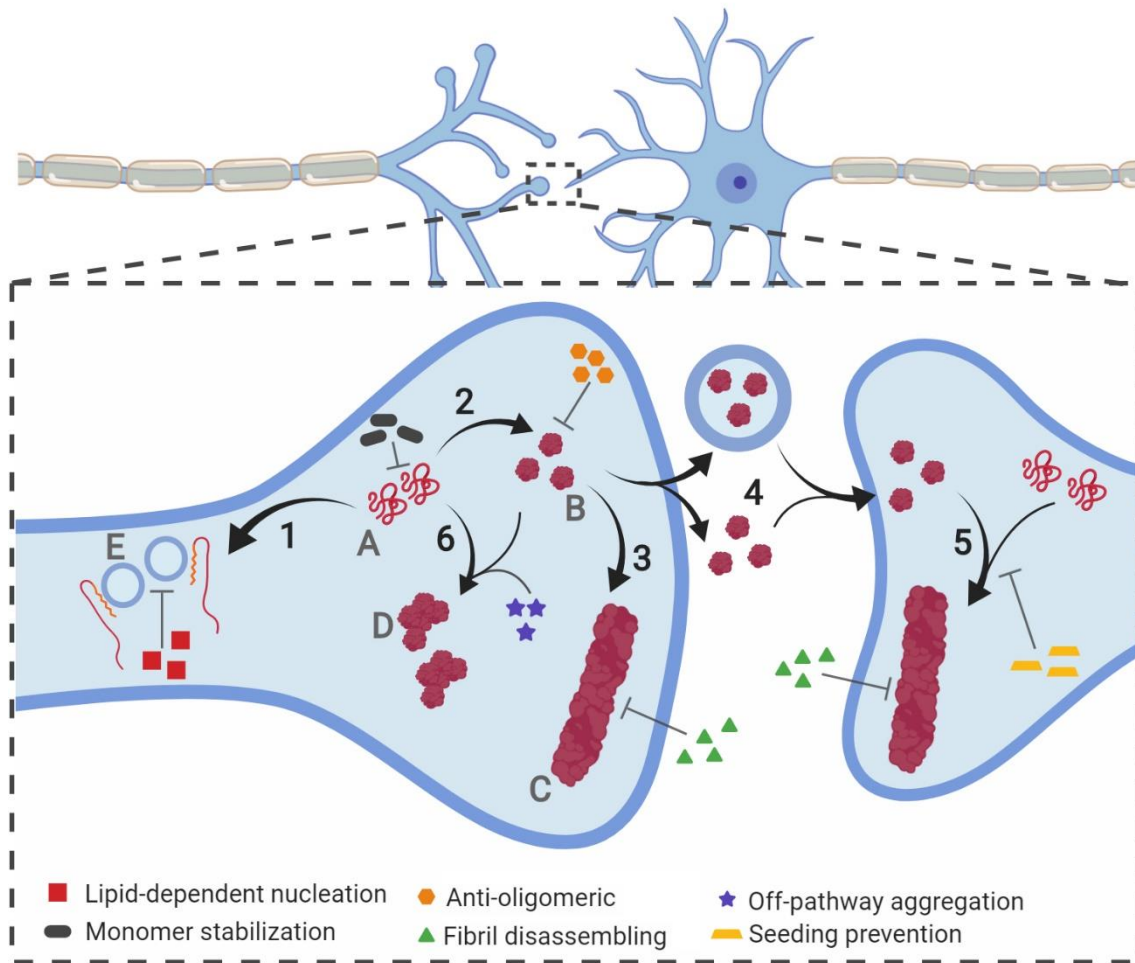
Figure 10. Schematic representation of the aggregation reaction of α-syn. The aggregation of α-syn follows a slow process that comprises different assemblies of the protein: monomeric (A), oligomeric (B) and fibrillar (C) or amorphous (D) aggregates. Accordingly, α-syn undergoes different processes during the onset and progression of PD: protein-lipid interaction (1), oligomerization (2), fibril elongation (3), transmission (4), seeding (5), and amorphous aggregation (6). Colored symbols aim at illustrating the stages of α-syn aggregation by targeting small molecules, see legend.

Targeting α-synuclein aggregation

Different α-syn-targeted strategies and tools have been developed in recent years to modulate its misfolding and aggregation in the context of synucleinopathies. These strategies include:1) SNCA gene-silencing approaches to knock out or decrease the neuronal levels of the protein, 2) strategies to increase the clearance of soluble or aggregated α-syn by potentiating autophagic and proteasomal activities, and 3) agents aimed at preventing the formation and/or propagation of the toxic aggregated species [142,143]. Different agents targeting the aggregation of α-syn in the central and peripheral

nervous systems are also in early-stage development [144]. They include antibodies, vaccines, molecular chaperones and small molecules. Despite their selectivity, protein-based therapeutics usually find difficulties in passing the blood-brain barrier (BBB) and may trigger collateral immunological reactions. Therefore, small molecules are still the default option for targeting brain diseases [145]. Although conventional structure-based drug design initiatives are hampered by the disordered nature of monomeric α-syn and the lack of high-resolution structures of toxic oligomers, a handful of small molecules have already shown promising activity in animal models of PD, and some of them have already entered into clinical trials.

In the subsequent sections, the leading small-molecule-based initiatives to tackle α-syn aggregation in PD will be dissected (Fig. 11). These small molecules constitute a group of structurally diverse compounds that are discussed in terms of their molecular origin, their mechanism of action, the strategies used for their identification and their current status of development (Fig. 11).
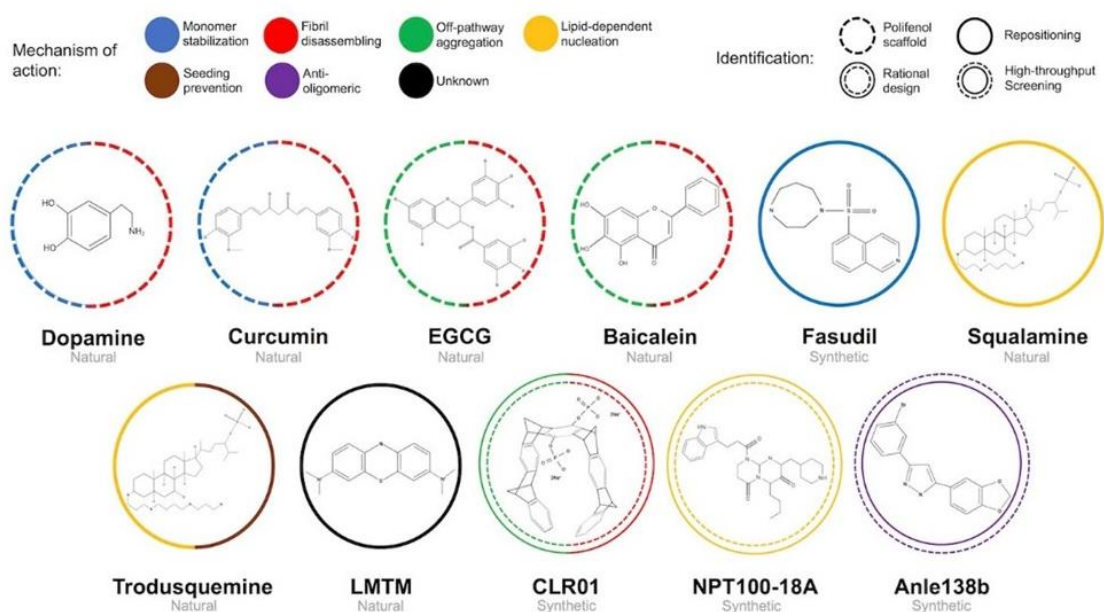

Figure 11. Chemical structures. Molecular structures of selected modulators of α-syn aggregation, grouped as a function of molecular class and the mechanism of action.

Natural compounds: establishing the foundation for α-syn-targeted therapies for Parkinson's disease

Natural compounds represent a major source of medications and important structural scaffolds to develop new drugs, many of them interacting specifically with biological macromolecules, mainly proteins. The capacity to interfere in α-syn aggregation with individual small molecules was reported for the first time in the early 2000s. The analysis of 169 molecules revealed that several natural catecholamines, including dopamine (DA) L-dopa, epinephrine, and norepinephrine, were able to inhibit the formation of α-syn amyloid fibrils [146]. Oxidized derivatives of DA diverted the aggregation of α-syn to off-pathway aggregates. This activity seems to depend on the interaction of DA with the $_{125}$YEMPS$_{129}$ sequence stretch at the C-terminus of the α-syn monomer [147,148]. Unfortunately, dopamine-induced oligomeric species triggered brain damage in animal models of the disease, which excluded dopamine and its derivatives as disease-modifying therapies for PD [149] while questioning the status quo of dopamine replacement as the prevailing symptomatic treatment for PD. Despite these limitations [150], the discovery of a first molecular entity modulating α-syn aggregation through a direct protein-compound interaction paved the way for future drug discovery programs.

In the following years, novel naturally existing small molecules acting as inhibitors of α-syn fibrillogenesis were discovered [151–153]. Among them, we highlight the polyphenols. These molecules display several phenyl rings and hydroxyl groups that seem to act coordinately, binding aggregates through π interactions and destabilizing them by disrupting the hydrogen bond network of the β-sheet. The number of phenolic moieties and the position of the hydroxyl groups are critical for their activity. Curcumin, epigallocatechin-3-gallate (EGCG), baicalein, rosmarinic acid, kaempferol, tanshinones, tannic acid, myricetin and quercetin all belong to this family [154,155]. Similar to DA, many polyphenolic molecules interact with the charged and disordered C-terminus of α-syn [156–158], and indeed, several of these compounds also display a catechol moiety. These compounds attenuate α-syn amyloidogenicity by facilitating intramolecular contacts and decreasing bimolecular self-assembly, driving the oligomerization process to the formation of nontoxic aggregates with low β-strand content, remodeling preformed toxic oligomers or disentangling mature fibrils. Remarkably, several polyphenols, such as curcumin, baicalein, myricetin and EGCG, combine two or more of these activities [158–161]. Several polyphenols have been assayed in animal models of the disease. As a general

trend, their administration resulted in neuroprotection and amelioration of motor deficits [154,162–167]. However, the data do not prove a direct connection between the observed protective effect and the anti-aggregation properties of polyphenols, mainly because the large majority of these studies have been performed in neurotoxin-induced models of PD, and there is only weak evidence of target engagement. Indeed, it cannot be disregarded that the antioxidant and anti-inflammatory properties of polyphenols counteract neurotoxin-induced brain damage, in addition to or instead of anti-aggregation activities [168,169]. Curcumin and its derivatives deserve special attention since they are one of the few polyphenols tested in a transgenic animal model of PD (Syn-GFP mice). The study reported motor and behavioral amelioration, although target engagement remained unclear [170]. Remarkably, the neuroprotective properties of EGCG were validated in 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP)-intoxicated monkeys [171], and the molecule reached phase II clinical trials for the treatment of PD[I]. The study comprised a total of 480 PD-suffering individuals who were randomized and treated with up to 1.2 grams/day of green tea polyphenols (GTP) or placebo. After 6-month treatment with GTP, the unified PD rating scale (UPDRS) score, which evaluates both motor and non-motor symptoms, was significantly improved, indicating that GTP provide a mild symptomatic benefit in early untreated PD. These results illustrate the potential of natural molecules as therapeutics for PD.

Drug repositioning: a shortcut for α-syn-targeted therapies

Drug discovery is characterized by high attrition rates, mainly attributed to a lack of pharmacological efficacy, deficient safety profiles or both. In addition to its intrinsic high risk, the economic investment required for such research initiatives is huge, averaging $5 to $10 million per molecule. In this context, drug repositioning offers an opportunity to reduce the risks and costs of therapeutic development. Successful therapies in cancer and cardiovascular and neurodegenerative diseases rely on this strategy [172]. Briefly, drug repositioning consists of the identification of existing compounds that are already developed or even licensed for a different indication [173]. Accordingly, the safety of these molecules is guaranteed since all of them have undergone phase I clinical trials. Taking advantage of this rationale, fasudil, squalamine, trodusquemine and leuco-methylthioninium bis(hydromethanesulphonate) (LMTM) have been repurposed as therapeutic candidates for PD.

Fasudil is an inhibitor of the human Rho kinase [174], whose overactivation is responsible for several types of vascular disorders [175]. This small molecule has been approved to treat cerebral vasospasm in China and glaucoma in Japan [176]. This compound crosses the BBB and was shown to induce neuroprotection in MPTP-intoxicated mice, increasing the survival of dopaminergic neurons, an effect that was initially attributed to its activity as an inhibitor of brain Rho kinase [177]. A subsequent study demonstrated that fasudil interacts with the aromatic side chains of Tyr133 and Tyr136 at the C-terminus of monomeric α-syn. Fasudil binding to α-syn decreases the nucleation and elongation constants of the *in vitro* aggregation reaction and reduces intracellular accumulations of α-syn in a cellular model of the disease, all in a concentration-dependent manner [178]. The administration of the molecule to A53T α-syn transgenic mice demonstrated target engagement, with a significant reduction in α-syn deposits and cognitive rescue, although motor improvement was poor.

Squalamine and trodusquemine are a pair of steroid-polyamine molecules originally isolated from the dogfish shark *Squalus acanthias* due to their antimicrobial activity [179,180]. Remarkably, they account for more than 15 clinical trials, including a phase III trial to test squalamine efficiency in the treatment of age-related macular degeneration [181]. These compounds behave as cationic lipids, interacting with the inner leaflet of the plasma membrane [182,183] and destabilizing protein-membrane contacts [184]. Polyamines also have the ability to interact with the negatively charged C-terminus of α-syn [185]. Thus, it was not surprising that both compounds were able to interfere with the oligomerization of α-syn at the membrane moiety, restoring its random coil conformation, promoting its dissociation and reducing the amount of membrane-bound oligomers, with all these effects contributing to avoid subsequent lipid-induced α-syn aggregation [186,187]. Their activity was confirmed in a neuronal cell model, where they alleviated mitochondrial dysfunction. In addition, both molecules reduced the number of intracellular α-syn aggregates and increased motility and lifespan in a *Caenorhabditis elegans* (*C. elegans*) model of PD. However, only trodusquemine can bind to the surface of fibrils, interfering with secondary nucleation events, which makes this compound more effective. In addition, in contrast to squalamine, trodusquemine can cross the BBB and is better suited for therapeutic indications.

An additional case of successful repositioning is methylthioninium chloride (MTC), also known as methylene blue. This phenothiazine, together with its reduced stable LMTM

variant, is an inhibitor of the aggregation of the Tau protein [188,189], which has reached phase III clinical trials as a therapy for AD [190]. Based on this evidence, MTC and LMTM were also validated as potent inhibitors of α-syn in a cellular model of PD and in the L58 and L62 *wild type* α-syn transgenic mice lines, where their administration reduced brain α-syn aggregates [191]. Despite the lack of mechanistic information on the nature of the protein-drug interaction, the fact that LMTM enhances both autophagy and mitochondrial respiration [192], while displaying optimal toxicological and pharmacokinetic properties, makes this molecule a particularly interesting candidate for further clinical study in PD.

It is likely that the safety profiles of the above-described drugs would allow them to enter into the clinics very soon.

Rational drug design: de novo discovery of α-syn aggregation inhibitors

The intrinsically disordered nature of α-syn and the complexity of the aggregation process preclude the use of conventional drug discovery strategies, which typically exploit 3D protein structures to design drugs that dock into protein binding pockets or protein-protein interaction sites [193–195]. In the case of α-syn, rational discovery usually targets protein residues involved in the weak non-covalent intermolecular contacts that initiate α-syn dimerization and/or oligomerization. CLR01 and NPT100-18A are two structurally unrelated small molecules that illustrate the success of the approach.

The molecular tweezer CLR01 displays a curved structure that creates a concave negatively charged cavity. This design allows CLR01 to bind specifically to exposed Lys side chains through transient interactions. CLR01 has been shown to efficiently inhibit several amyloidogenic proteins [196,197], being able to block α-syn aggregation in a dose-dependent manner at substoichiometric ratios [198]. In addition, mature α-syn fibrils became dismantled after incubation with a 10-fold excess of CLR01. The primary site of interaction for CLR01 is well defined and involves Lys10 and Lys12 at the N-terminus of α-syn. The compound reduces the inherent toxicity of α-syn aggregates in cellular models and in a transgenic zebra fish model of PD. The intracerebroventricular and subcutaneous administration of CLR01 in a transgenic mouse model of PD (Thy1-αsyn) ameliorated motor dysfunction. However, it did not reduce aggregated α-syn levels [199]. CLR01 exhibits favorable pharmacokinetics and toxicity profiles [200], but the lack of demonstrable target engagement in a mammalian model of the disease might limit its further development.

Another remarkable example of rational design is the NPT100-18A molecule from Neuropore and its lead derivative NPT200-11, developed in partnership with UCB Pharma [201,202]. NPT100-18A is a compound aimed at interfering with α-syn dimerization, developed on the assumption that α-syn oligomerization could be initiated by the formation of dimers on membrane surfaces [203–205]. Molecular dynamic simulations suggested the existence of a pharmacophore responsible for dimerization, comprising residues 96-102 at the α-syn C-terminus [204,205]. NPT100-18 is the best performer from a chemical library of 34 peptidomimetic compounds targeting the predicted protein-protein interaction region. NPT100-18A precludes the formation of α-syn mature aggregates *in vitro* and releases monomeric α-syn by destabilizing the oligomers formed in the presence of a membrane mimetic environment. NPT100-18A reduces the accumulation of aggregates and their subsequent toxicity in primary neuronal cells overexpressing α-syn. The administration of this molecule to E57K α-syn transgenic mice revealed neuroprotection in several brain regions, including the neocortex, hippocampus, and striatum, as a result of an overall decrease in the aggregated and oligomeric species of α-syn. Unfortunately, no concluding results were obtained for the substantia nigra [201]. Despite encouraging target engagement and symptomatic amelioration, NPT100-18A pharmacokinetic properties were far from optimal, with poor BBB permeability and a low concentration of the bioactive molecule in the brain. In an effort to bypass these hit limitations, Neuropore developed the lead compound NPT200-11, a nonpublic structure, increasing the C_{max} and brain plasma exposure ratio by 129- and 66-fold, respectively. This allowed the attainment of optimal brain concentrations in mice after an oral administration of 10 mg/kg. This pharmacokinetic optimization did not significantly affect target engagement or symptomatic amelioration [202]. Notably, in 2016, NPT200-11 successfully completed a clinical trial phase I involving 55 healthy participants[II].

Expanding the molecular repertoire: high-throughput screening of large drug-like collections

A powerful alternative for the discovery of novel active molecules is the high-throughput screening of large collections of chemically diverse compounds. These initiatives necessarily rely on optimized assays that attempt to minimize time and cost per test while preserving specificity and sensitivity. The stochastic nature of protein aggregation impacts experimental reproducibility and thus compromises hit selection using this kind of "brute force" approach. Notably, by optimizing α-syn aggregation protocols and/or

integrating new detection methods, several groups managed to overcome this limitation and discovered a set of novel α-syn aggregation inhibitors [206–209]. The following presents the examples of Anle138b and SynuClean-D.

Anle138b was discovered after two consecutive rounds of screening. A first assay evaluated the capacity to inhibit prion protein (PrP) amyloid assembly, rendering 3,5-diphenyl-pyrazole (DPP) the most active compound among approximately 20,000 chemically diverse drug-like compounds [210]. Keeping the DPP scaffold, the authors designed a set of 150 molecules in a structure-activity relationship (SAR) effort intended to balance brain permeability and anti-aggregation activity. Anle138b was selected as the leading compound of this initiative. This molecule does not interact with the α-syn monomer but instead binds to a hydrophobic pocket in oligomeric assemblies, interfering with β-sheet formation. Remarkably, the hydrophobicity of the binding site impacts the fluorescence emission spectrum of Anle138b, which allows its use to detect early-stage aggregates, with potential implications for PD diagnosis [211]. The therapeutic performance of Anle138b was evaluated in three different PD mouse models, in which oral administration of the compound ameliorated PD-related symptoms, including motor deficits, gut motility, weight, neuroprotection and survival. In the A30P transgenic model, Anle138b was shown to reduce α-syn aggregation in the brain, linking this phenotype with the observed neuroprotection. Importantly, a therapeutic effect was also observed in symptomatic late-stage rodents, suggesting that Anle138b might be used not only for prevention or in early-stage patients but also for symptomatic cases of PD [212]. Moreover, preclinical analysis suggests that the molecule is innocuous, even after long-term and high-dose treatment. In addition, it displays adequate pharmacokinetics, brain distribution and pharmacodynamic properties. Thus, it is likely that we will soon see clinical trials for this promising compound.

Work 4

High-throughput screening methodology to

identify α-synuclein aggregation inhibitors

High-throughput screening methodology to identify alpha-synuclein aggregation inhibitors

Jordi Pujols[1,2,‡], Samuel Peña[1,2,†], María Conde-Giménez[3], Francisca Pinheiro[1,2], Susanna Navarro[1,2], Javier Sancho[3], Salvador Ventura[1,2,*]

[1] Department of Biochemistry and Molecular Biology , Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain

[2] Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain

[3] Department of Biochemistry and Molecular and Cellular Biology, Instituto Universitario de Investigación en Biocomputación y Física de Sistemas Complejos, Zaragoza, Spain

[*] Correspondence: salvador.ventura@uab.cat; Tel.: +34-93-5868956

[‡] These authors contributed equally to this work.

Abstract

An increasing number of neurodegenerative diseases are being found to be associated with the abnormal accumulation of aggregated proteins in the brain. In Parkinson's disease, this process involves the aggregation of alpha-synuclein (α-syn) into intraneuronal inclusions. Thus, compounds that inhibit α-syn aggregation represent a promising therapeutic strategy as disease-modifying agents for neurodegeneration. The formation of α-syn amyloid aggregates can be reproduced in vitro by incubation of the recombinant protein. However, the in vitro aggregation of α-syn is exceedingly slow and highly irreproducible, therefore precluding fast high throughput anti-aggregation drug screening. Here we present a simple and easy to implement in-plate method for screening large chemical libraries in the search for α-syn aggregation modulators. It allows to monitor aggregation kinetics with high reproducibility, while being faster and requiring lower protein amounts than conventional aggregation assays. We illustrate how the approach allowed to identify strong aggregation inhibitors in a library of more than 14.000 compounds.

Keywords: High-throughput screening; α-synuclein; Parkinson disease; Amyloid; Protein aggregation

Introduction

Protein misfolding and amyloid aggregation is behind a growing number of human diseases, including Parkinson's disease (PD) [1]. PD is the second most common neurodegenerative disorder, after Alzheimer's disease (AD) and is still incurable. PD is characterized by protein deposition in intraneuronal inclusions, the so-called Lewy bodies (LB) and Lewy neurites (LN) [2,3], whose major component is α-synuclein (α-syn) [4], a 140 amino acid presynaptic protein encoded by the SNCA gene and normally found in both soluble and membrane-associated fractions of the brain [5,6]. The protein α-syn is a central component in PD pathogenesis and accordingly a privileged target for therapeutic intervention. In vitro, under physiological conditions, α-syn assembles into aggregates that are structurally similar to those found in the inclusions of disease-affected brains [7,8] . The aggregation process is thought to start from soluble monomers that polymerise into ring-shaped and string-like oligomers. These small structures coalesce to form protofibrils that assemble into insoluble fibrils[9,10]. The precise nature of the toxic α-syn species is still unclear, although it is believed that specific oligomeric species play a key role in neuronal toxicity, rather than the mature aggregates [11,12]. It is though that the population of these small oligomeric species is also associated with the spread of the disease between different structures in the brain [13,14].

There is strong interest in the discovery of small compounds that can act as chemical chaperones modulating the aggregation of α-syn [15-20]. In the absence of a defined 3D-structure to target, screening of large collections of chemically diverse compounds is a useful approach toward the discovery of novel bioactive molecules exhibiting an α-syn anti-aggregational effect. Chemical kinetics approaches would allow the quantitative detection of the effects of potential therapeutic molecules on aggregation [21]; however the application of this type of analysis is hampered by the low reproducibility of aggregation reactions, resulting in dissimilar kinetic parameters and/or high errors even within replicates in the same aggregation assay. This especially true for α-syn, a protein displaying a very slow aggregation reaction, usually taking several days, which is highly influenced by factors like pHs, temperature, agitation or the presence of impurities [18-20,22-31]. The lack of reproducibility between aggregation curves is a strong limitation to identify *bona fide* aggregation inhibitors, since their potency becomes hidden in overlapping errors bars, especially at the beginning of the reaction, where the more toxic

oligomeric species are expected to be formed. The slow aggregation kinetics of α-syn is also an important time limitation for large-scale screening, where several thousands of potential inhibitors should be tested. Due to the dependence of the reaction on the initial protein concentration, the aggregation of α-syn can be accelerated by increasing this parameter. However, this means that very large amounts of protein will be necessary for high-throughput screening assays.

The aim of the present work is to provide a detailed aggregation kinetics protocol suitable for the large-scale screening of aggregation modulators that can be used without requiring extensive previous expertise in protein aggregation and/or in the manipulation of α-syn. By ensuring a high purity of the recombinant protein and performing protein aggregation assays in 96-well plates in presence of teflon polyballs, the fibrillation reaction is boosted, requiring times and protein quantities compatibles with high-throughput screening. After optimizing agitation and temperature, we obtained highly reproducible kinetics that allowed to derivate accurate aggregation constants. We illustrate how the approach permitted identifying strong inhibitors after screening a library of more than 14.000 compounds.

Results

2.1 Protein expression and purification

For protein expression and purification, we adapted a protocol from Volles and Lansbury [32], including an additional sonication step during cell lysis and, more importantly, a final anion exchange chromatography (Fig. 1). This purification step is crucial, since not only it increases homogeneity, but also avoids the co-elution of nucleic acids. α-syn binds to nucleic acids, which concentrations and identities might vary from preparation to preparation. Because most labs monitor the purity of their α-syn preparations using SDS-PAGE and protein staining, nucleic acids are not visualized. The use of spectrophotometry to discard nucleic acid contaminations is highly advisable, since, in our hands, their presence results in a large heterogeneity in the kinetics of aggregation reactions.

Following this protocol, we obtained high amounts of nucleic acid free α-syn (35 mg/L culture). Purity of α-syn was checked by SDS-PAGE (Fig. 2) and mass spectrometry, obtaining, a band and a peak that corresponds to the theoretical weight of α-syn, respectively, without any trace of contaminant proteins.

Aggregation kinetics

One of the most important hallmarks of α-syn aggregation is the long times it needs to be completed. When compared with other amyloid-forming proteins, the time required for the assembly of detectable mature fibrils is significantly longer, in such a way that at concentrations below 100 μM, the reaction exhibits a large lag phase, is only completed in one or more weeks and displays low reproducibility [18]. Because the aggregation ratios are extremely dependent on the initial protein concentration, the aggregation of α-syn has been usually accelerated by dramatically increasing its concentration in the assay, up to 400 μM. High concentration also reduces erratic fibrillation [9]. However, the drawback is that very large amounts of protein are necessary for large-scale screening (0,75 mg of α-syn for a single 150 μL reaction at 400 μM) [23,24].
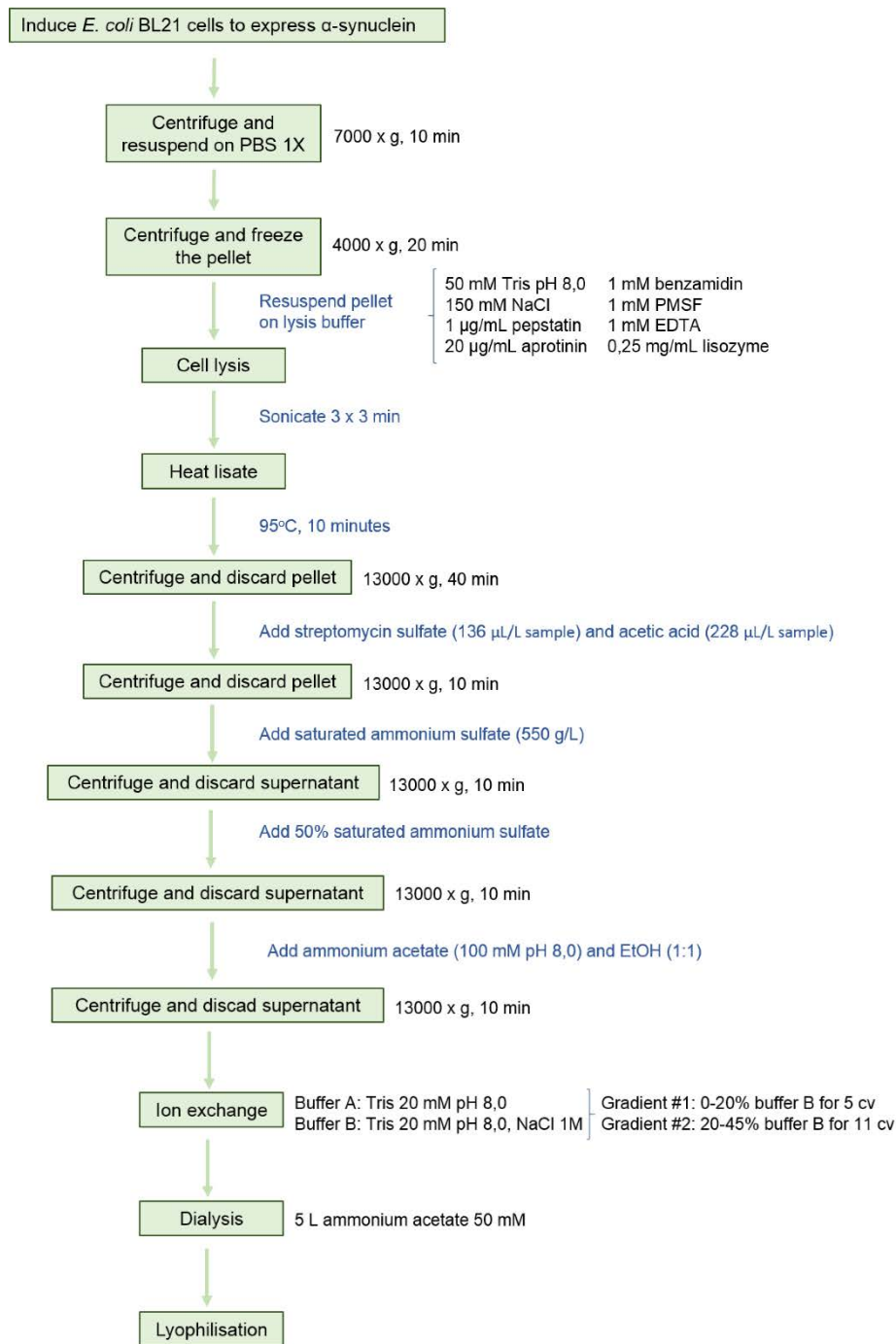
Figure 1. General strategy for the purification of α-synuclein and SDS-PAGE of the purified protein

Here we present a protocol, in which, despite we use concentrations of purified α-syn <100 μM, the sigmoidal aggregation reactions are completed in approx. 24h. To track the aggregation progress, we took advantage of the amyloid specific reporter Th-T, added to each sample at 40 μM as final concentration. One of the main problems when performing kinetics, is the loss of the mixture homogeneity as the reaction progresses. Most of the assays are performed on Eppendorf or glass tubes, from where aliquots are removed at certain times and diluted for fluorescence measurements in a quartz cuvette. Dilutions do

not necessarily represent the original species population, and even when samples are sonicated or re-mixed, it is impossible to predict how representative is the amount of aggregated we are taking out from the solution. To reduce variability, caused by both sample manipulation and aggregates heterogeneity, in our case, measurements were directly recorded in a 96-well plate (black plastic) using a plate reader.
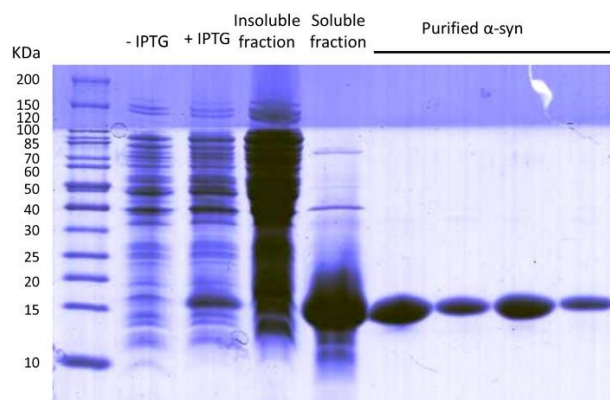

Figure 2. SDS-PAGE of the expression, fractionation and purification process. –IPTG and +IPTG lines correspond to-non induced and induced cell extracts. The insoluble and soluble fractions of induced cells are shown as well as different fractions eluting from the anion exchange column.

Agitation has been shown to accelerate α-syn aggregation up to 1 order of magnitude. The effect of the agitation is implicitly or explicitly attributed to mass transfer or fibril fragmentation. However, for α-syn, despite agitation with only air resulted in somehow accelerated kinetics, in our hands, the reactions were inconsistent from well to well and even resulted in different aggregates morphology. The use of glass beads during agitation potentially increases the air–water interface area, however their effect on the speed and reproducibility of the reaction is controversial [33,34]. Investigations have demonstrated the sensitivity of aggregation reactions to hydrophobic-water interfaces and in particular to Teflon-water interfaces [35] and indeed Teflon beads have been used in a number of α-syn aggregation studies [36,37]. Therefore, we incorporated a Teflon bead (2 mm diameter) in each well and incubated the plates under continuous agitation in an orbital shaker, since it provides more homogeneous ball agitation than a linear one. The presence of these little spheres accelerated dramatically the kinetics of α-syn aggregation, allowing to work with 70μM α-syn in a final sample volume of 150 uL, thus reducing the amount of protein needed per reaction (0,145 mg/well). No aggregates were observed to form at the same protein concentration and conditions during the first 24 h in the absence of

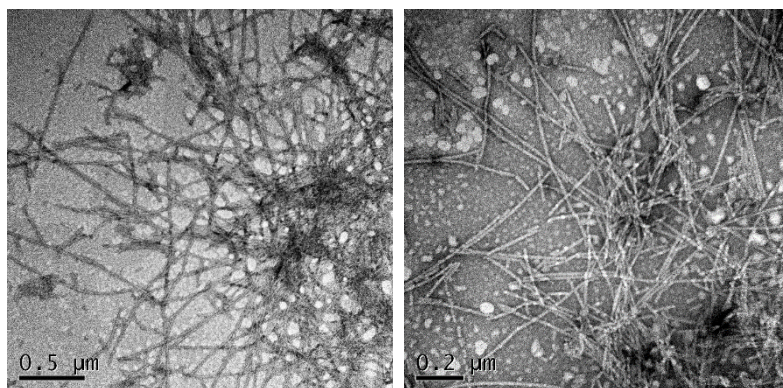mixing balls, whereas in their presence typical long unbranched amyloid fibrils could be observed by TEM (Fig. 3).

Figure 3. α-synuclein fibrils formed in the presence of Teflon beads. TEM images were collected upon incubation of 70μM soluble α-syn for 24 h with agitation in the presence of beads. The samples were briefly sonicated before imaging.

One remarkable advantage of the protocol is the capacity to simultaneously handle a high number of plates by fixing them on a conventional orbital culture incubator, instead of using the plate reader itself to incubate the plates one by one. In this way, we can achieve a smooth, prolonged and continuous agitation, in contrast to the short and intermittent one (commonly only 15 seconds before each measure) performed by conventional plate readers. Moreover, it is possible to adjust accurately the temperature (37ºC) and agitation conditions (100 rpm), enabling to acquire well-to-well and plate-to-plate reproducible aggregation curves. Plates are only pulled out from the incubator for fluorescence reads and rapidly returned to their original location. Together, the method reduces both protein concentration and aggregation times, abolishes most sample manipulation, allows simultaneous assays and maximizes reproducibility (Fig. 4A). The accuracy of the protocol can be observed in Figure 4B, where it is shown how plotting Th-T fluorescence signal as a function of time for two different aggregation reactions, recorded in non-consecutive days, results in sigmoidal curves that overlap almost perfectly. In this way, when k1 (nucleation rate constant) and k2 (growth rate constant) were calculated using the Finke–Watzky two-step model for the two curves [38] we obtained $k1_1 = 0.00640 \pm 0.0013$ h−1, $k1_2 = 0.00645 \pm 0.0016$ h−1 and $k2_1 = 0.3623 \pm 0.0264$ h−1, $k2_2 = 0.3365 \pm 0.0336$ h−1, with correlation coefficients R> 0.998, in both cases. As expected the relative standard error is larger for k1 than for k2, indicating that nucleation is inherently more stochastic than elongation. This is the reason why the use of preformed amyloid seeds usually results in highly reproducible kinetics [39]; however, this precludes

the identification of compounds that target the nucleation phase of amyloid formation. The applicability of this methodology was validated by screening a large chemical library. We used the Maybridge HitFinder Collection, containing 14,400 compounds, in the search for putative inhibitors of α-syn aggregation. Each compound was tested in triplicate and each plate contained a triplicated α-syn control devoid of compound. Discarding false positives caused by Th-T fluorescence quenching during the data collection was essential. For this purpose, we recorded the absorbance spectra from 400 nm to 600 nm for any putative positive compound, and discarded those absorbing either at the Th-T excitation or emission wavelengths, 450 and 480 nm respectively. The goodness of the approach is illustrated in Figure 4 C, with two compounds (D and G) that exhibit medium and high effect on final Th-T fluorescence. It can be observed the low error bars obtained for all the reactions, which results in high statistical significance when assessing the inhibitory potency of the compounds at any time point. Accordingly, fitting the experimental data to the Finke–Watzky curve resulted in the large majority of cases in correlation coefficients R> 0.985. The ability of the method to identify bona fide inhibitors requires an orthogonal validation. In our case it was confirmed by visualizing control and compound-treated reactions end points by TEM. As an example, it can be seen in Figure 4 D-F how the presence of compounds D and G results in an important decrease in both number and size of α-syn fibrils. Centrifugation assays indicated higher levels of soluble α-syn at the end of the reaction in the presence of inhibitors. The screening rendered 47 novel compounds able to efficiently modulate the aggregation and inhibit fibril formation of α-syn. These compounds belong to different chemical families, supporting the versatility of the assay. The chemical structures and the activities of 10 representative molecules are illustrated in Table 1. After fitting their respective aggregation curves, the 47 compounds fall into one of two classes: those that decrease the final Th-T fluorescence without reducing K1 and K2, like compound G, and those that decrease both the fluorescence signal and the kinetic constants, like compound D. They are those last compounds that exhibit pharmacological interest, since they are expected to delay the onset of the aggregation reaction. A convenient way to discriminate between these two types of molecules in large-screening assays is to compare the halftime of aggregation (t50) [40].
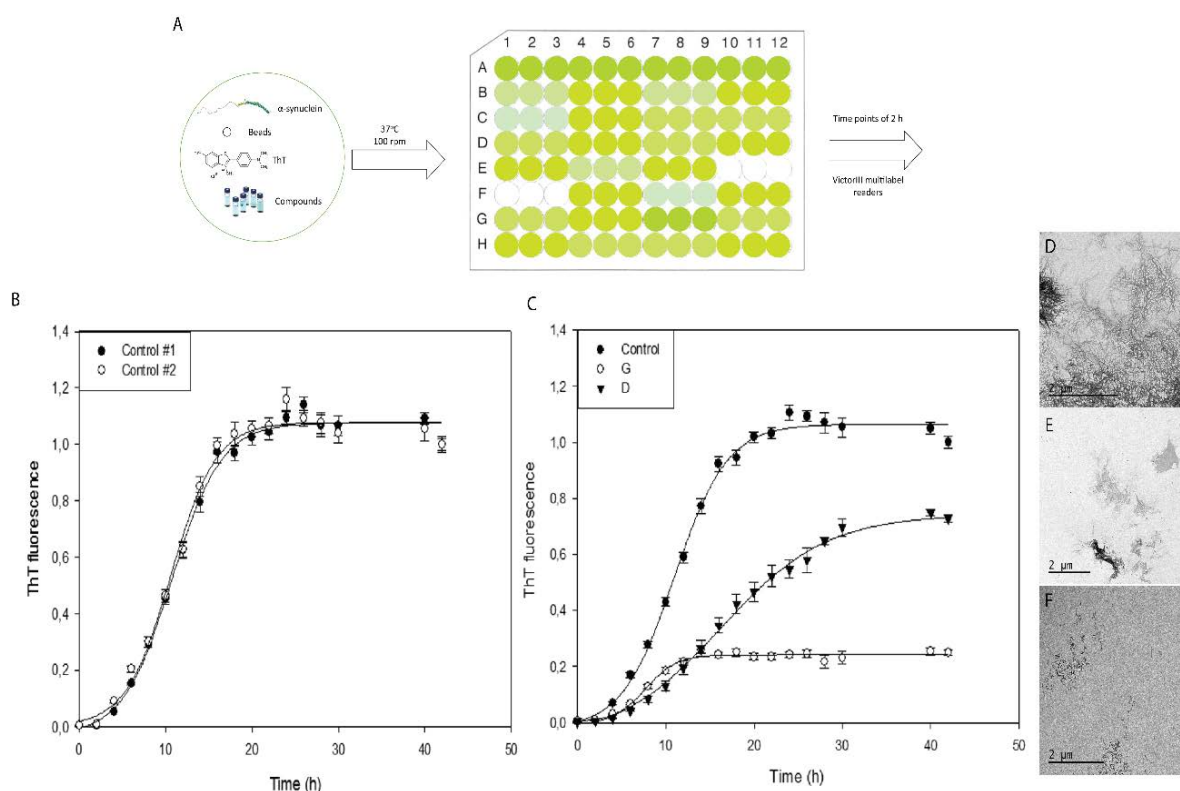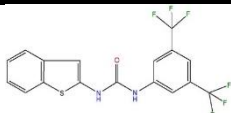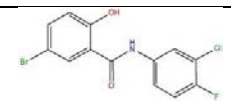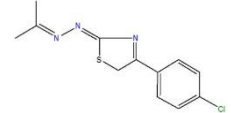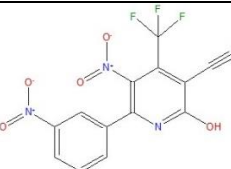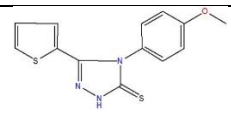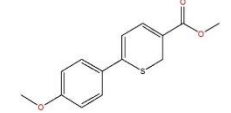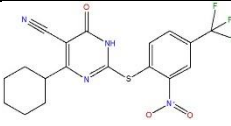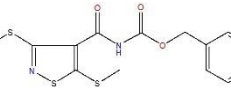
Figure 4. Aggregation kinetics, from plate preparation to putative inhibitors. A) General scheme for plate preparation and incubation. B) α-synuclein aggregation kinetics performed on non-consecutive days. C) α-synuclein aggregation kinetics in presence of putative inhibitors. Measured by Th-T fluorescence emission, represented as normalised means. Error bars are represented as standard error. D), E), F) TEM images of α-synuclein fibrils in absence (D) and presence of inhibitors D (E) and G (F).

The present method not only provides an effective system for screening, but can be used to set up a variety of experiments. For instance, when developing new anti-aggregational drugs, it is essential to titrate their concentration dependent activity. Many promising potential drugs fail because they are not effective at a concentration compatible with their pharmaceutical application. The high reproducibility of the method allows to perform accurate tit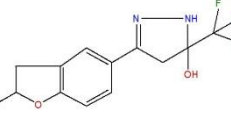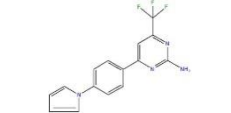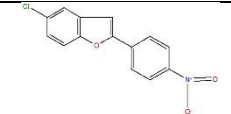ration assays, as illustrated for compound D in Figure 5 where a clear dose-dependent anti-aggregation activity could be tracked for this molecule, which turns to be active even at sub-stoichiometric protein:compound ratios. However, we should clarify that the method only allows to detect α-syn species that can be stained with Th-T and thus it fails to monitor the impact of the compounds in the formations of early oligomers.

Table1. Representative active compounds identified in the high-throughput screening

Name	Code	Structure	% Inhibition[a]	Δt_{50} (h)[b]

N-(1-benzothiophen-2-yl)-N'-[3,5-bis(trifluoromethyl)phenyl]urea	A		79,6	1
5-bromo-N-(3-chloro-4-fluorophenyl)-2-hydroxybenzamide	B		74,4	5
1-[4-(4-chlorophenyl)-2,5-dihydro-1,3-thiazol-2-yliden]-2-(1-methylethylidene)hydrazine	C		76,4	-3
2-hydroxy-5-nitro-6-(3-nitrophenyl)-4-(trifluoromethyl)nicotinonitrile	D		32,4	3
4-(4-methoxyphenyl)-5-(2-thienyl)-2,4-dihydro-3H-1,2,4-triazole-3-thione	E		59,8	3
methyl 6-(4-methoxyphenyl)-2H-thiopyran-3-carboxylate	F		93,0	1
4-cyclohexyl-2-{[2-nitro-4-(trifluoromethyl)phenyl]thio}-6-oxo-1,6-dihydropyrimidine-5-carbonitrile	G		72,7	-2
3-(2-methyl-2,3-dihydro-1-benzofuran-5-yl)-5-(trifluoromethyl)-4,5-dihydro-1H-pyrazol-5-ol	H		31,7	1
4-[4-(1H-pyrrol-1-yl)phenyl]-6-(trifluoromethyl)pyrimidin-2-amine	I		0	11
5-chloro-2-(4-nitrophenyl)benzo[b]furan	J		44,2	1
benzyl N-{[3,5-bis(methylsulfanyl)-4-isothiazolyl]carbonyl}carbamate	K		67,7	5
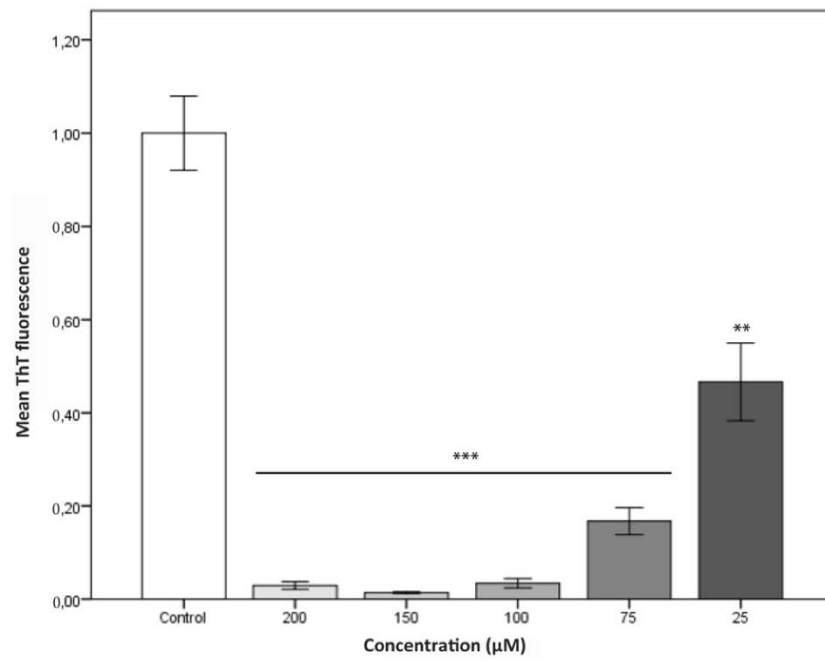
Figure 5. Inhibition of α-synuclein aggregation at different concentrations of the compound D. Error bars are represented as standard error.

Discussion

The present work provides a simple and accurate method for screening large chemical libraries in the search of lead compounds with the potential to become enrolled in future treatments for PD. It is conceived as a friendly-to-use, easy to implement, protocol for non-specialized research labs, with the aim of reducing both time and resources, without losing accuracy. It allows to monitor aggregation kinetics with high reproducibility and low errors, permitting to identify true positives among large collections of putative candidates and, in addition, assessing specific drug features such as the mechanism of action or the concentration and time dependence of the compound activity. Despite it is true that compounds able to inhibit *in vitro* α-syn in the presence of Teflon polyballs would not necessarily behave in the same manner *in vivo* and thus that they are still far from being biologically effective therapeutics for synucleinopathies, the possibility to recruit novel labs in the search for such molecules, without the need for an initial large investment in specialized equipment, would likely boost the finding of novel candidates with the potential to halt the onset or the progression of these devastating disorders.

Materials and Methods

Expression and purification of human α-synuclein

Human α-synuclein was expressed and purified adapting a previous protocol from Volles and Lansbury [32]. Escherichia coli BL21 (DE3) cells were transformed with a pET21a plasmid (Novagen) containing the α-syn cDNA, grown in LB medium containing 100 µM/mL ampicillin and induced with 1 mM IPTG for 4 hours at an optical density at 600nm of 0.6. After cell centrifugation at 7000 x g for 10 min at 4°C, the pellet was resuspended in 20 mL Phosphate Buffered Saline (PBS) buffer, centrifuged again at 4000 x g for 20 min at 4°C and frozen at -80°C. When needed, the pellets were defrosted and resuspended in 10 mL lysis buffer (50 mM Tris pH 8, 150 mM NaCl, 1 µg/mL pepstatin, 20 µg/mL aprotinin, 1mM benzamidine, 1mM PMSF, 1mM EDTA and 0,25 mg/mL lysozyme) prior to sonication using a LabSonic®U sonicator (B. Braun Biotech International) with a power level of 40 W and a repeating duty cycle of 0.7 sec for 3 intervals of 3 min. Resultant cell extract was boiled at 95°C for 10 min and centrifuged at 20000 x g for 40 min at 4°C. To the obtained supernatant 136 µL/mL of 10% w/v streptomycin sulfate and 228 µL/mL of pure acid acetic were added and centrifuged at 4°C (20000 x g, 10 min). The resulting soluble fraction was diluted with saturated ammonium sulfate (550 g/l) 1:1 (v/v) and centrifuged at 4°C (20000 x g, 10 min). Then, the pellet was resuspended in 50 % ammonium sulfate and centrifuged at 4°C (20000 x g, 10 min). The pellet was washed with 100 mM pH 8 ammonium acetate (5 mL per culture litre) and pure EtOH 1:1 (v/v), then, the mixture was centrifuged at 4°C (20000 x g, 10 min). The pellet was resuspended in 20 mM pH 8 Tris and filtered with a 0,45 mm filter. Anion exchange column HiTrap Q HP was coupled to an ÄKTA purifier high performance liquid chromatography system in order to purify α-synuclein. Tris 20 mM pH8 and Tris 20mM pH8, NaCl 1 M were used as buffer A and buffer B respectively. After column equilibration with buffer A, sample was injected by using a Pump Direct Loading P-960 and the weak bonded proteins were washed with 5 column volumes (cv) of Buffer A. To properly isolate α-syn, a step gradient was applied as follows: i) 0-20 % buffer B, 5 cv; ii) 20-45 % buffer B,11 cv; iii) 100 % buffer B, 5 cv, obtaining pure α-syn between 25-35 % buffer B concentration. The collected peaks were dialyzed in 5 L ammonium sulfate 50 mM overnight. α-syn concentration was determined measuring the absorbance at 280 nm and using the extinction coefficient 5960 M−1cm−1. Purity was

checked using 15% SDS-PAGE and unstained Protein Standard markers from Thermo Fisher Scientific. The gel was stained with comassie brilliant blue. Identity was checked by mass spectrometry. 2 µL of protein were dialysed for 30 minutes at room temperature using 20 mL of 50 mM ammonic bicarbonate and a 0.025 µm pore membrane (Millipore). After that, MALDI-TOF was analysis was performed with a ground steel plate and 2,6-dihidroxiacetophenone acid as a matrix, in a MALDI-TOF UltrafleXtreme (Bruker Daltonics). A 1:1 sample:matrix mixture was used, adding just 1 µL of these sample to the plate. For the analysis, a lineal mode was used with an accelerated voltage of 25kv. Finally, after lyophilisation, the protein was kept at -80ºC

Quenching analysis of compounds

The different compounds were dissolved at 50 mM in pure 100% DMSO solution. In order to check the interference of the compounds with thioflavin-T (Th-T) excitation or emission, the absorption spectra for each compound at 100 µM was measured from 400 to 600 nm in a spectrophotometer Cary100.

α-synuclein aggregation and Thioflavin-T assays

Previously lyophilised α-synuclein was carefully dissolved in PBS buffer to a final concentration of 210 µM and filtered through a Millipore s 0.22-☐m filter. α-syn aggregation assay was performed in a 96 wells plate (non-treated, black plastic) containing in each well a Teflon polyball (1/8'' diameter), 40 µM thiofalvin-T, 70 µM α-synuclein, 100 µM of the tested compounds and PBS up to a final volume of 150 µL. Plates were fixed into an orbital culture shaker Max-Q 4000 Thermo Scientific to keep the incubation at 37°C, 100 rpm. Every 2 hours, the fluorescence intensity was measured using a Victor3.0 Multilabel Reader by exciting the mixtures with 430-450 filter and collecting the emission intensity with 480-510 filter (triplicates for each measurement). Each plate contained 3 α-syn controls in the absence of any compound. The averaged Th-T fluorescence obtained for these wells at the end of the experiment was normalized to 1 and the kinetic curves in the different wells re-scaled accordingly. Re-scaled curves were used to compare the controls with the effect of compounds and to ensure that the controls were reproducible between different experiments.

For the titration assay the following concentrations for all selected compounds (200, 150, 100, 75 and 25 µM) were used.

Transmission Electron Microscopy (TEM)

α-synuclein fibers from final point reaction (either in absence or presence of the final concentration inhibitors) were collected in Eppendorfs. After diluting the aggregated α-syn to a concentration of 10 μM α-synuclein, each sample was sonicated for 10 minutes. 5 μL of these samples were placed on carbon-coated copper grids and allowed to stand for 5 minutes. Then, samples were carefully dried with filter paper to remove the excess of sample. Grids were washed twice with MiliQ water by immersion and stained by incubating grids with 5 μL 2 % (w/v) uranyl acetate for 2 minutes for the negative staining. After removal of the uranyl acetate excess with filter paper, grids were let to air-dry for 10 minutes. The samples were imaged using a Jeol 1400 Transmission Electron Microscopy operating at an accelerating voltage of 120 kV. 30 fields were screened at least, to obtain representative images.

Statistical Analysis

Data were analysed by ANOVA Tukey test using SPSS software. All data are shown as means and standard error. $p < 0.05$ was considered statistically significant and indicated by ** and *** if $p < 0.005$ and $p < 0,0005$, respectively.

Work 5

A novel small molecule inhibits α-synuclein aggregation, disrupts amyloid fibrils and prevents degeneration of dopaminergic neurons

A novel small molecule inhibits α-synuclein aggregation, disrupts amyloid fibrils and prevents degeneration of dopaminergic neurons

Jordi Pujols[a,b,‡], Samuel Peña-Díaz[a,b,‡], Diana F. Lázaro[c], Francesca Peccati[d,e], Francisca Pinheiro[a,b], Danilo González[e], Anita Čarija[a,b], Susanna Navarro[a,b], María Conde-Jimenez[f], Jesús García[g], Xavier Salvatella[g,h], Javier Sancho[f], Mariona Sodupe[e,h], Tiago F. Outeiro[c,i,j], Esther Dalfó[k,l] and Salvador Ventura[a,b,*]

[a] Institut de Biotecnologia i Biomedicina. Universitat Autònoma de Barcelona, 08193-Bellaterra, Spain.

[b] Departament de Bioquímica i Biologia Molecular. Universitat Autònoma de Barcelona, 08193-Bellaterra, Spain.

[c] Deparment of Experimental Neurodegeneration, Center for Biostructural Imaging of Neurodegeneration, Center for Nanoscale Microscopy and Molecular Physiology of the Brain, University Medical Center Göttingen, Waldweg 33, 37073 Göttingen, Germany.

[d] Departament de Química, Universitat Autònoma de Barcelona , 08193 Bellaterra, Spain

[e] Sorbonne Universités, CNRS, Laboratoire de Chimie Théorique, LCT, F-75005 Paris, France

[f] Department of Biochemistry and Molecular and Cell Biology, Institute for Biocomputation and Physics of Complex Systems (BIFI), University of Zaragoza, 50018 Zaragoza, Spain

[g] Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Baldiri Reixac 10, 08028, Barcelona, Spain

[i] Max Planck Institute for Experimental Medicine, Göttingen, Germany.

[j] Institute of Neuroscience, The Medical School, Newcastle University, Framlington Place, Newcastle Upon Tyne, NE2 4HH, UK

[k] Medicine, M2 , Universitat Autònoma de Barcelona (UAB), Bellaterra Campus , Cerdanyola del Vallés, Barcelona, Spain. ORCID: 0000-0003-4677-8515

[*] Correspondence: salvador.ventura@uab.cat; Tel.: +34-93-5868956

[‡] These authors contributed equally to this work.

Abstract

Parkinson's disease (PD) is characterized by a progressive loss of dopaminergic neurons, a process that current therapeutic approaches cannot prevent. In PD, the typical pathological hallmark is the accumulation of intracellular protein inclusions, known as Lewy bodies and Lewy neurites, which are mainly composed of α-synuclein. Here, we exploited a high-throughput screening methodology to identify a small molecule (SynuClean-D) able to inhibit α-synuclein aggregation. SynuClean-D significantly reduces the *in vitro* aggregation of *wild-type* α-synuclein and the familiar A30P and H50Q variants in a substochiometric molar ratio. This compound prevents fibril propagation in protein misfolding cyclic amplification assays and decreases the number of α-synuclein inclusions in human neuroglioma cells. Computational analysis suggests that SynuClean-D can bind to cavities in mature α-synuclein fibrils and, indeed, it displays a strong fibril disaggregation activity. The treatment with SynuClean-D of two PD *Caenorhabditis elegans* models, expressing α-synuclein either in muscle or in dopaminergic neurons, significantly reduces the toxicity exerted by α-synuclein. SynuClean-D treated worms show decreased α-synuclein aggregation in muscle and a concomitant motility recovery. More importantly, this compound is able to rescue dopaminergic neurons from α-synuclein induced degeneration. Overall, SynuClean-D appears as a promising molecule for therapeutic intervention in Parkinson's Disease.

Keywords: Parkinson's Disease, α-synuclein, amyloid, aggregation inhibition, protein aggregation, dopaminergic degeneration, *C. elegans*, high-throughput screening, protein misfolding cyclic amplification, disaggregation, molecular dynamics.

Introduction

Parkinson's Disease (PD) is the second most prevalent neurodegenerative disorder after Alzheimer's Disease and is still incurable (1). PD is the most common synucleinopathy, a group of neurodegenerative disorders that includes Dementia with Lewy bodies and Multiple System Atrophy among others (2-4). Although the synucleinopathies are multifactorial disorders, the molecular events triggering the pathogenic breakthrough of the disease converge to the abnormal aggregation of α-synuclein (α-Syn), a process that starts in dopaminergic neurons (5, 6). α-Syn is an intrinsically disordered protein expressed at high levels in the brain whose function is thought to be related to vesicle trafficking (7). This *wild type* protein is the main component of cytoplasmic Lewy bodies and Lewy neurites in sporadic PD (8). In addition, dominantly inherited mutations in α-Syn, as well as multiplications of the gene encoding for α-Syn (*SNCA*), cause familial forms of PD (9).

Interfering with α-Syn aggregation has been envisioned as a promising disease-modifying approach for the treatment of PD (1). However, the disordered nature of α-Syn precludes the use of structure-based drug design and optimization for the discovery of novel molecules able to modulate α-Syn aggregation. Therefore, many efforts have focused on the analysis of large collections of chemically diverse molecules to identify lead compounds (10). Recently, we have developed an accurate and robust high-throughput screening methodology to identify α-Syn aggregation inhibitors (11). Here, we describe the properties of SynuClean-D (SC-D), a small molecule identified with this approach (Fig. 1).
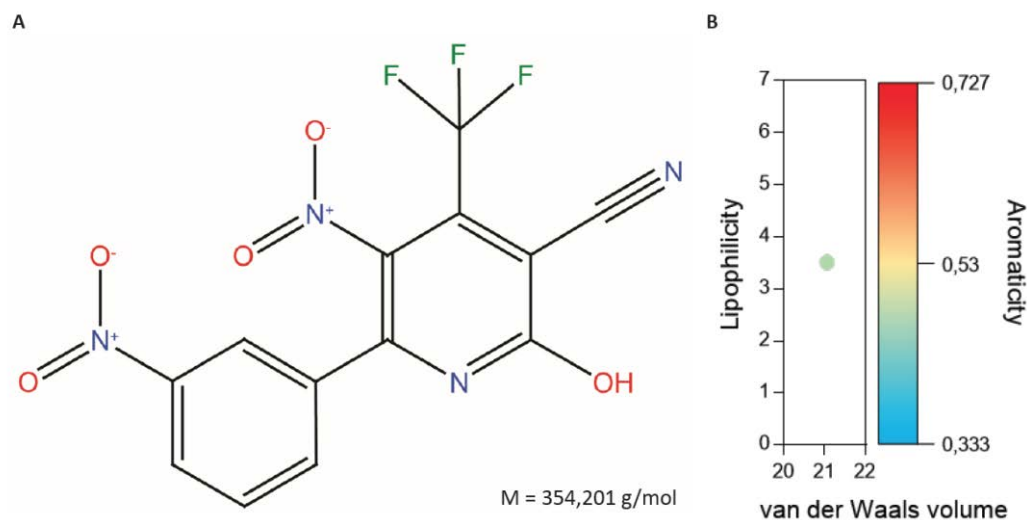
Figure 1. 2-hydroxy-5-nitro-6-(3-nitrophenyl)-4-(trifluoromethyl)nicotinonitrile or SynuClean-D (SC-D). Chemical structure (A) and properties (B) of the compound. Lipophilicity is defined as the Ghose–Crippen octanol/water coefficient (22). The volume corresponds to the sum of atomic van der Waals volumes (23). Aromaticity is defined as the ratio of the number of aromatic atoms to the total number of atoms.

We first performed a detailed *in vitro* biophysical characterization of the inhibitory and disaggregation activities of SC-D and tested its performance in human neuronal cells. Finally, we validated the effects *in vivo* in two well-established *Caenorhabditis elegans* models of PD, which express α-Syn either in muscle cells or in dopaminergic neurons. The inhibitor reduced α-Syn aggregation, improved motility and protected against neuronal degeneration.

Results

SynuClean-D inhibits α-synuclein aggregation in vitro

The formation of α-Syn amyloid fibrils can be reproduced in vitro by incubating the recombinant protein. However, fibril growth is very slow and highly variable, complicating drug screening (12, 13). We have implemented a robust high-throughput kinetic assay to screen large chemical libraries in the search for α-Syn aggregation inhibitors (10, 11). The assay uses Thioflavin-T (Th-T) as a readout of amyloid formation, completing highly reproducible reactions in 30 h. Approximately 14,400 chemically diverse compounds of the HitFinderTM Collection from Maybridge were screened with this approach.

SC-D (2-hydroxy-5-nitro-6-(3-nitrophenyl)-4-(trifluoromethyl)nicotinonitrile), a small aromatic compound, was identified as one of the molecules of potential interest (Fig. 1). Evaluation of SC-D structure according to medicinal chemistry criteria (14-17) predicted good metabolic stability, blood-brain barrier permeability and oral bioavailability required for long-term therapy. Incubation of 70 µM α-Syn with 100 µM SC-D impacted α-Syn aggregation, as monitored by Th-T fluorescence (Fig. 2A). The analysis of the aggregation curves indicated that the autocatalytic rate constant in the presence of the compound ($ka = 0.25$ h-1) was 25 % lower than in its absence ($ka = 0.33$ h-1). SC-D increases t50 by 1.5 h and reduces by 53% the amount of Th-T positive material at the end of the reaction. By measuring light scattering, we confirmed that the observed changes in Th-T fluorescence reflected an effective decrease in the levels of α-Syn aggregates, with a reduction of 48% and 58% in the scattering signal at the end of the reaction in the presence of SC-D, when exciting at 300 nm and 340 nm, respectively (Fig. 2B). Nanoparticle tracking analysis indicated that the presence of SC-D increased the number of particles < 100 nm and decreased the formation of large aggregates (150-500 nm) (Fig. S1). Finally, Transmission Electron Microscopy (TEM) images confirmed that those samples incubated with SC-D contained smaller and much less fibrils per field than untreated samples (Fig. 2C and 2D). The inhibitory activity of SC-D was dose dependent and still statistically significant at 10 µM (1:7 compound:α-Syn ratio), where it reduces the final Th-T signal in a 34% (Fig. 1E).

We further investigated whether SC-D was active against the aggregation of α-Syn variants associated with PD (1). SC-D was able to reduce by 45% and 73% the amount

160

of Th-T positive aggregates at the end of the reaction for the H50Q and A30P α-Syn familial variants, respectively (Fig. 2F).

The inhibitory activity of SC-D was also assessed using protein misfolding cyclic amplification (PMCA) (18, 19). Conceptually based on the nucleation-dependent polymerization model for prion replication, PMCA has been recently adapted to amplify α-Syn amyloid fibrils (20). The PMCA technique combines cycles of incubation at 37 °C, to grow fibrils, and sonication, to break fibrils into smaller seeds. In our conditions, a single cycle of amplification was sufficient to generate amyloid-like protease K (PK) resistant α-Syn assemblies, but the highest levels of protection were attained after four rounds (Fig. 2G). When the same experiment was performed in the presence of SC-D, we observed a substantial decrease in the amount of PK resistant material (Fig. 2H), indicating that the molecule was interfering with α-Syn template seeding amyloid formation.
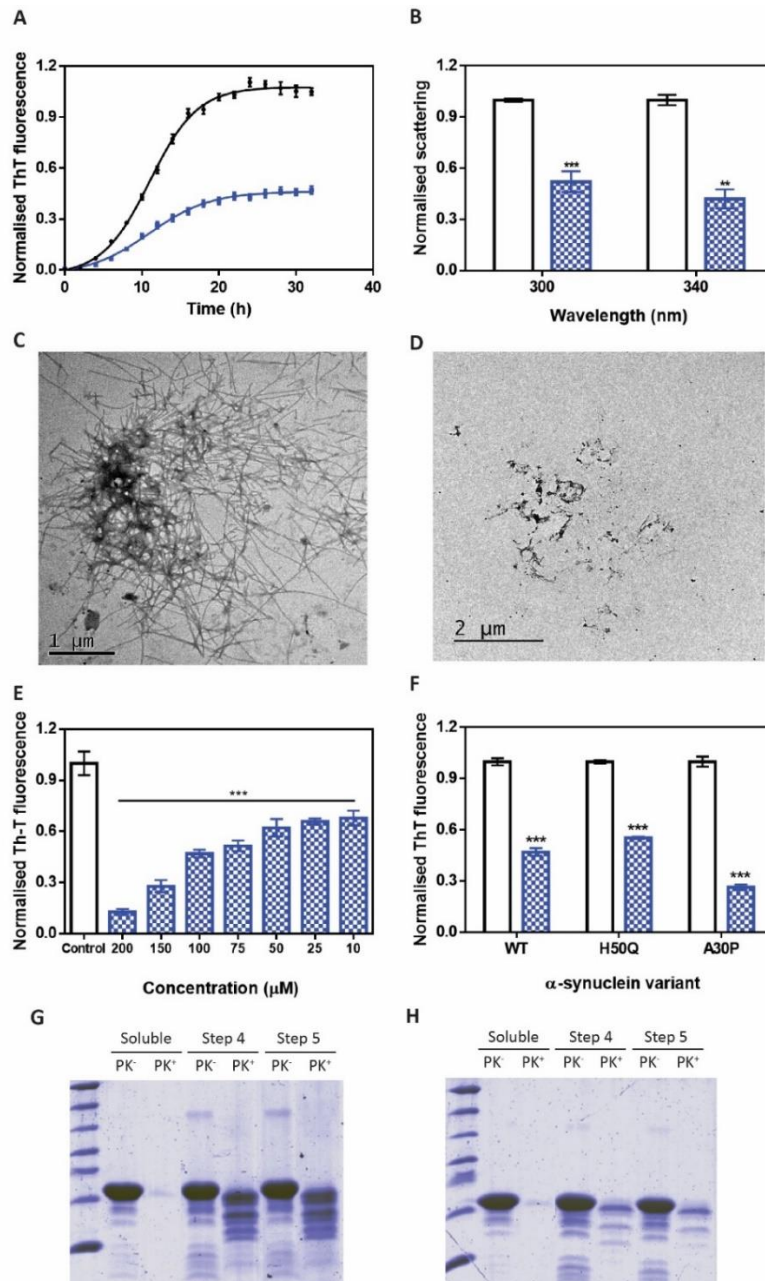
Figure 2. Effect of SynuClean-D in the aggregation of α-synuclein *in vitro*. (A) α-Syn aggregation kinetics in absence (black) and presence (blue) of SC-D followed by Th-T derived fluorescence. (B) Light scattering signal at 300 nm and 340 nm, both in absence (white) and presence (blue) of SC-D. (C and D) Representative TEM images in absence (C) and presence (D) of SC-D. (E) Inhibition of α-Syn aggregation in presence of different concentrations of SC-D. (F) H50Q and A30P α-Syn variants aggregation in absence (white) and presence (blue) of SC-D. (G and H) Bis/Tris gels of PMCA samples in absence (G) and presence (H) of SC-D, both analysed after PK digestion. For all panels: Th-T fluorescence is plotted as normalized means. Final points were obtained at 48 h. of the aggregation reaction. Error bars are represented as standard error of means values, where $p < 0.01$ and $p < 0.001$ are indicated as ** and *** respectively.

SynuClean-D disrupts preformed α-synuclein fibrils

The progress of α-Syn PMCA reactions can be also monitored by using the Th-T signal as the readout for fibril assembly (21). Consistent with PK resistance analysis, Th-T fluorescence of α-Syn increased significantly after four cycles of PMCA (Fig. 3A). Surprisingly, with SC-D, the Th-T signal not only did not increase but began to decrease after the third cycle. This suggested that SC-D might disrupt newly formed amyloid fibrils.

To address the time window in which SC-D is active, we set up aggregation reactions with constant amount of SC-D at different time intervals. As presented in Fig. 3B, the effect of SC-D on the final amount of amyloid structures was independent on whether it was added at the beginning (4 h), in the middle (12 h), at the end (18 h) of the exponential phase or when the reaction had already attained a plateau (24 h). These results again suggested an ability to disrupt/destabilize newly formed fibrils.

To confirm the fibril disrupting activity of SC-D, four days mature α-Syn fibrils were incubated for 24 h in the absence or presence of the compound. Incubation with SC-D promoted a 42.63% reduction in Th-T fluorescence emission (Fig. 3C). Moreover, light scattering measurements indicated a reduction in the amount of detectable aggregates of 29% and 39% at 300 and 340 nm, respectively (Fig. 3D). Consistently, TEM images illustrated how four-days-incubated α-Syn tended to form big fibrillary clusters (Fig. 3E) that became completely disrupted in the presence of SC-D (Fig. 3F).
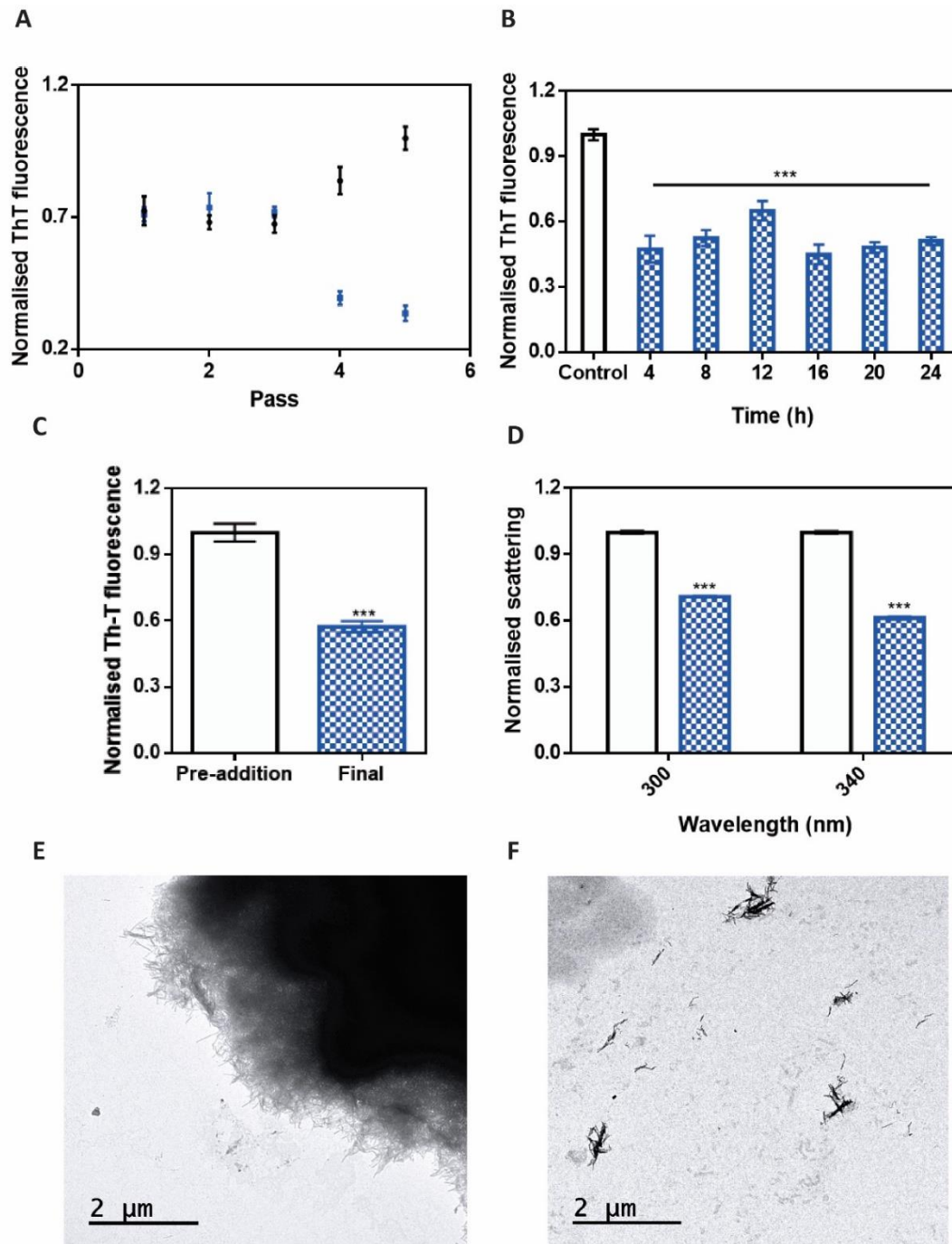
Figure 3. Disaggregational capacity of SynuClean-D. (A) Th-T fluorescence of the different PMCA passes of both, treated (blue) and untreated (black) samples with SC-D. (B) Aggregation kinetics of α-Syn after the addition of SC-D at different time points. (C and D) Th-T derived fluorescence (C) and light scattering (D) assays before and after the addition of SC-D to pre-formed α-Syn fibrils. (E and F) Representative TEM images in absence (E) and presence (F) of SC-D. For all panels: Th-T fluorescence is plotted as normalized means. Error bars are represented as standard error of means values, where $p < 0.001$ is indicated as ***.

α-synuclein fibrils can accommodate SynuClean-D

To assess if SC-D can bind monomeric and soluble α-Syn, the recombinant protein was isotopically labelled and NMR 1H-15N-HSQC spectra of 70 μM 15N-α-Syn were recorded in the absence and presence of SC-D. We did not detect any perturbations in chemical shifts or peak intensities respect the original α-Syn spectrum in the presence of 100 μM of the molecule (Fig. S3), indicating that SC-D does not bind α-Syn monomers.

Induced fit docking simulations of α-Syn-SC-D revealed four major poses for the interaction of the SC-D with α-syn fibrils (22), two internal, with SC-D fully inserted in the fibril, (poses 1 and 2) and two external, with SC-D partially exposed (poses 3 and 4) (Fig. 4). In the internal poses, the ligand is sandwiched between two parallel β-sheets of the greek-key motif and interacts with the side chains of ALA53, VAL55, THR59, GLU61, THR72 and GLY73. The only difference between pose 1 and 2 lies in the orientation of the compound in the binding pocket. PELE (23-26) interaction energies are stronger for the internal poses, where SC-D binds essentially through dispersion interactions into a solvent-excluded cavity, than for external ones, where SC-D inserts in a surface groove of the fibril. In light of this, we predict that SC-D will bind into the core of α-Syn deposits, where its flexibility and interference with the native pattern of stabilizing contacts may initiate the disaggregation process.

MM/GBSA calculations (Table S1 of ESI) (27-30) also show that the internal binding pose 1 (Fig. 4A and 4B) is the one that exhibits a larger binding energy with the fibril, the computed ΔGbind being -18.4 ± 4.1 kcal mol-1. Note that this binding energy does not include the positive (destabilizing) entropic term, which, according to previous simulations on similar compounds interacting with amyloid fibers (31), is expected to be in the 12-18 kcal mol-1 range. Decomposition of the binding energy (Table S1) reveals that its main contribution is the van der Waals term, representing roughly 80% of the binding interaction.

Table S1. Interaction energies values involved in different predicted binding sites.

Binding pose	1	2	3	4
ΔG_{gas}	-53.7 (5.4)	-47.1 (9.5)	-40.6 (8.3)	-52.3 (16.6)
vdW	-36.2 (4.5)	-37.1 (4.7)	-25.0 (3.9)	-31.0 (5.0)
E_{El}	-17.5 (4.7)	-10.0 (10.3)	-15.6 (7.0)	-21.3 (13.3)
$D\Delta G_{solv}$	35.3 (4.6)	32.5 (8.1)	29.2 (6.8)	36.2 (12.2)
ΔG_{bind}	-18.4 (4.1)	-14.6 (3.4)	-11.4 (3.1)	-16.1 (5.7)

* MM/GBSA gas phase binding energy ΔG_{gas}, van der Waals (vdW) and electrostatic contributions (E_{El}), solvation free energy (DDG_{solv}), and binding energy in solution (ΔG_{bind}) for the two internal (1 and 2) and two external (3 and 4) poses of SC-D interacting with the α-Syn fibril model. All energies are in kcal mol^{-1}. Standard deviation is showed in parentheses.

This is not surprising given the nature of SC-D, a planar aromatic molecule. Furthermore, plots of the reduced density gradient versus the density (Fig. 4C) provide information on the nature of the non-covalent interactions present in the system (32). That is, peaks in the negative region of the x-axis are indicative of attractive interactions, while peaks in the positive regions are indicative of repulsive (non-bonded) contacts. The region around zero of the x-axis brings information about the weakest non-covalent interactions present, which correspond to van der Waals contacts. Though weak, these interactions are present in a large number and involve the whole body of the molecule, resulting in the most important contribution to the binding energy. Their spatial extension can be appreciated in Fig. 4D, where these contacts are represented in green. For pose 1, with the largest binding affinity (Fig. 4A), the NCI plot shows that besides van der Waals, a H-bond contact is responsible for the binding of SC-D (Fig. 4C).
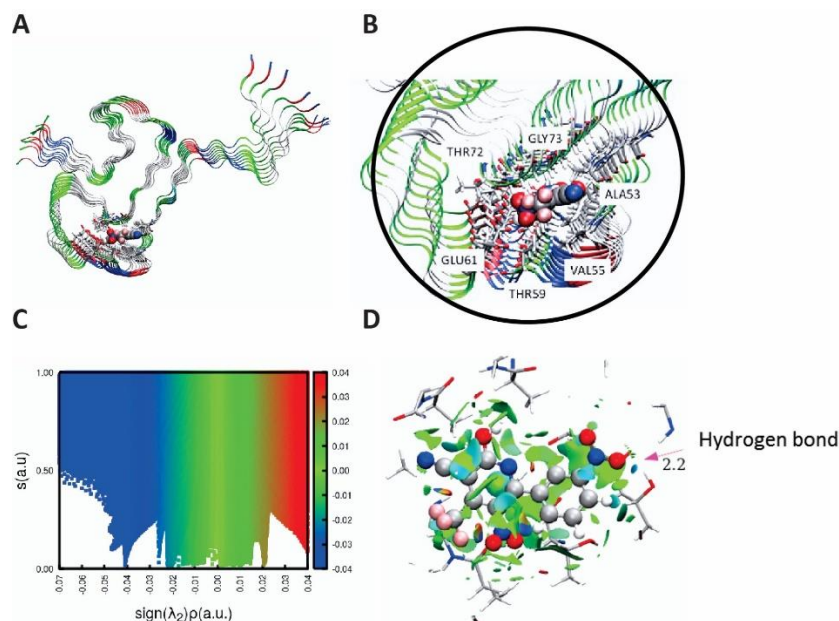
Figure 4. Characterisation of SynuClean-D-fibril interaction. (A and B) Most stable binding pose of SC- D on the α-syn fibril model. (C and D) Non-covalent interaction plot (C) and the representation of the non-covalent interactions involving SC-D in preferred binding pose at the PELE geometry (D). Distances in Å.

SynuClean-D inhibits the formation of intracellular α-synuclein aggregates in cultured cells

We tested the potential toxicity of SC-D for human neuroglioma cells (H4). After confirming that the molecule was innocuous at concentrations as high as 50 μM (Fig. 5A), we used a well-established cell model that enabled us to assess α-Syn inclusion formation. H4 cells were transiently transfected with α-Syn (synT) and synphilin-1, which results in the formation of LB-like inclusions, as we and others previously described (9). The formation of α-Syn inclusions was assessed 48 h after transfection by immunofluorescence (Fig. 5D). Upon the treatment with 1 μM and 10 μM SC-D we observed a significant increase in the number of transfected cells devoid of α-Syn inclusions (SC-D 1 μM: 42.4 ± 1.0 %, SC-D 10 μM: 49.5 ± 4.5 %) relative to untreated samples (ctrl: 28.7 ± 2.0 %) (Fig. 5B). SC-D treatment also promoted a decrease in the number of transfected cells displaying more than 10 aggregates (SC-D 1 μM: 25.4 ± 5.6 %, SC-D 10 μM: 25.1 ± 8.2 %) relative to control cells (ctrl: 39.2 ± 6.7 %) (Fig. 5C and D).
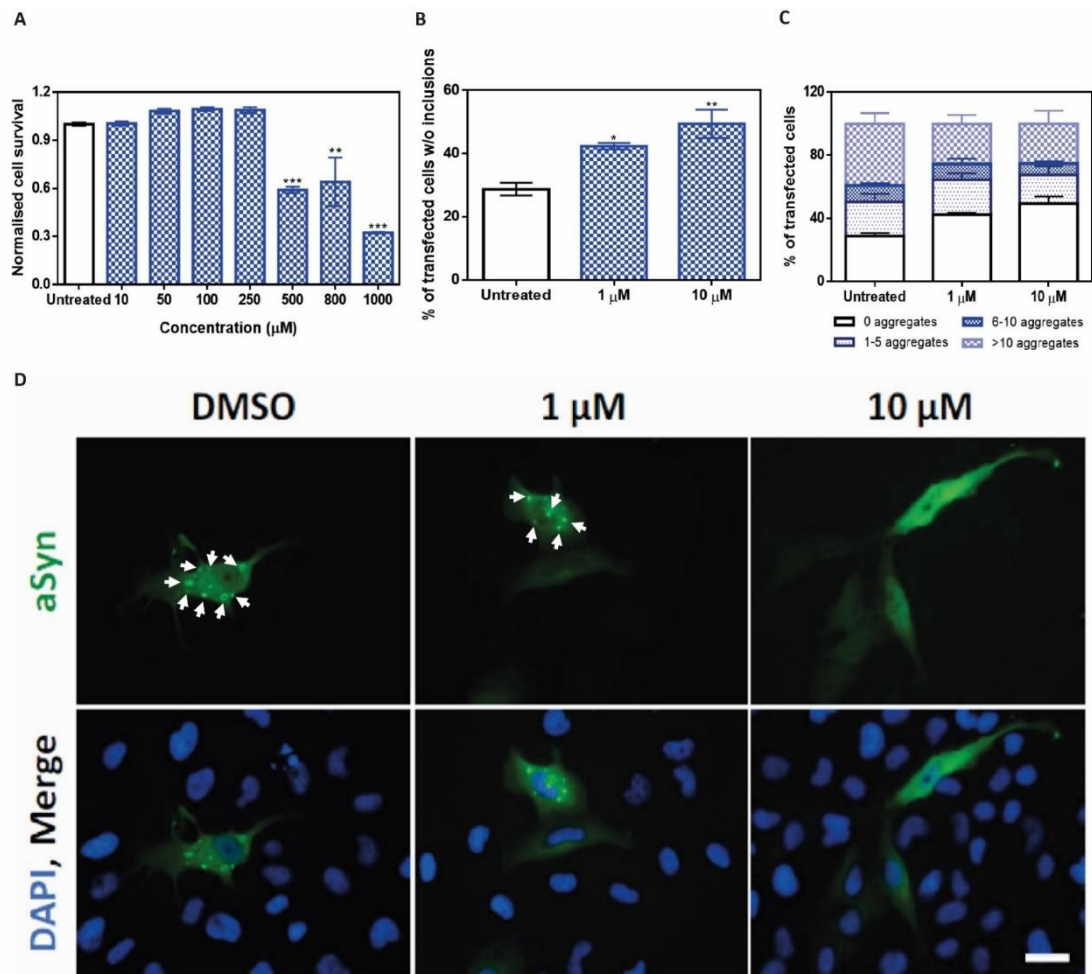
Figure 5. Inhibition of α-synuclein aggregates formation in cultured cells. (A) Human neuroglioma cells (H4) survival when incubated with different concentrations of compound (blue) and without (white) the compound. (B) Reduction of α-Syn inclusions formation in human cultured cells in presence of different SC-D concentrations. (C) Inclusion quantification of, at least 50-100 cells per condition and classification according to the number of inclusions per cell. (D) Representative epifluorescent images from cells treated with SC-D. n = 3. Scale bar: 30 μm. For all panels: Error bars are represented as standard error of means, where $p < 0.05$, $p < 0.01$ and $p < 0.001$ are indicated as *, ** and *** respectively.

SynuClean-D inhibits α-synuclein aggregation in a C. elegans model of PD

Next, we tested SC-D in a living system. We used a well-studied nematode model of PD, in which human α-Syn fused to the yellow fluorescent protein (YFP) is under control of the muscular unc-54 promoter, transgene pkIs2386 [Punc-54::α-SYN::YFP] (33). Muscle expression has been used successfully to model protein-misfolding diseases and to identify modifier genes without considering neuronal effects (33-35). To determine the effects of SC-D in α-Syn accumulations, animals at fourth larval stage (L4) (36) were incubated with and without the compound, to analyze the inhibitor efficiency in aged worms at 7 days post-hatching, which mimic aged human PD (37). We avoided a

168

compound burst at L1 (38) because this treatment mimics a preventive rather than a disease-modifying intervention. Quantification of the number of α-Syn aggregates (39) revealed that in treated animals the number of visible α-Syn aggregates was decreased by 13.1 units in comparison with non-treated worms (23.0 ± 2.7 vs. 36.1 ± 1.7 respectively) (Fig. 6A and D).

Major defects in regular bending have been used to identify modifiers of protein aggregation (33, 40). Indeed, C. elegans thrashing can be measured in liquid media by counting the number of body bends per unit of time (41). By using this method, we confirmed an improved motility in SC-D treated animals in comparison with non-treated worms (Fig. 6B and E). We observed a decrease on bending of 72.2 % in YFP::α-SYN animals when compared to N2 wild-type strain (18.1 ± 2.5 vs 90.3 ± 6.7 respectively). This motility decrease was reverted in YFP::α-SYN animals treated with SC-D. In these animals, the average bending increased by a 2.7-fold when compared to non-treated animals (18.1 ± 2.5 vs 49.1 ± 4.4).

SynuClean-D prevents degeneration of dopaminergic neurons in a C. elegans model of PD

PD is characterized by the degeneration of dopaminergic (DA) neurons. There exist four pairs of DA neurons in C. elegans hermaphrodites, three of them (CEPD, CEPV and ADE) located in the anterior part, and one pair, the PDE, located in the posterior part of the nematode (42). To investigate the neuroprotective role of SC-D in dopaminergic cell death, we sought to analyze its effect in a C. elegans model of PD in which DA neurons undergo age-dependent neurodegeneration (43). In this model, animals express both α-Syn and GFP in DA neurons, according to the simplified genotype Pdat-1::GFP; Pdat-1::α-SYN (the full genotype is detailed in the methodology section). This strain has been successfully used for the investigation of human PD related mechanisms (37). When human α-Syn is expressed in these animal's DA neurons, the six DA neurons within the anterior region of the worm display progressive degenerative characteristics (44). To model the aging contribution to PD, we determined the inhibitor capacity of SC-D in dealing with DA cell death induced by human α-Syn at 9 days (L4 + 7) post hatching. Cell bodies and neuronal processes were assessed to determine whether these structures displayed morphology changes. At 9 days post hatching, only 13.0 ± 3.0 % of non-treated animals showed 6 wild type DA neurons. In contrast, 41.7 ± 1.2 % of treated animals

showed the six intact DA neurons (Fig. 6C and F), which evidenced the ability of SC-D to protect against α-Syn induced DA neurons degeneration.
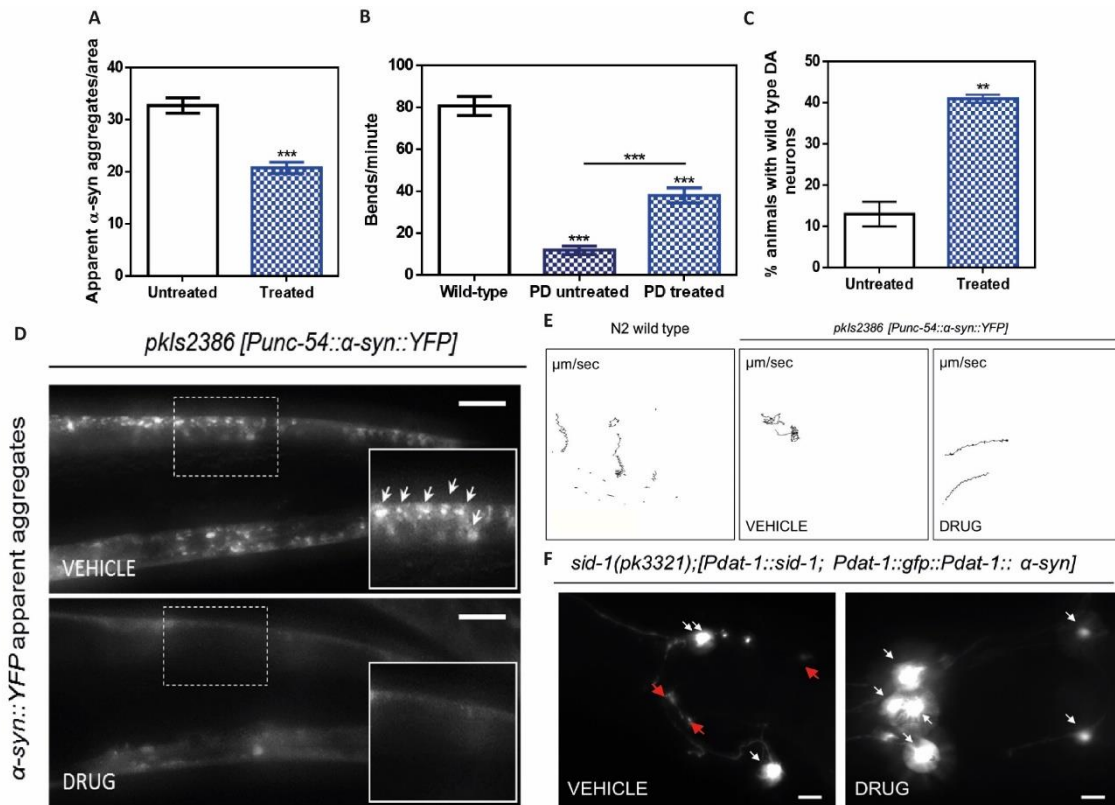

Figure 6: Inhibition effect of the compound in the formation of α-synuclein inclusions and protection from the α-synuclein induced dopaminergic cell death in *C. elegans* models of PD. (A) Quantification of α-Syn muscle inclusions per area of NL5901 worms in absence (white) and presence (blue) of SC-D. (B) Worm thrashing representation as number of bends per minute of N2 wild type and NL5901 worms treated without (white) and with SC-D (blue). (C) Percentage of UA196 worms that maintain a complete set of dopaminergic neurons (4 pairs of them) after a treatment without (white) and with SC-D (blue) for 7 days after L4 stage. (D) Aggregates Representative images of α-Syn muscle aggregates by epifluorescence microscopy of NL5901 worms treated without (top panel, vehicle) and with SC-D (bottom panel, drug). Bar=10 μm. Between 40-50 animals were analyzed per condition. (E) Path representation of the mobility of N2 wild type (left panel, vehicle) and NL5901 worms grown without (middle panel, vehicle) and with SC-D (right panel, drug). (F) Representative worms expressing GFP-α-Syn specifically in DA neurons without (left panel, vehicle) and with SC-D (right panel) for 7 days after L4. Healthy neurons are labeled with white arrows whereas neurodegenerated or missing neurons are labeled with red arrows. Bar = 30 μm. Between 40-50 animals were analyzed per condition in each experiment. For all panels: data are shown as means and error bars as standard error of mean, where $p < 0.01$ and $p < 0.001$ are indicated as ** and *** respectively.

Discussion

α-Syn aggregation plays a major pathophysiological role in the development of PD. Since its discovery and subsequent identification as the most abundant protein in Lewy bodies (3), α-Syn was shown to be important for a number of cellular processes (4, 7). Therapeutic strategies targeting the aggregation of α-Syn thus hold the promise to result in disease-modification and mitigate pathology in PD (45).

The screening of large chemical libraries has rendered promising molecules that inhibit the progression of PD by targeting the aggregation of α-Syn, like BIOD303 (46) or anle138b (14). Here, by exploiting a previously developed high-throughput methodology, we identified SC-D as a compound able to inhibit > 50% of α-Syn amyloid formation *in vitro* when employed in a 0.7:1 (protein:SC-D) ratio. The ability of SC-D to inhibit α-Syn aggregation, was confirmed by light scattering and TEM assays. SC-D was also active against the aggregation of familial forms of α-Syn.

The activity of SC-D was concentration dependent and still evident at a substoichiometric 7:1 protein:compound ratio. This already suggested that, unlike other compounds, SC-D does not bind to monomeric α-Syn, supported by NMR experiments. From a therapeutic perspective, this is an important advantage, as SC-D is not expected to interfere with the physiological functions of the soluble protein.

PMCA experiments indicated that SC-D might disrupt α-Syn amyloid assemblies, a property that was confirmed by the ability of SC-D to reduce the amount of formed amyloid fibrils, almost independently of the stage of the aggregation reaction at which it was added. Indeed, SC-D is very effective disassembling clusters of aged α-Syn amyloid fibrils, conceptually similar to the α-Syn amyloid inclusions recurrently observed in the dopaminergic neurons of PD patients. This is important because the disassembly of preformed amyloid fibrils has been traditionally challenging and very few molecules have been reported to break down amyloid fibrils (47) because of their high stability. Computational analysis suggest that the disrupting activity of SC-D is mediated by its binding to an inner cavity in α-Syn fibrils. The disassembly of large fibrils has been seen traditionally as a risky strategy for the amelioration of aggregation-linked diseases, because it may increase the population of smaller toxic species (12). However, our *in vivo* experiments demonstrate that this is not the case for SC-D.

The ability of SC-D to target preformed fibrils might have important implications for the prion-like pathological spreading of α-Syn aggregates in the brain (5). By disentangling transmissible fibrillary assemblies, SC-D might reduce templated seeding and thus aggregates-catalysed conversion of soluble α-Syn molecules into their insoluble forms. It has been suggested, that to attain a sustainable spreading and prevent dilution of aggregates as they propagate from cell to cell, a process of aggregate amplification is required, in addition to templated seeding (48). PMCA emulates these particular conditions *in vitro*. The potency of SC-D in blocking PMCA induced amplification of α-Syn fibrils is thus promising.

SC-D displayed low toxicity for neuronal cells and had drug-like chemical properties that anticipated its cellular permeability. Indeed, SC-D treatment at concentrations as low as 1 μM significantly reduced the number of α-Syn inclusions and the number of inclusions per cell. These data prompted us to assess the effects of SC-D treatment in *in vivo* models of α-Syn aggregation. First, we selected a well-validated *C. elegans* model of PD that expresses human α-Syn in muscular cells (33). When worms were treated with SC-D at pre-adult stages, we observed an important decrease in the number of α-Syn inclusions and a significant recovery of motility in adult PD worms. However, independently if a compound targets the early stages of aggregation or it disrupts mature protein inclusions, or both, the final aim of a PD oriented therapy is not to interfere with α-Syn aggregation *per se*, but instead to prevent the neuronal degeneration associated to this phenomenon. Is in this therapeutic context, where SC-D stands out, since it is able to increase by more than three-fold the number of animals with intact DA neurons in a *C. elegans* model in which the expression of human α-Syn is directly connected to dopaminergic degeneration (43).

Materials and Methods

Protein Purification

cDNAs corresponding to WT α-Syn and A30P and H50Q mutants were cloned in a pET28(a) plasmid and transformed into an *E. coli BL21 DE3* strain. WT α-Syn and its variants were expressed and purified as described previously (1). Samples were lyophilised and kept at -80ºC until assayed.

Metabolic stability in human liver microsomes

SC-D at 0.1 μM final concentration was pre-incubated with pooled human liver microsomes in phosphate buffer (pH 7.4) for 5 min in a 37 ºC shaking water-bath (final microsomal protein concentration: 0.1 mg/mL) in duplicate. The reaction was initiated by adding NADPH-generating system and incubated for 0, 15, 30, 45, and 60 min. The reaction was stopped by transferring the incubation mixture to acetonitrile/methanol (1:1, v/v). Samples were then vortexed and centrifuged at 4 ºC. Supernatants were used for UPLC-MS/MS analysis on an SCIEX 5500 triple-quadrupole MS *(SCIEX, Framingham, Massachusetts, USA)*. Two reference compounds were tested in parallel: propranolol, which is relatively stable, and verapamil, which is readily metabolized in human liver microsomes.

In vitro aggregation of α-synuclein

Lyophilized α-Syn was dissolved in sterile PBS 1X to a final concentration of 210 μM and filtered through 22 μm membranes to remove α-Syn small protein aggregates. The aggregation reaction was carried out in a 96-well sealed plate. Each well contained 70 μM α-Syn (WT, A30P or H50Q), 40 μM Th-T in PBS 1X, a 1/8" diameter teflon polyball (*Polysciences Europe GmbH, Eppelheim, Germany*) and the absence or presence of compound at 100 μM in a total volume of 150 μL. The plate was fixed in an orbital culture shaker Max-Q 4000 (*ThermoScientific, Waltham, Massachusetts, USA*) and agitated at 100 rpm and 37ºC. Th-T fluorescence was measured every 2 h by exciting the samples through a 430–450 nm filter and collecting the emission signal with a 480–510 filter using a Victor3.0 Multilabel Reader (*PerkinElmer, Waltham, Massachusetts, USA*). A minimum of three plate replicates were made for each condition and also measurement triplicates were taken for each time point. After compiling the fluoresce signals for each

condition and measurement, means and standard error of mean (SEM) were used to fit the aggregation kinetics with equation (1),

Equation (1)
$$\propto = 1 - \frac{1}{k_b(e^{k_a t} - 1) + 1}$$

where k_b and k_a indicate the homogeneous nucleation rate constant and the secondary rate constant, accounting for fibril elongation and secondary nucleation, respectively (2).

For the analysis of the disaggregation capacity of SC-D, 70 µM α-Syn was incubated as previously described for four days and the Th-T signal measured. Then, 3.75 µL of DMSO or SC-D 4 mM in DMSO were added to attain a 100 µM compound concentration in the assay. The plate was incubated for an additional 24 h and Th-T derived fluorescence signals were measured again.

Transmission Electron Microscopy (TEM)

α-Syn samples were diluted 1:10 in PBS 1X, sonicated for 5 minutes and 5 µL of the resulting mixture immediately placed on a carbon-coated copper grid. After 5 min, samples were carefully dried with a piece of filter paper to remove the excess of liquid and washed with MiliQ water twice. Then, 5 µL of a solution of 2% (w/v) uranyl acetate was placed on the top of the grid for 2 min. Uranyl acetate excess was removed with filter paper. Finally, grids were left to air-dry for 10 min. Images were obtained using a Transmission Electron Microscopy Jeol 1400 (*Peabody, Massachusetts, USA*) operating at an accelerating voltage of 120 kV. A minimum of 30 fields were screened for each sample to obtain representative images.

Light scattering

Total aggregate formation was measured by light scattering adding 120 µL of pre-aggregated α-Syn into a quartz cuvette. Samples were previously resuspended by carefully vortexing and pipetting and then excited at 300 and 340 to collect and 90° scattering collected between 260 to 400 nm in a Cary Eclipse Fluorescence Spectrophotometer (*Agilent, Santa Clara, California, USA*).

Nanoparticle tracking analysis

End point α-Syn aggregation reactions were collected and diluted 1:100 with MiliQ water to a final volume of 1 mL. The sample was measured by using a Nanosight NS3000 (*Malvern Instruments Ltd, Malvern, United Kingdom*) recording the light scattering and

trajectory of any particle in solution for 1 minute. All samples were measured in triplicate and analysed with NTA3.1 software (*Malvern Instruments Ltd, Malvern, United Kingdom*).

Protein Misfolding Cyclic Amplification (PMCA)

Lyophilised α-Syn was dissolved in Conversion Buffer (PBS 1X, 1% Triton X-100, 150 mM NaCl), supplemented with Complete Protease Inhibitor Mixture (*Roche Applied Science, Penzberg, Germany*), to a final concentration of 90 µM as previously described. Then, 60 µL of the α-Syn solution was placed in 200-µL PCR tubes containing 1.0 mm silica beads (*Biospec Products, Bartlesville, OK, USA*) and the mixture was subjected to 24-hour cycles of 30 s sonication and 30 min incubation at 37ºC, using a Misonix 4000 sonicator setted at 70% power. Every 24 h, 1 µL of PMCA-incubated sample was transferred to a fresh soluble α-Syn samples, repeated for 5 days. Treated samples were prepared by adding SC-D into 200-µL tubes to a final concentration of 128 µM, in order to maintain the 0.7:1 α-Syn:SC-D ratio used in original kinetic assays. Control samples were prepared with the corresponding concentration of DMSO (0.26%). All reactions were performed in triplicate.

After each cycle of PMCA, 10 µL of α-Syn aggregates were diluted to a final volume of 100 µL with PBS 1X and 40 µM Th-T. Th-T fluorescence was recorded by exciting at 445 nm and collecting the emission fluorescence between 460 to 600 nm in a Cary Eclipse Fluorescence Spectrophotometer (*Agilent, Santa Clara, California, USA*).

Proteinase K Digestion

18 µL of α-Syn aggregates coming from each cycle of PMCA were mixed with 6 µL of Proteinase K (5 µg/mL final concentration). After an incubation of 30 minutes at 37ºC, 8 µL of loading buffer containing 1% β-mercaptoethanol were added and the enzyme inactivated by an incubation for 10 min at 95ºC in a Thermocell Cooling&Heating Block (*BIOER, Hangzhou, Zhejiang Province, China*). Afterwards, 7 µL of each sample were loaded into a Tricine-SDS-PAGE gel together with unstained Protein Standard markers (*ThermoFisher Scientific, Waltham, Massachusetts, USA*). Gels were stained with Blue safe.

Nuclear Magnetic Resonance

WT ^{15}N-labeled α-Syn was expressed in *E. coli BL21 DE3* strain. Cells were previously grown in LB medium until an OD_{600} of 0.6 was reached. The cultures were centrifuged at 3000 rpm for 15 min and cells resuspended in minimal medium; 750 mL of miliQ water containing 100 μL $CaCl_2$ 1M, 2 mL $MgSO_4$ 2 M, 20 mL glucose 20%, 100 mL vitamins 100x (*Sigma-Aldrich, Darmstadt, Germany*), 200 mL salts M9 and 1 g [15]NH_4 (*Cambridge Isotope Laboratories, Inc., Tewksbury, Massachusetts, USA*). After 1 h of incubation at 37°C, protein expression was induced with 1 mM IPTG. Protein was purified as previously described (1).

^{1}H-^{15}N HSQC spectra were obtained at 20 °C on a Bruker 600 MHz NMR spectrometer equipped with a cryoprobe in a sample containing 70 μM ^{15}N-labeled α-Syn, PBS buffer (pH 7.4), 2.5% d6-DMSO and 10% D_2O in the absence or in the presence of 100 μM SC-D.

Characterisation of SynuClean-D – Fibril interaction

α-synuclein fibril model

To predict the binding energy of SC-D on human α-Syn fibrils, a model composed of 10 strands was built based on PDB structure 2N0A (Fig. S8A) (3). The central part of the structure is characterized by parallel, in-register β-sheets with a greek-key topology, while the terminals are flexible loops. The full-length α-Syn strand comprises 140 residues; however, SC-D, with its planar aromatic structure, will interact with the ordered β portion of the fibril rather than with the unstructured flexible loops. For this reason, a reduced model of α-Syn fibril was constructed including only residues 34 to 99 of PDB 2N0A (Fig. S8B). This reduced model, whose validity as a model for the full fibril has been tested with molecular dynamics simulations, has been used in all calculations.

To justify the validity of this approximation, 10 ns molecular dynamics simulations were performed and the RMSD of the backbone was computed for residues 34 to 99 on both the full-length α-synuclein fibril model (Fig. S8A) and the reduced model (Fig. S8B). Additionally, for the reduced model, a 1 ns and a 100 ns simulations were performed, to assess the stability of the fibrillar structure as a function of the simulation time. Results are reported in Fig. S9, and indicate that along a 10 ns simulation the RMSD evolution is identical for the full and the reduced model, validating our approximation. As far as the stability of the fibril as a function of time is concerned, our simulations revealed that the model of α-synuclein fibril is only partially stable, as along the 100 ns simulation the high

fluctuation of RMSD indicates that important structural deformations take place. This instability is likely due to the small size of the model, which only comprises 10 strands and is hardly representative, in terms of network of weak interactions, of a fully grown α-synuclein fibril. Within the 10 ns simulation, however, the RMSD fluctuation is lower. This observation determined the maximum time of production simulation, which was set to 3 ns just to avoid unphysical structural deformations.

Binding site prediction

To predict the interaction of SC-D with the model of α-Syn fibril we used the Protein Energy Landscape Exploration (PELE) methodology (4, 5). PELE combines a Monte Carlo stochastic approach with protein structure prediction algorithms, and is able, among several options, to predict the binding poses of a small ligand on a protein without any beforehand crystallographic information. The general workflow of PELE is composed of three steps: i) a localized perturbation (random translation and rotation of the ligand, but it may also involve the backbone of the protein); ii) a side chain sampling based on a rotamer library; iii) a minimization, involving the full system. The OPLS2005 force field is employed and solvation effects are accounted for with the Generalized Born method (6, 7). Owing to the stochastic nature of PELE, several independents runs need to be performed on the same protein-ligand pair to ensure a full exploration of the conformational space. We ran 150 independent trajectories for the α-Syn/compound pair, providing a satisfactory sampling.

Molecular dynamics simulations indicate that despite the fact that SC-D is fully inserted between the parallel β-sheets, it is still in contact with water molecules along the whole simulation. This is not surprising owing to the wealth of polar group that are attached to the aromatic ring, and again corroborates the hypothesis that the main interaction of SC-D with the fibril is of van der Waals type. These interactions are represented as green patches (Fig. S5).

Binding energy calculation

We evaluated the binding energy of the poses predicted by PELE with the MM/GBSA method as implemented in the Amber16 suite (8, 9). In the MM/GBSA method, the binding energy is estimated based on a collection of frames, which were generated with molecular dynamics simulations run with the ff14SB force field, with the ligand topology generated using the General Amber Force Field and atomic partial charges fitted on the

electrostatic potential computed at the B3LYP/6-31+G(d,p) level of theory with the Gaussian 09 package (10, 11). The PELE structures are neutralized with the appropriate number of counterions (Na^+/Cl^-) and water molecules are added up to a minimum distance of 8 Å from the protein. The protocol for the molecular dynamics simulations is as follows: i) 2000 steps of minimization; ii) a 200 ps equilibration run in the NVT ensemble rising the temperature from 0 to 100 K, with a 4.0 kcal mol^{-1} constraint on the backbone, to avoid strong deformations; iii) a 2 ns run in the NPT ensemble, with the temperature rising from 100 to 300 K in the first half, and kept constant to 300 K in the second half, with a 2.0 kcal mol^{-1} constraint on the backbone, again to avoid strong deformations; iv) 5 independent 3 ns production runs in the NPT ensemble with the temperature kept constant to 300 K, with no constraints. A Langevin thermostat and a Monte Carlo barostat were employed. 20 geometries were evenly sampled from the last 2 ns of each trajectory of the production run, and the resulting 100 structures were used for binding free energy calculations neglecting the entropy term. The free energy values were computed on the ligand/fibril complex trajectory. An ionic strength of 100 mM, which is compatible with that of a biological buffer, was employed, and the igb5 generalized Born model of Amber16 was used. In this model, the Born radii are re-scaled to obtain a better agreement with the electrostatic component of solvation energy calculated by the Poisson-Boltzmann treatment in proteins (12).

Non-covalent interactions prediction

Non-covalent interactions were studied with the NCIPLOT program according to the methodology proposed by Yang and co-workers (13). The reduced density gradient s(r) (Eq. 2) is plotted against the electron density ρ multiplied by the sign of the second eigenvalue of the density Hessian. Peaks in the negative region of the x-axis in Fig. 4C are indicative of attractive interactions, while peaks in the positive regions are indicative of repulsive (non-bonded) contacts.

$$\text{Equation (2)} \qquad s(\boldsymbol{r}) = \frac{1}{2(3\pi^2)^{1/3}} \frac{|\nabla\rho|}{\rho^{4/3}}$$

Cytotoxicity assay

For human neuroglioma cells (H4), 40,000 cells were plated in a 24-well plate. After 24h, cells were treated with different concentrations of the compound or with vehicle (DMSO). Cells were the fixed after 24h with 4% PFA for 10 min, permeabilized with 0,1% triton

X-100 for 20 min, and stained with DAPI (1:5000) for 10 min. With the Olympus IX81-ZDC microscope system, 36 fields were randomly taken with 10x objective, in 6 independent experiments. The number of cells was counted using the Olympus Scan^R Image Analysis Software and the results analyzed using Graphpad Prism software (GraphPad Software, La Jolla, California, USA).

For human neuroblastoma cells (SH-SY5Y), 4000 cells/well were seeded into 96 well-plates and cultured in DMEM medium supplemented with 10% FBS. Then, SC-D was added at a range from 10 μM to 1 mM into each well. Treated and untreated cells were incubated at 37 ºC for 72 h. Then, 10 μL of PrestoBlue® reagent (*ThermoFisher Scientific, Waltham, Massachusetts, USA*) was added and, after an incubation for 10 minutes at 37 ºC, the fluorescence emission was recorded at 590 nm by exciting at 560 nm.

In-cell aggregation of endogenous α-synuclein

Cell Culture

Human neuroglioma cells (H4) were maintained in Opti-MEM I Reduced Serum Medium (*Life Technologies-Gibco, Thermo Fisher Scientific, Waltham, Massachusetts, USA*) supplemented with 10% fetal bovine serum Gold (FBS) (*PAA, Cölbe, Germany*) and 1% Penicillin-Streptomycin (*PAN, Aidenbach, Germany*). Cells were grown at 37 °C with 5% of CO_2 atmosphere.

Cell Transfection

Twenty-four hours prior to transfection, H4 cells were plated in 12-well plates (*Costar, Corning, New York*). The cells were transiently transfected according to the calcium phosphate method. Equal amounts of plasmids encoding a C-terminally modified α-Syn (SynT construct) were diluted in H_2O and 2.5 M $CaCl_2$. The mixture was then added dropwise and vigorously mixed into 2xBES-buffered saline solution containing phosphate ions (50 mM BES, 280 mM NaCl, 1.5 mM $Na_2HPO_4xH_2O$, pH 6.98). Fresh media was added to the cells during the 20 min of incubation. 16 h after transfection, cells were treated with different concentrations of the compound (1 or 10 μM). DMSO was used as the vehicle. Finally, at 24 h the cells were fixed and subjected to immunocytochemistry to examine α-Syn inclusions formation.

Immunocytochemistry

After the treatment, cells were fixed with 4% paraformaldehyde for 10 min at room temperature followed by 20 min of permeabilization with 0.1% Triton X-100 (*Sigma-Aldrich, Darmstadt, Germany*) at room temperature. The cells were blocked in 1.5% normal goat serum (PAA)/PBS for 1 h, and then incubated for 3 h with mouse anti-α-Syn primary antibody (1:1000, *BD Transduction Laboratories, USA*). Afterwards, cells were washed three times with PBS and incubated with a secondary antibody (Alexa Fluor 488 donkey anti-mouse IgG) for 2 h at room temperature. Finally, cells were stained with Hoechst 33258 (*Life Technologies- Invitrogen, Carlsbad, California, USA*) (1:5000 in PBS) for 5 min and maintained in PBS for epifluorescence microscopy.

Quantification of α-synuclein inclusions

Quantifications were performed as previously described (14). Briefly, transfected cells were detected and scored based on the α-Syn inclusions pattern. Results were expressed as the percentage of the total number of transfected cells. At least 50-100 cells were counted *per* condition.

Caenorhabditis elegans models of PD

Media

Nematodes were cultured and maintained following standard procedures (17). Briefly, animals were synchronized by hypochlorite bleaching, hatched overnight in M9 buffer (3 g/L KH_2PO_4, 6 g/L Na_2HPO_4, 5 g/L NaCl, 1 M $MgSO_4$), and subsequently cultured at 20 °C on nematode growth medium (NGM) [1 mM $CaCl_2$, 1 mM $MgSO_4$, 5 μg/mL cholesterol, 250 M KH_2PO_4 (pH 6.0), 17 g/L Agar, 3 g/L NaCl] plates seeded with the *E coli* OP50 strain (OD600 values between 1.3-1.6).

C. elegans strains

NL5901, *unc-119(ed3)* III; *pkIs2386 [Punc-54::α-SYN::YFP; unc-119(+)]* and N2 (*Bristol*) wild type strains were obtained from the *Caenorhabditis elegans Genetic Center* (CGC). To analyse α-Syn induced dopaminergic degeneration, we used the strain UA196 (15), gifted generously by the laboratory of Dr Guy Caldwell (*Department of Biological Science, The University of Alabama, Tuscaloosa, USA*), (*sid-1(pk3321)*; *baIn33* [*Pdat-*

1::sid-1, Pmyo-2::mCherry]; *baIn11* [*Pdat-1::α-SYN; Pdat-1::GFP*]). This strain was named *Pdat-1::GFP; Pdat-1::α-SYN* in the main text.

SynuClean-D administration

SynuClean-D was administered within NGM agar media and within the food source E. coli *OP50* strain. Briefly, 100 μM of a stock SC-D in 0.2% Dimethyl sulfoxide (DMSO) was added to liquid warm NGM to a final concentration of 10 μM. 250 μL of the *E. coli OP50* strain, containing 10 μM of SC-D or 0.2% of DMSO as control were added to dried plates, allowing the liquid to soak into the plates. Seeded plates were dried 24 hours before transferring the worms. Worms were transferred onto SC-D-seeded plates directly at larval stage L4, exposed to SC-D for 7 days and transferred daily to avoid cross progeny. EGCG was administered in the same manner.

Aggregate quantification

The quantification of aggregates was performed as previously described (16, 17). Briefly, NL5901 *pkIs2386* [*Punc-54::α-SYN::YFP; unc-119(+)*] animals were age-synchronized and left overnight to hatch. L1 animals were grown onto individual NGM plates containing *E. coli* OP50 to reach L4 developmental stage. Afterwards, L4 animals were transferred onto SC-D containing plates, while NGM plates containing only DMSO were used as negative control. Worms were transferred daily to avoid cross contamination. Aggregates were counted for each animal staged at L4+7 days in the posterior part. For each independent experiment, thirty 7-days old worms of each treatment were examined under a Nikon Eclipse E800 epifluorescence microscope equipped with an Endow GFP HYQ filter cube (*Chroma Technology Corp, Bellows Falls, Vermont USA*) and each experiment was performed by triplicate. Aggregates were defined as discrete, bright structures, with boundaries distinguishable from surrounding fluorescence. Measurements of the inclusions were performed using ImageJ software considering the area dimensions.

Thrashing assays

Animals from the strain NL5901 [*unc-119(ed3)* III; *pkIs2386* [*Punc-54::α-Syn::YFP; unc-119(+)*]] (SC-D treated and non-treated) at L4+7 days of development, were placed individually in a drop of M9 buffer and allowed to recover for 120 s (to avoid observing behaviour associated with stress). Afterwards, the number of body bends was recorded

for 1 min. Non-treated N2 wild type animals were used as control. Movies of swimming worms were recorded using a Leica MZFFLIII stereomicroscope at nominal magnification of 30X and the Hamamatsu ORCA-Flash 4.0LT camera at 17 frames per second (17 fps) for 1 min. Bends per minute were obtained with the Worm Tracker pluggin (wrMTrck), from the ImageJ software. Thirty animals were counted in each experiment unless stated otherwise. Experiments were carried out in triplicate. Statistical analysis was performed using Graphpad Prism software (*GraphPad Software, La Jolla, California, USA*).

Blinding of experiments and replicates

Thrashing, or body bending studies were completed such that the experimenter was blind to the genotype of the worms. Strains were given letter codes by another member of the laboratory and the code was not broken until all of the replicates for a particular assay were completed. For all assays, we completed a minimum of three biological replicates per strain.

C. elegans neurodegeneration assays

Worms were analyzed for α-Syn -induced dopaminergic neurodegeneration as described previously (17). Briefly, 20-30 L4 staged animals from the strain UA196 were transferred to SC-D-NGM plates and were grown for seven days after which the dopaminergic cell death induced by the over-expression of α-Syn was analysed by fluorescence. Plates containing only 0.2% DMSO, without SC-D, were used as control. Worms were transferred daily to avoid cross contamination.

The six anterior DA neurons (four CEP and two ADE DA neurons) were scored for neurodegeneration according to previously described criteria (18). Worms were considered normal when all six anterior DA neurons (four CEP (cephalic) and two ADE (anterior deirid) were present without any visible signs of degeneration. If a worm displayed degeneration in at least one of the six neurons, it was scored as exhibiting degeneration. For each independent experiment, thirty worms of each treatment were examined under a Nikon Eclipse E800 epifluorescence microscope equipped with an Endow GFP HYQ filter cube (*Chroma Technology Corp, Bellows Falls, Vermont, USA*).

Microscopy and imaging

Animals were placed in a 1 mM solution of sodium azide and mounted with a coverslip on a 4% agarose pad. Animals were visualized with a Nikon Eclipse E800 epifluorescence microscope. The system acquires a series of frames at specific Z-axis position (focal plane) using a Z-axis motor device. Animals were examined at 100× magnification to examine α-Syn induced DA cell death and at 40X to examine α-Syn apparent aggregates.

Statistical analysis

All graphs were generated with GraphPad Prism 6.0 software (*GraphPad Software Inc, La Jolla, California, USA*). Data were analysed by two-way ANOVA Tukey test using SPSS software version 20.0 (*IBM Analytics, Armonk, NY, United States*). All data are shown as means and standard error of mean (SEM). $p < 0.05$ was considered statistically significant. In the graphs *, ** and *** indicate $p < 0.05$, $p < 0.01$ and $p < 0.001$, respectively.

Supplementary Information for:

A novel small molecule inhibits α-synuclein aggregation, disrupts amyloid fibrils and prevents degeneration of dopaminergic neurons

Jordi Pujols[a,b,‡], Samuel Peña-Díaz[a,b,‡], Diana F. Lázaro[c], Francesca Peccati[d,e], Francisca Pinheiro[a,b], Danilo González[e], Anita Čarija[a,b], Susanna Navarro[a,b], María Conde-Jimenez[f], Jesús García[g], Xavier Salvatella[g,h], Javier Sancho[f], Mariona Sodupe[e,h], Tiago F. Outeiro[c,i,j], Esther Dalfó[k,l] and Salvador Ventura[a,b,*]

[a] Institut de Biotecnologia i Biomedicina. Universitat Autònoma de Barcelona, 08193-Bellaterra, Spain.

[b] Departament de Bioquímica i Biologia Molecular. Universitat Autònoma de Barcelona, 08193-Bellaterra, Spain.

[c] Deparament of Experimental Neurodegeneration, Center for Biostructural Imaging of Neurodegeneration, Center for Nanoscale Microscopy and Molecular Physiology of the Brain, University Medical Center Göttingen, Waldweg 33, 37073 Göttingen, Germany.

[d] Departament de Química, Universitat Autònoma de Barcelona , 08193 Bellaterra, Spain

[e] Sorbonne Universités, CNRS, Laboratoire de Chimie Théorique, LCT, F-75005 Paris, France

[f] Department of Biochemistry and Molecular and Cell Biology, Institute for Biocomputation and Physics of Complex Systems (BIFI), University of Zaragoza, 50018 Zaragoza, Spain

[g] Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Baldiri Reixac 10, 08028, Barcelona, Spain

[i] Max Planck Institute for Experimental Medicine, Göttingen, Germany.

[j] Institute of Neuroscience, The Medical School, Newcastle University, Framlington Place, Newcastle Upon Tyne, NE2 4HH, UK

[k] Medicine, M2 , Universitat Autònoma de Barcelona (UAB), Bellaterra Campus , Cerdanyola del Vallés, Barcelona, Spain. ORCID: 0000-0003-4677-8515

[l] Faculty of Medicine. University of Vic-Central University of Catalonia (UVic-UCC), Can Baumann, 08500, Vic, Spain

[*] Correspondence: salvador.ventura@uab.cat; Tel.: +34-93-5868956

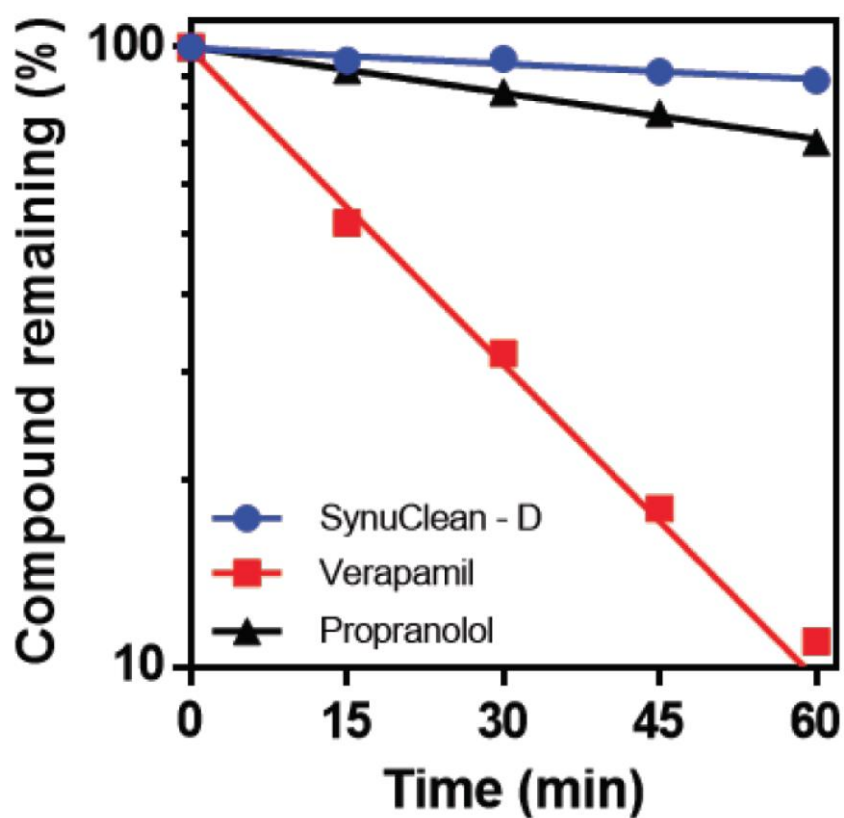[‡] These authors contributed equally to this work.

Figure S2. Metabolic stability of SC-D in human liver microsomes. The percentage of remaining compounds is plotted as a function of the incubation time at 37 ºC. Propranolol and verapamil are slowly and rapidly metabolized reference compounds, respectively. The data correspond to the mean of two experiments. The calculated $T_{1/2}$ of SC-D, propranolol and verapamil are 357.2, 119.5 and 19.0 min, respectively. The calculated intrinsic clearance values for SC-D, propranolol and verapamil are 3.9, 11.6 and 73.1 μL/min/mg, respectively.
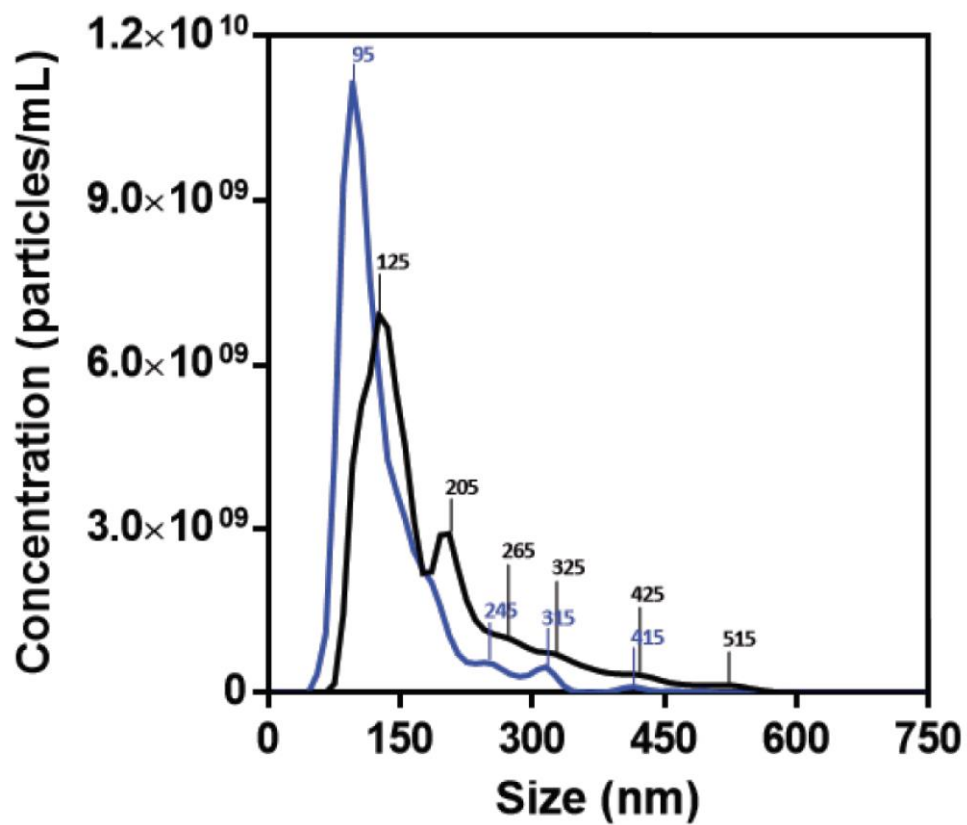
Figure S3. Fibril size characterization. Determination of the different fibril size population in absence (black) and presence (blue) of SC-D by using nanosight system.
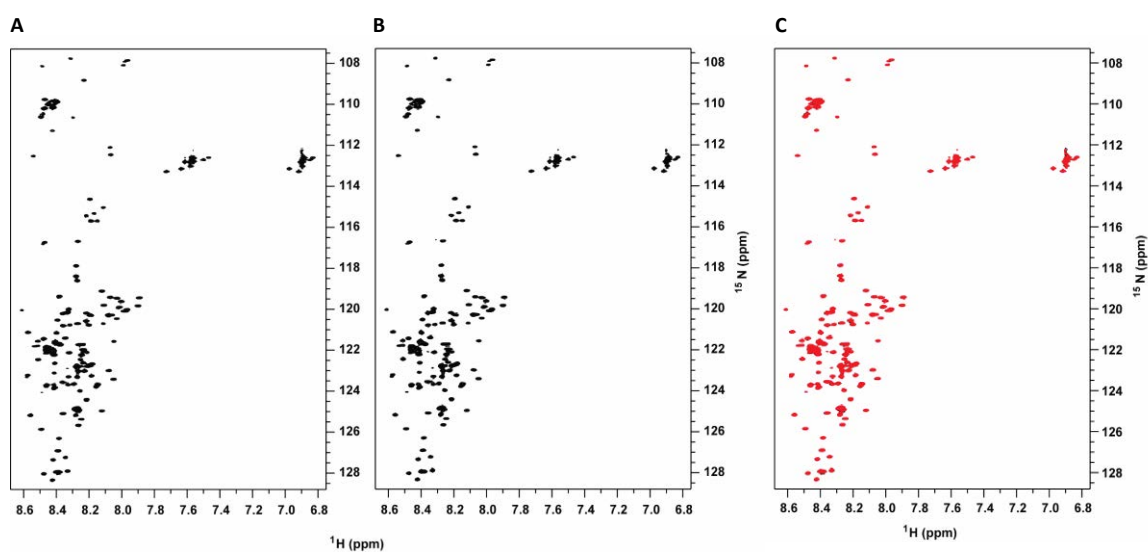
Figure S4. NMR characterisation of SynuClean-D non interaction to α-synuclein monomer. ^{1}H-^{15}N HSQC NMR spectra of ^{15}N-labeled α-Syn (70 **μ**M) in the absence (A) and in the presence (B) of SC-D (100 μM). Superposition of the spectra (C) shown in panels (A) and (B) illustrating that α-Syn exhibits identical NMR spectra in the absence (black contours) and in the presence (red contours) of SC-D.
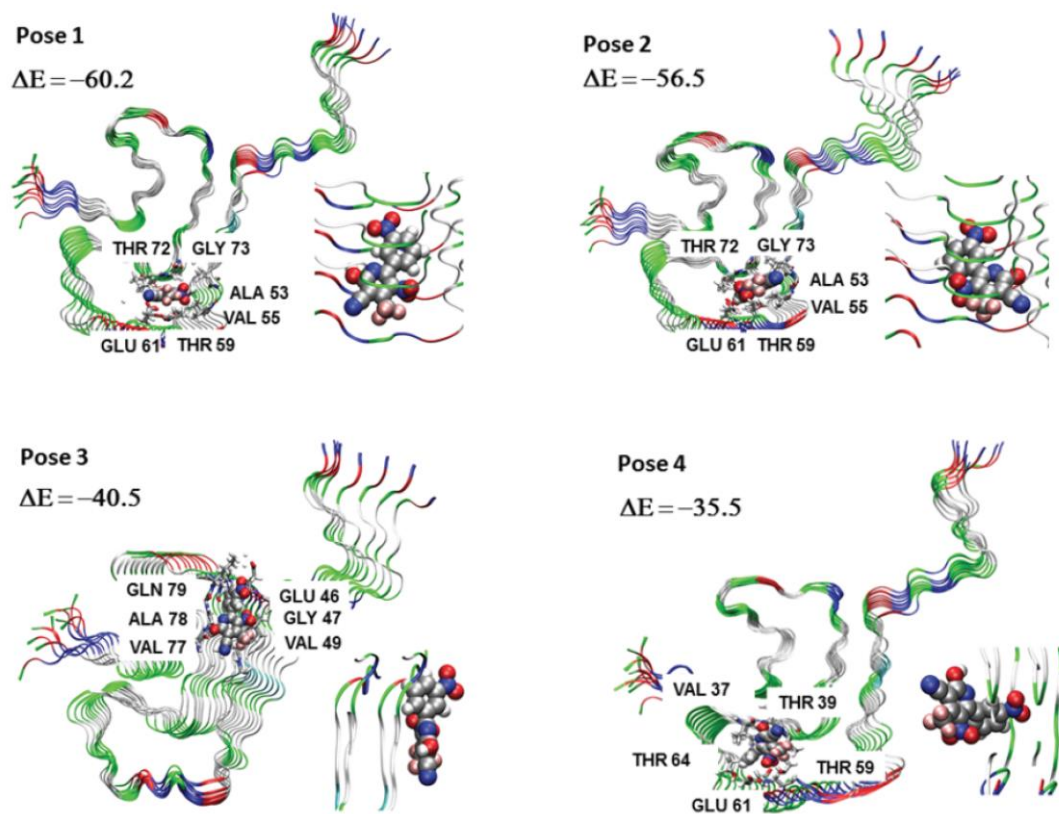
Figure S5. Binding sites prediction. Binding sites predicted by the Protein Energy Landscape Exploration (PELE) methodology. Interaction energies (DE) are expressed in kcal mol^{-1}.
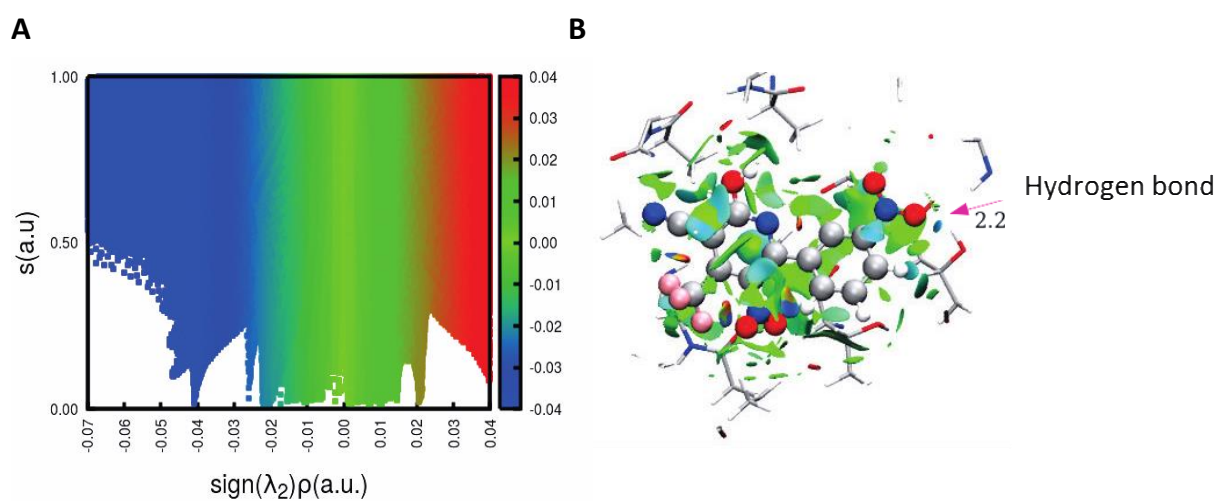
Figure S6. Characterisation of SynuClean-D-fibril interaction. Non-covalent interaction plot (A) and the representation of the non-covalent interactions involving SC-D in preferred binding pose at the PELE geometry (B). Contacts are represented in green. Distances in Å.
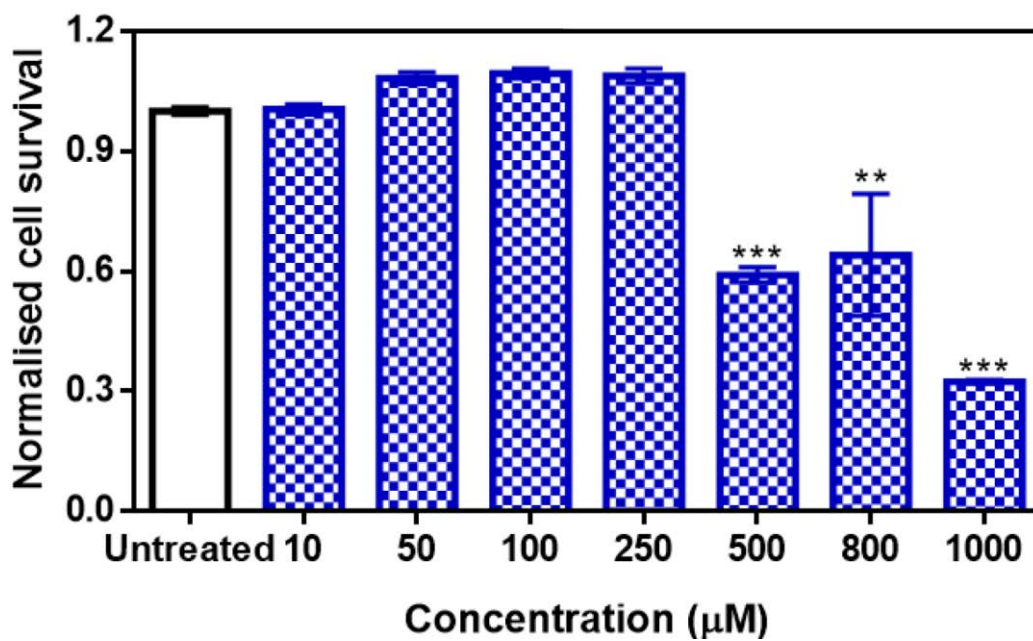
Figure S7. Toxicity of SC-D for neuroblastoma cells. Normalized SH-SY5Y cells survival in the presence of increasing concentrations of SC-D (blue) and without the compound (white).
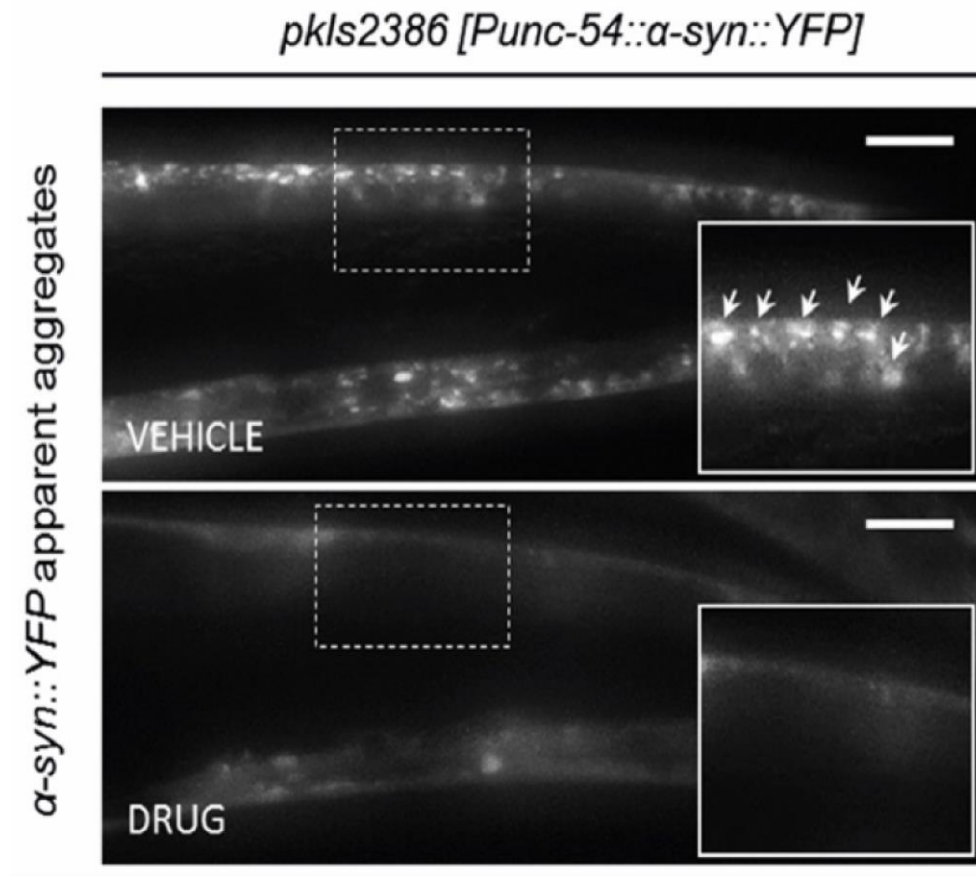
Figure S8. Inhibition effect of the compound in the formation of α-synuclein inclusions in a *C. elegans* model of PD. Selected images of α-Syn muscle aggregates obtained by epifluorescence microscopy of NL5901 worms treated without (top panel, vehicle) and with SC-D (bottom panel, drug).
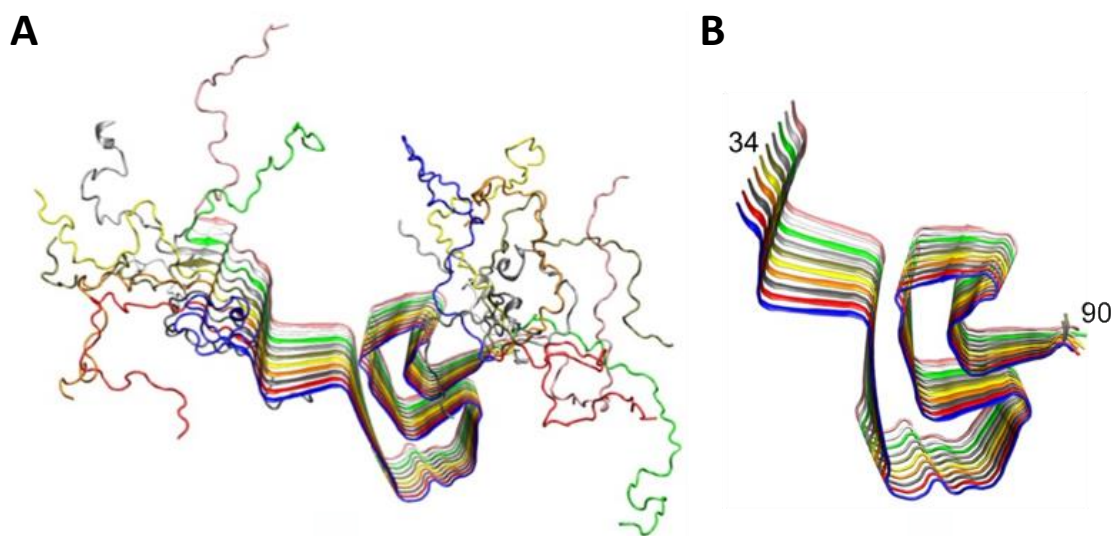
Figure S9. Aggregated structure of human α-synuclein (PDB code 2N0A). NMR full-length model of human α-Syn fibril (A). Reduced model (B), including residues 34 to 99 of PDB 2N0A employed in the binding site search and binding energy calculations.

A

B
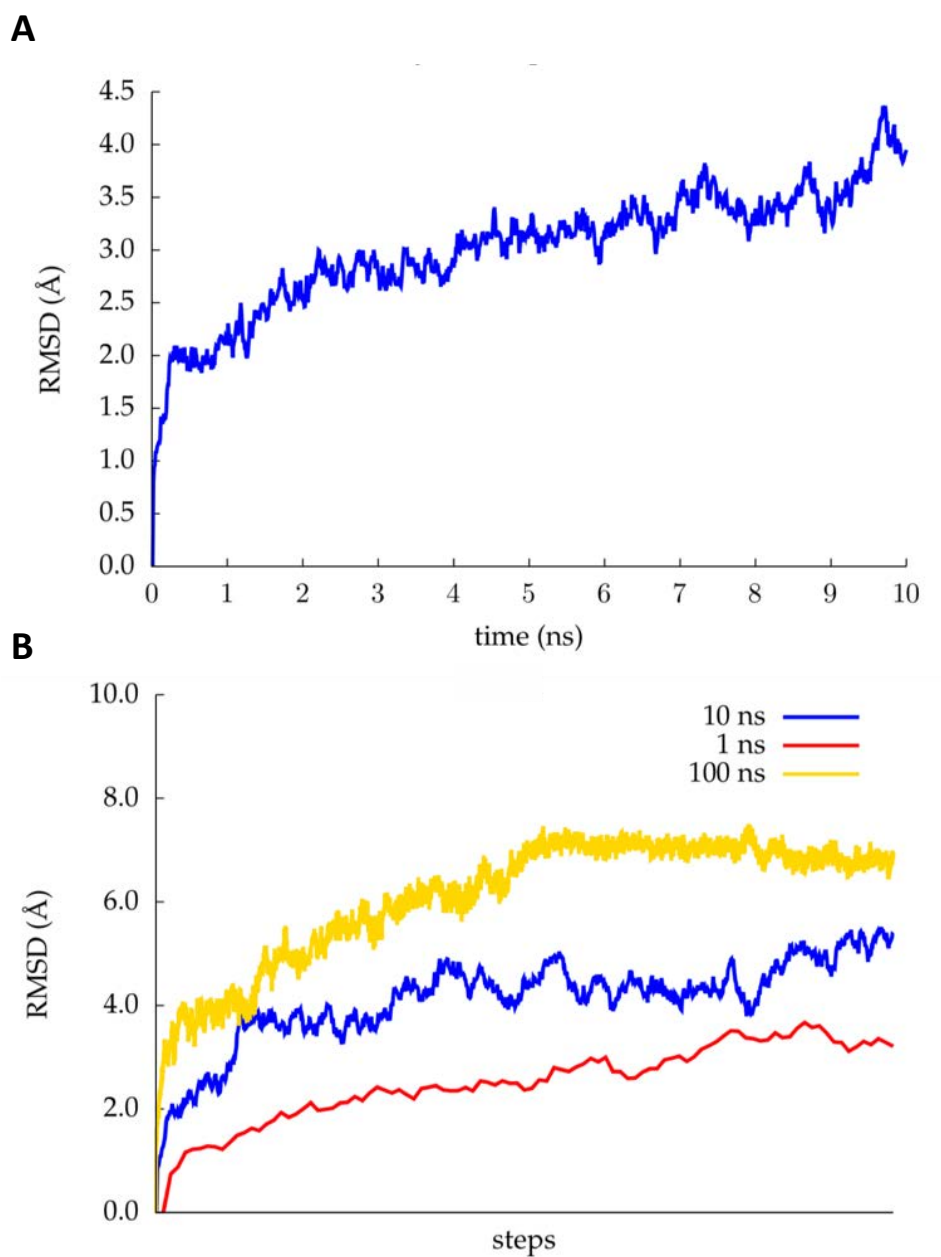
Figure S10. RMSD results of NPT molecular dynamics simulations. Backbone RMSD along 10 ns of NPT molecular dynamics simulation (A) at 300 K of the full α-synuclein fibril from PDB structure 2N0A, with RMSD computed on residues 34 to 99. Backbone RMSD computed along 1, 10 and 100 ns of NPT molecular dynamics simulations (B) of the reduced α-Syn fibril model comprising residues 34 to 99.

Work 6

ZPD-2, a small compound that inhibits

α-synuclein amyloid aggregation

and its seeded polymerization

ZPD-2, a small compound that inhibits α-synuclein amyloid aggregation and its seeded polymerization

Samuel Peña-Díaz[a,b, ‡], Jordi Pujols[a,b, ‡], María Conde-Giménez[c], Anita Čarija[a,b], Esther Dalfo[e,f], Jesús García[d], Susanna Navarro[a,b], Francisca Pinheiro[a,b], Jaime Santos[a,b], Xavier Salvatella[d], Javier Sancho[c] and Salvador Ventura[a,b,*]

[a] Institut de Biotecnologia i Biomedicina. Universitat Autonoma de Barcelona, 08193-Bellaterra, Spain.

[b] Departament de Bioquimica i Biologia Molecular. Universitat Autonoma de Barcelona, 08193-Bellaterra, Spain.

[c] Department of Biochemistry and Molecular and Cell Biology, Institute for Biocomputation and Physics of Complex Systems (BIFI), University of Zaragoza, 50018 Zaragoza, Spain

[d] Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Baldiri Reixac 10, 08028, Barcelona, Spain

[e] Medicine, M2, Universitat Autonoma de Barcelona (UAB), Bellaterra Campus, Cerdanyola del Valles, Barcelona, Spain. ORCID: 0000-0003-4677-8515

[f] Faculty of Medicine. University of Vic-Central University of Catalonia (UVic-UCC)

[*] Correspondence: salvador.ventura@uab.cat

[‡] These authors contributed equally to this work

Abstract

α-synuclein forms toxic intracellular protein inclusions and transmissible amyloid structures in Parkinson's disease. Preventing α-synuclein self-assembly has become one of the most promising approaches in the search for disease-modifying treatments for this neurodegenerative disorder. Here, we describe the capacity of a small molecule (ZPD-2), identified after a high-throughput screening, to inhibit α-synuclein aggregation. ZPD-2 inhibits the aggregation of wild-type α-synuclein and the A30P and H50Q familial variants in vitro at substoichiometric compound:protein ratios. In addition, the molecule prevents the spreading of α-synuclein seeds in protein misfolding cyclic amplification assays. ZPD-2 is active against different α-synuclein strains and blocks their seeded polymerization. Treating with ZPD-2 two different Parkinson's Disease Caenorhabditis elegans models which express α-synuclein either in muscle or in dopaminergic neurons substantially reduces the number of α-synuclein inclusions and decreases synuclein-induced dopaminergic neurons degeneration. Overall, ZPD-2 is a hit compound worth to be explored in order to develop lead molecules for therapeutic intervention in Parkinson's Disease.

Keywords: Parkinson's disease, α-synuclein, amyloid, protein aggregation, aggregation inhibitor, Caenorhabditis elegans, neurodegeneration.

Introduction

Parkinson's disease (PD) is a neurodegenerative disorder that affects about 0.3 % of the population and more than 1 % of people over 60 years of age (4 % over 80 years) [1; 2]. It is characterised by the loss of dopaminergic (DA) neurons in *substantia nigra pars compacta*, which compromises the motor capacity of PD-suffering patients, producing tremor, rigidity and bradykinesia [3]. Additionally, since the disease spreads to the cerebral cortex [4], symptoms could include emotional and cognitive impairment [3]. Nowadays, treatments are focused on alleviating the above mentioned motor symptoms, mostly using dopamine replacement by administration of dopamine precursor (L-DOPA), combined with carbidopa, a L-DOPA decarboxylase inhibitor, and/or catechol-O-methyl transferase inhibitors and monoamine oxidase-B inhibitors [2]. However, these treatments do not prevent the progression of PD and they lose efficacy as the disease advances.

PD is pathologically characterized by the accumulation of protein aggregates in the neuronal body, Lewy bodies (LB), and/or fibrils deposited in neuronal processes, Lewy neurites (LN), of affected neurons [5]. These inclusions are mainly composed of α-synuclein (α-Syn), a protein predominantly expressed in the synaptic termination of dopaminergic neurons [6]. This evidence, together with the identification of mutations in the gene that encodes for this protein (SNCA) as the cause behind familial cases of PD [7] and the observation that duplications and triplications of the SNCA gene lead to highly penetrant forms of the disease [8; 9] directly connect PD and α-Syn. In fact, the presence of aggregated α-Syn in the brain is a common feature of a group of diseases named synucleinopathies, which, in addition to PD, include Dementia with Lewy's Bodies (DLB) and Multiple System Atrophy (MSA), among others [3].

In solution, α-Syn is a 140 amino acid intrinsically disordered protein whose function seems to be related with vesicle trafficking [6]. *In vitro* it forms thermodynamically stable amyloid aggregates [10] that can display different conformational features [11]. The formation of amyloids by α-Syn follows the typical sigmoidal kinetics, reflecting a nucleation-polymerization process [12]; although secondary nucleation reactions might also occur [13]. *In vivo*, α-Syn assemblies exert a toxic effect [14] and could be transmitted from cell to cell in a prion-like manner by seeding native α-Syn aggregation in previously unaffected neurons [15].

Preventing α-Syn aggregation seems to hold the potential to achieve significant therapeutic impact. Several strategies have been developed towards this objective: SNCA gene-silencing approaches to decrease the protein levels [16], methods to increase the clearance of aggregated α-Syn by autophagic and proteasomal machineries [17], and molecules intended to avoid the formation and/or propagation of aggregated α-Syn [18; 19]. One of the main limitations of this last strategy is the absence of a well-defined structure of monomeric α-Syn in solution, due to its intrinsically disordered nature, which hampers the rational design of inhibitors. High-throughput screening protocols have been developed to circumvent this problem [20; 21]. A number of promising small molecules have been discovered with this approach, including anle138b [20], BIOD303 [22], fasudil [23], squalamine [24] or SynuClean-D [25]. In this context, we have developed a robust screening and validation protocol to analyse large chemical libraries in the search for effective inhibitors of α-Syn aggregation [26]. The *in vitro* pipeline integrates thioflavin-T (Th-T) fluorescence and light scattering measurements, transmission electron microscopy and protein misfolding cyclic amplification assays (PMCA). This approach allowed us to identify ZPD-2 (**Fig. 1**) as a novel small molecule able to inhibit the aggregation of *wild-type* (WT) α-Syn, as well as that of the A30P [27] and H50Q [28] familial mutants, being active against the seeded polymerization of different α-Syn strains. The compound displayed low toxicity for neuronal human cells and demonstrated significant inhibitory capacity in two well-established *Caenorhabditis elegans* models of PD [29; 30].
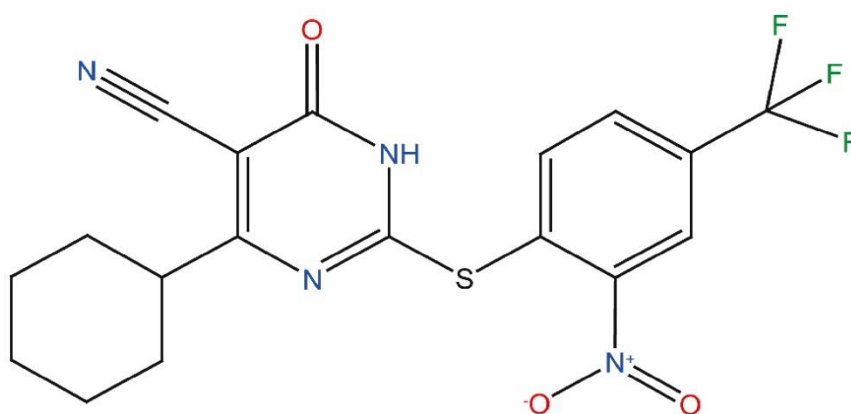

Figure 1. Chemical structure of the compound ZPD-2. ZPD-2 corresponds to 4-cyclohexyl-2-((2-nitro-4-(trifluoromethyl)phenyl)thio)-6-oxo-1,6-dihydropyrimidine-5-carbonitrile.

Results

ZPD-2 reduces and delays the aggregation of human α-synuclein in vitro

We designed and optimized a screening protocol that allows to follow the aggregation kinetics of α-Syn by monitoring Thioflavin-T (Th-T) fluorescence emission for 32 h. This approach permitted us to study the inhibitory potential of more than 14,000 compounds [25; 26]. The activity of molecules able to reduce significantly the final amount of Th-positive material and/or impact the nucleation or elongation rates of the reaction was further confirmed using light scattering and TEM measurements at the end of the reaction. This allowed us to identify 30 active compounds, most of which seem not to be connected in terms of structure, precluding QSAR studies. We have previously described the properties of SynuClean-D (SCD) a molecule that acts preferentially on top of α-Syn proto-fibrillar or fibrillar assemblies [25]. Here, we describe the properties of ZPD-2 (Fig. 1), a compound that differs in its mechanism of action. SCD and ZPD-2 share a benzotrifluoride group, which suggested that it could constitute the minimal inhibitory unit; however, this group is devoid of any anti-aggregation activity by itself (unpublished), indicating that, most likely, it only acts as a framework for the different active groups in the two molecules.

The incubation of 70 μM of α-Syn in the presence and absence of 100 μM of ZPD-2 revealed that the compound modulated the protein aggregation, reducing the formation of Th-T positive structures at the end of the reaction by an 80 %, while extending t50 by 8 h (Fig. 2A). The analysis of the kinetics revealed a reduction in the nucleation rate constant in presence of ZPD-2 (kb= 0.008833) by three-fold, when compared to the control reaction (kb=0.02754). The autocatalytic rate constant was also lower in the treated sample (ka=0.2432 h-1) than in the control (ka= 0.3230 h-1). Light scattering measurements at 300 nm confirmed that the observed reduction in Th-T fluorescence corresponds to an effective decrease in the levels of α-Syn aggregates, with a 67 % decrease in the dispersion of light in the presence of ZPD-2 (Fig 2B). Transmission Electron Microscopy (TEM) images corroborated that the samples incubated with ZPD-2 (Fig. 2D) contained less fibrils per field than the non-treated ones (Fig. 2C). In good agreement with these data, quantification of soluble α-Syn at the end of the aggregation reaction indicated that its level was three-fold higher in ZPD-2 treated samples (Fig. S1A).
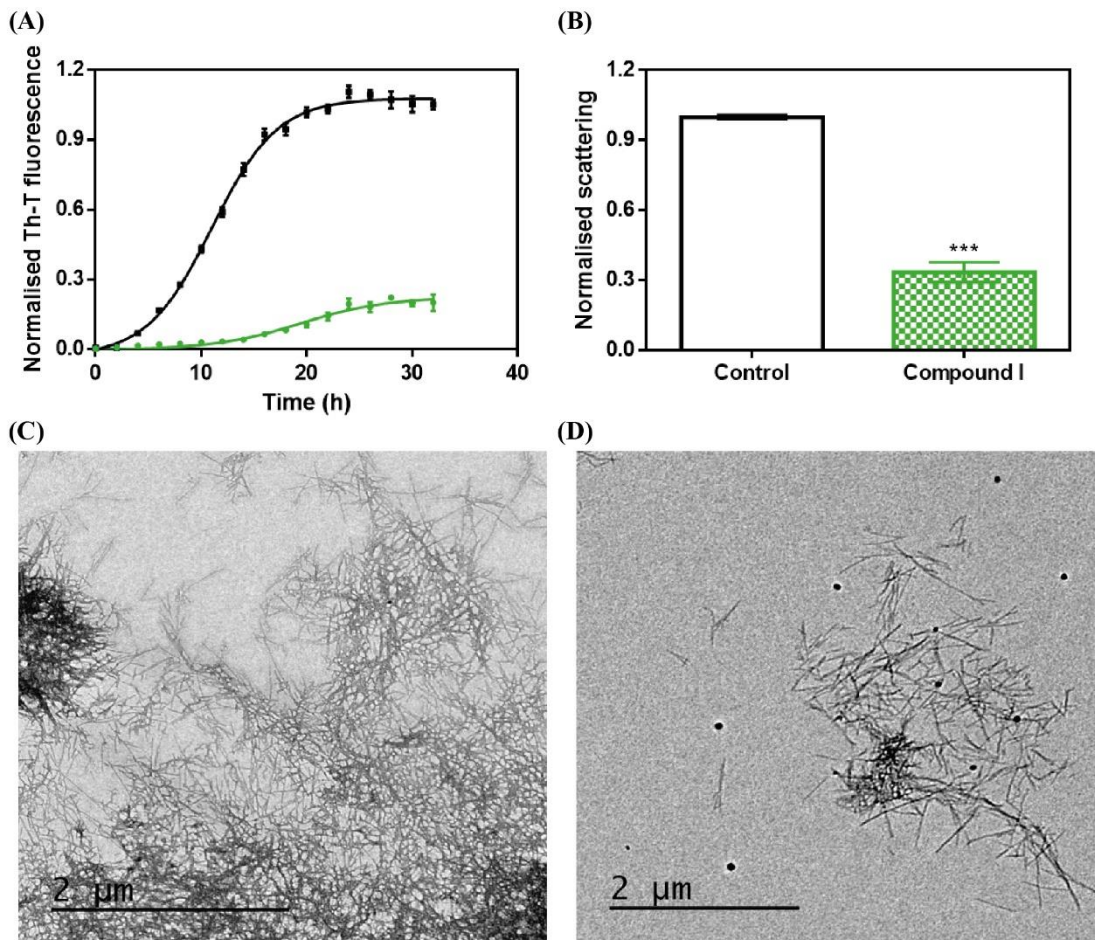
Figure 2. ZPD-2 inhibits the aggregation of wild-type α-synuclein in vitro. (A) Aggregation kinetics of α-Syn in absence (black) and presence (green) of ZPD-2. Intensity of Th-T fluorescence is plotted as a function of time. (B) Light scattering of end-point aggregates is measured at 300 nm for untreated (white) and ZPD-2 treated samples (green). (C and D) Representative TEM images of untreated (C) and ZPD-2 treated (D) samples. Th-T fluorescence is expressed as normalized means. Final points were obtained at 48 h after the aggregation reaction begin. Error bars are shown as standard errors of means values, *** indicates *p < 0.001*.

Further analysis of the inhibition capacity of ZPD-2 indicated that it exhibited a dose-dependent effect, displaying a statistically significant effect even at 10 μM (1:7 compound:protein ratio) (Fig. 3A), where the final Th-T signal was reduced by 49 %.
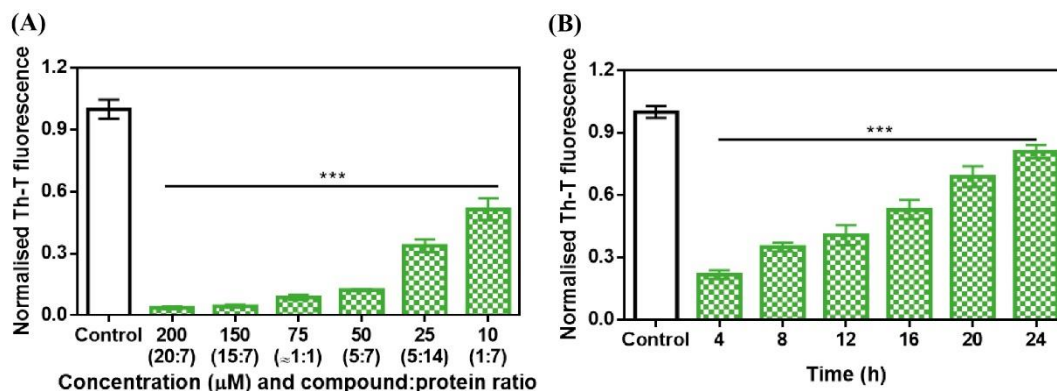
Figure 3. Analysis of the inhibitory capacity of ZPD-2. (A) Titration of the effect of different concentrations of ZPD-2 on 70 μM α-Syn aggregation. (B) Th-T fluorescence of α-Syn end-point aggregates after the addition of ZPD-2 at different time points during the aggregation kinetics. Th-T fluorescence is plotted as normalized means. End-points were obtained at 48 h of α-Syn incubation. Error bars are shown as standard errors of mean values, *** indicates $p < 0.001$.

To address the time window in which ZPD-2 is active, we set up aggregation reactions with a constant amount of ZPD-2 added at different time points after the reaction begins. A time-dependent response was observed (Fig. 3B), with a very significant inhibition when ZPD-2 was added at early (4 – 8 h) and intermediate (12 – 16 h) times, and a less pronounced effect when it was added at the plateau phase (20 – 24 h). This indicates that ZPD-2 is mostly active against the species formed early in the aggregation reaction, consistent with its highest impact on the nucleation rate constant kb. Importantly, NMR studies using isotopically labelled monomeric and soluble α-Syn indicated that ZPD-2 does not interact with its native form, since we could not detect any perturbations in chemical shifts or peak intensities in α-Syn in the presence of a molar excess of the molecule (Fig. S2).

Several α-Syn single point mutations are connected with the onset of familial cases of PD [27; 28]. We studied the ability of ZPD-2 to prevent the aggregation of two of the most frequent and aggressive variants, H50Q and A30P. The molecule was also active against these α-Syn forms in kinetic assays (Fig. 4A). According to the relative Th-T signal at the end of the reaction in ZPD-2 treated and non-treated samples, the molecule inhibited the aggregation of A30P and H50Q by 96 % and 94 %, respectively (Fig. 4B).
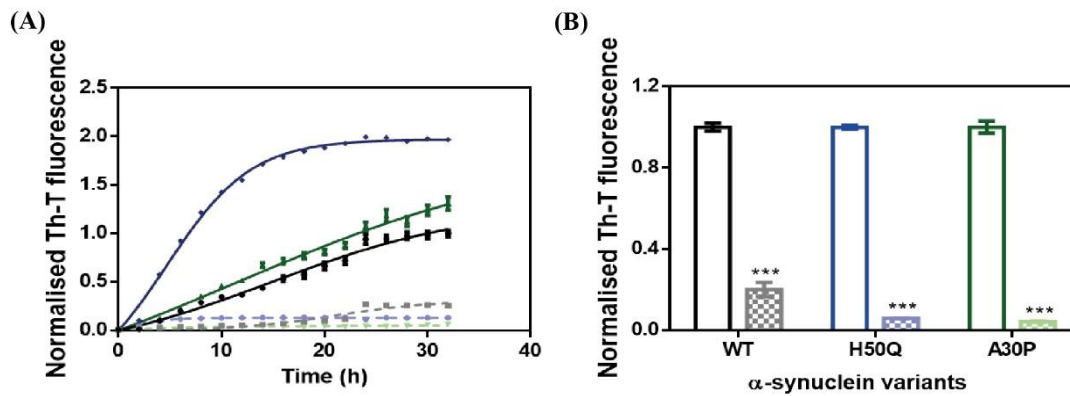
(A)

(B)

Figure 4. ZPD-2 inhibits the aggregation of α-synuclein familial variants. (A) Aggregation kinetics of WT (black), H50Q (blue) and A30P (green) variants of α-Syn in presence (dotted) and absence (continuous) of ZPD-2, using Th-T as reporter. (B) End-point measurements of the aggregation of WT, H50Q and A30P variants of α-Syn in presence (dotted) or absence (continuous) of ZPD-2 Th-T fluorescence is expressed as normalized means. Error bars are shown as standard errors of mean values.

ZPD-2 prevents a-syn seeded aggregation in Protein Misfolding Cyclic Amplification assays

Protein misfolding cyclic amplification (PMCA), initially developed to study the polymerization and propagation process of the prion protein [39; 40], has been recently adapted for α-Syn amyloid aggregation [35]. Essentially, cycles of incubation at 37oC are followed by vigorous sonication in order to allow fibril growth and subsequent fibrillar rupture, thus producing α-Syn seeds. These preformed seeds are used to trigger the aggregation of fresh protein in the following cycle, amplifying the fibrillar content. At 90 μM of α-Syn, PMCA produced amyloid structures resistant to protease K (PK) digestion, as observed by SDS-PAGE, with the maximum protection arising after 4 rounds (Fig. 5A middle panel). Th-T fluorescence measurements of the same samples indicated that this protection correlates with an increasing presence of amyloid-like assemblies (Fig. 5B). In sharp contrast, in the presence of ZPD-2, the amount of PK resistant protein after 4 rounds is negligible (Fig. 5A right panel), Th-T fluorescence signal being also significantly low relative to control samples at this stage (Fig. 5B). These results suggested that ZPD-2 was strongly interfering with the PMCA-promoted seeding of α-Syn amyloids. The fact that Th-T decrease becomes significant only at pass 4, likely indicates that the aggregated non PK-resistant species generated at early steps still retain certain Th-T binding ability, since SDS-PAGE analysis indicates that the levels of PK-resistant protein is already decreased in treated samples at passes 1 to 3 (Fig. S3).
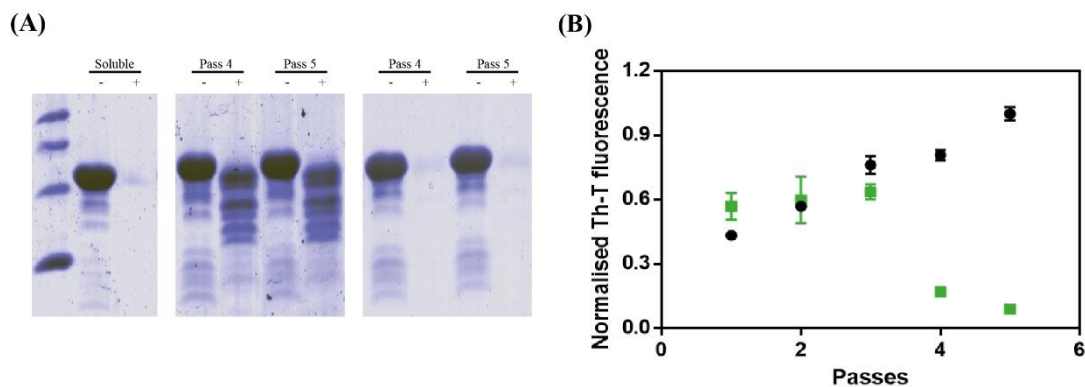
(A)

(B)

Figure 5. PMCA of α-synuclein in presence of ZPD-2. (A) Tricine-SDS-PAGE gels of untreated (middle) and ZPD-2 treated (right) PMCA samples before (-) and after (+) being digested with proteinase K. (B) Th-T fluorescence of different PMCA cycles of treated (green) and untreated (black) samples. Soluble α-Syn and PMCA steps 4 and 5 are shown. Th-T fluorescence is plotted as normalized means. Error bars are shown as standard errors of mean values.

ZPD-2 prevents the aggregation of different α-synuclein amyloid conformations

The aggregation of α-Syn has been described to lead to the formation of different amyloid conformations, or strains, depending on the environmental conditions [11]; a property that has been linked with its spreading in the brain and the manifestation of different synucleinopathies [34]. We analysed the capacity of ZPD-2 to prevent the aggregation of α-Syn into different previously described amyloid conformations [32; 33]. We refer them as strain B (buffer B, 50 mM Tris-HCl pH 7.0) and strain C (buffer C, 50 mM Tris-HCl pH 7.0 supplemented with 150 mM NaCl), to keep the original strain nomenclature. ZPD-2 was active in both cases (Fig. 6A and 6E), inhibiting by up to 90 % the formation of the amyloid strains B and C, as monitored by Th-T fluorescence. Light scattering measurements (Fig. 6B and 6F) and TEM imaging Fig. 6C, 6D, 6G and 6H) and soluble protein quantification at the end of the reaction (Fig. S1B) of the different samples at the end of the reaction confirmed the inhibitory activity of ZPD-2 against the two strains. Nonfibrillar aggregates might be necessary for fibril formation (obligate), able to convert into fibrils, but not indispensable for fibril formation (on-pathway), or unable of converting directly to fibrils (off-pathway). The difference between the large reduction in Th-T fluorescence promoted by ZPD-2 in strain C aggregation kinetics and the moderate impact the molecule has in light scattering and soluble protein levels might indicate the formation of Th-T negative off-pathway aggregates in these conditions, since they do not evolve into fibrils. However, their size should be rather small, since we did not observe any large amorphous aggregate in ZPD-2 treated samples (Fig. 6H).
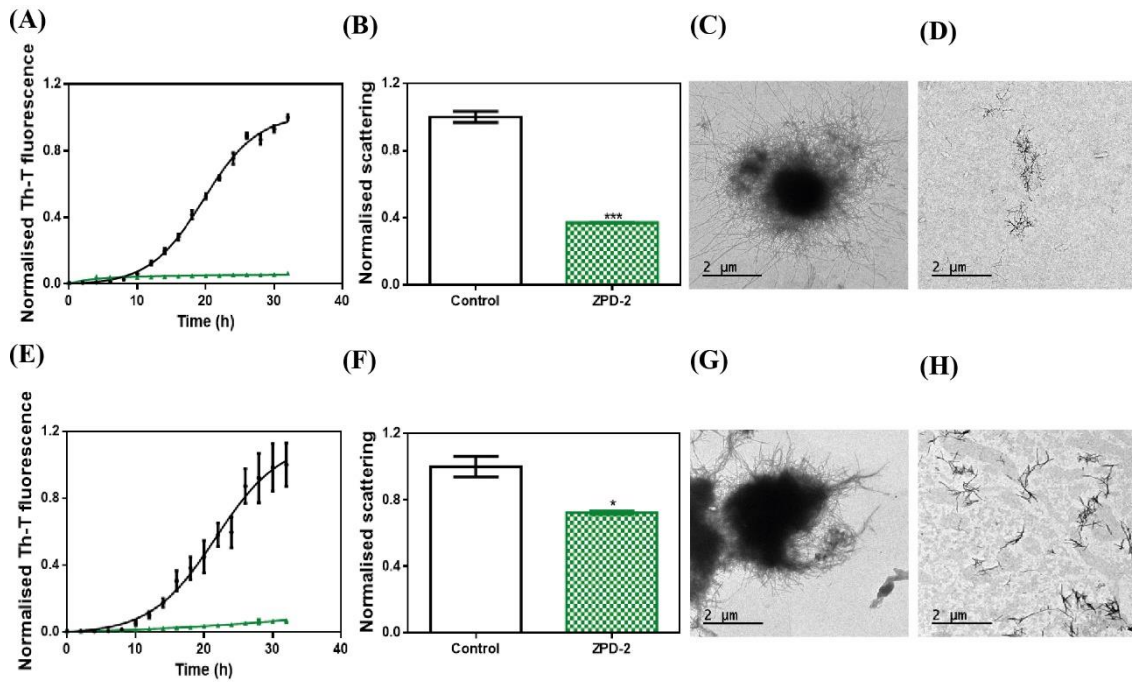
Figure 6. ZPD-2 blocks the aggregation of two different α-synuclein strains. (A and E) Aggregation kinetics of α-Syn strain B (A) and C (E) in absence (black) and presence (green) of ZPD-2. (B and F) Light scattering final point measurements at 300 nm of untreated (white) and ZPD-2 treated samples (green) of strain B (B) and C (F). (C, D, G, H) Representative TEM images of untreated α-Syn aggregates (C and G) and treated (D and H) samples for strains B and C, respectively. Th-T fluorescence is expressed as normalized means. Final points were obtained at 48 h after the aggregation reaction begins. Error bars are shown as standard errors of mean values, where $p < 0.05$, $p < 0.01$ and $p < 0.001$ were indicated by *, ** and ***, respectively.

We addressed whether the strong inhibitory capability of ZPD-2 at neutral pH can be overridden by the presence of preformed fibrils able to seed the aggregation reaction. The addition of 1 % (v/v) of seeds effectively accelerated the formation of both B and C strains (Fig. 7A and 7B). However, the presence of ZPD-2 abrogates this effect, reducing the final amount of amyloid-like structures in seeded reactions by an 87 % for strain B (Fig. 7A) and a 90 % for strain C (Fig. 7B), according to Th-T fluorescence. Again, light dispersion measured at 300 nm revealed a significant decrease of aggregates by 57 % and 70% in the case of strains B and C, respectively (Fig. 7C and 7D).
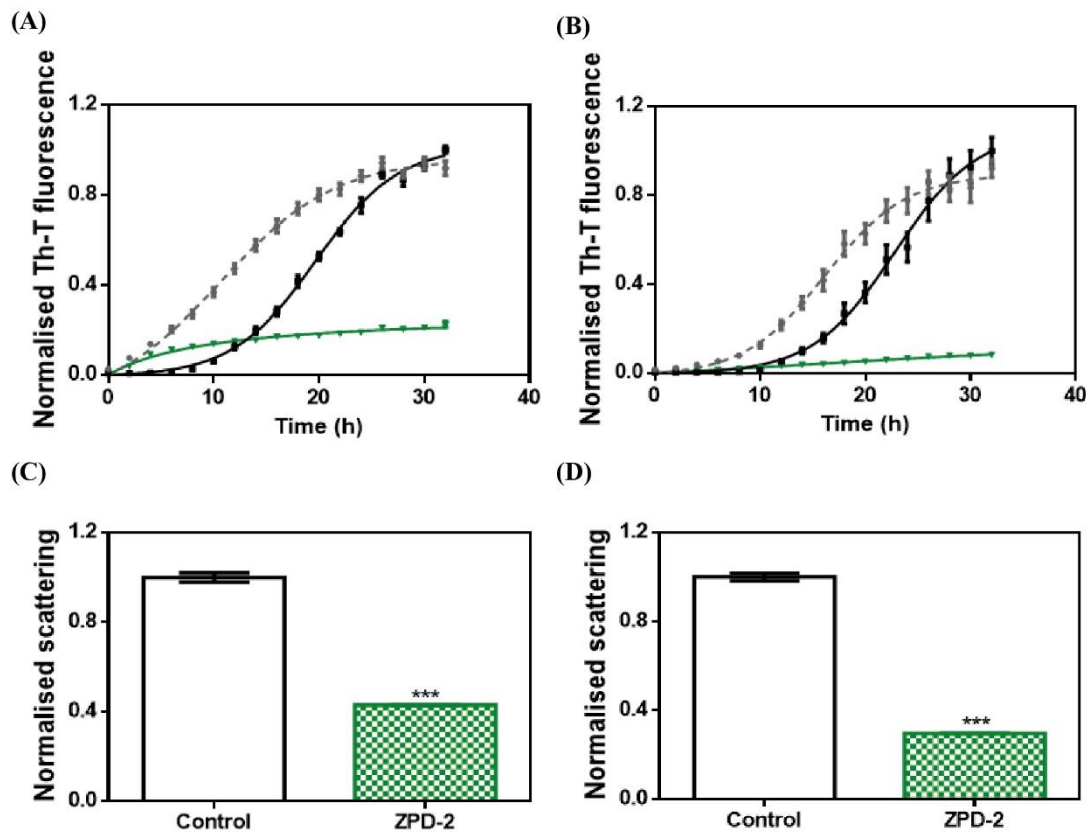
Figure 7. Seeding assays with three different strains. (A and B) Aggregation kinetics of α-Syn, buffer B (50 mM Tris-HCl pH 7.0) (A), or buffer C (50 mM Tris-HCl pH 7.0 supplemented with 150 mM NaCl) (B), reported by Th-T fluorescence, in absence of compounds and seeds (black), in presence of 1 % (v/v) of preformed seeds at the specific condition (grey dotted line) and in presence of seeds and 100 µM of ZPD-2 (green). Light dispersion of treated (green) and untreated (white) seeded samples at final point of strain B (C) and strain C (D). Error bars are shown as standard errors of mean values, *** indicates $p < 0.001$.

ZPD-2 reduces the formation of α-synuclein inclusions in a C. elegans model of PD

We assessed the toxicity of ZPD-2 for human neuroblastoma cells. No significant toxicity was observed when the molecule was added to the cell culture up to 80 µM (Fig. S4). We skipped efficacy studies on neuroblastoma cells, because, with more than 20 different compounds analysed, we could not find a straightforward connection between the potency of the molecules in cell cultures and that in our C. elegans models of PD. We first analysed the effect of ZPD-2 in the C. elegans strain NL5901. This strain over-expresses human α-Syn fused to the yellow fluorescent protein (YFP), under the control of the muscular unc-54 promoter, transgene phIs2386 [Punc-54::α-SYN::YFP]. The expression of human α-Syn in the muscle of this nematode has been successfully used to identify modifier genes [29; 41]. Animals at the fourth larval stage (L4) were incubated in the presence or absence of 10 µM ZPD-2 and analysed at 9 days post-hatching (L4+7). These aged worms, which

mimic aged PD patients, were then analysed by epifluorescent microscopy and the number of visible α-Syn inclusions was quantified (Fig. 8A-B). In these assays, ZPD-2 moderately, but significantly, reduced the number of apparent aggregates (25.7 ± 1.3) when compared to untreated worms (31.8 ± 1.7) (Fig. 8C). In addition, worms treated with ZPD-2 showed an increase in their mean lifespan of 14.2 %, relative to untreated animals (P value = 0.015, Wilcoxon unpaired test) (Fig. S5).
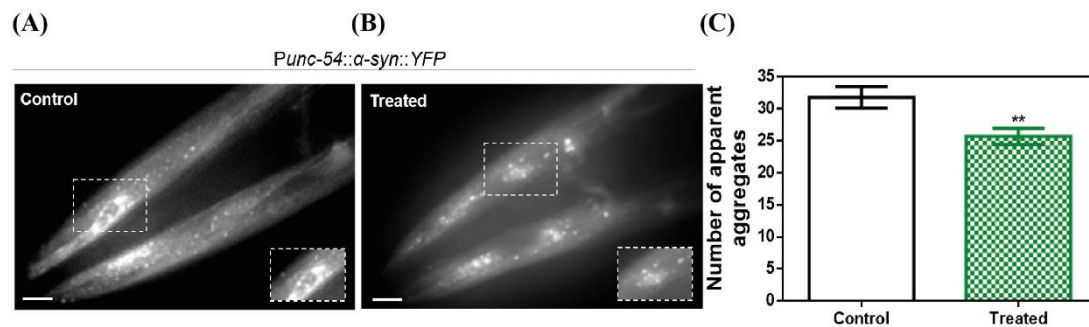
Figure 8. In vivo anti-aggregational assays in *Caenorhabditis elegans*. Representative images of apparent α-Syn aggregates in *C. elegans* body wall muscle cells obtained by epifluorescence microscopy of NL5901 worms treated without (A) and with ZPD-2 (B). Quantification of α-Syn muscle inclusions in the absence (white) and presence of ZPD-2 (green). ** indicates *p < 0.01*.

Neuroprotective role of ZPD-2 in a C. elegans model of PD

The loss of DA neurons is one of the most important characteristics of PD and an important target in the search for a future treatment for this disorder. C. elegans presents a total of four pairs of DA neurons, three of them in the anterior part (CEPD, CEPV and ADE) and one pair in the posterior part (PDE) [38]. The existence of 6 anterior DA neurons has been recently used to analyse PD-related processes in a model (strain UA196) that expresses both human α-Syn and GFP under the control of the dopamine transporter promoter (Pdat-1::GFP; Pdat-1::α-SYN) [42]. Human α-Syn expression in these DA neurons induces a progressive degeneration process [43]. At 9 days post-hatching, the number of remaining functional neurons of untreated (Fig. 9A) and ZPD-2 treated (Fig. 9B) worms was analysed. As an average, in control worms 48.1 % of DA neurons are non-functional, whereas in treated animals this value decreases to 40.4 % (P value = 0.038, Wilcoxon unpaired test). Despite the difference between both means is rather low, the distribution of the data indicated a displacement in the DA neurons survival profile (Fig. 9C and Fig. S6) in the presence of ZPD-2 when compared to the control worms. As a result, there is a significant increase in the number of worms containing more than 3

functional neurons in the anterior region in the presence of ZPD-2 (51.0 ± 4.8 %) when compared to the controls (29.1 ± 3.1 %) (Fig. 9D).
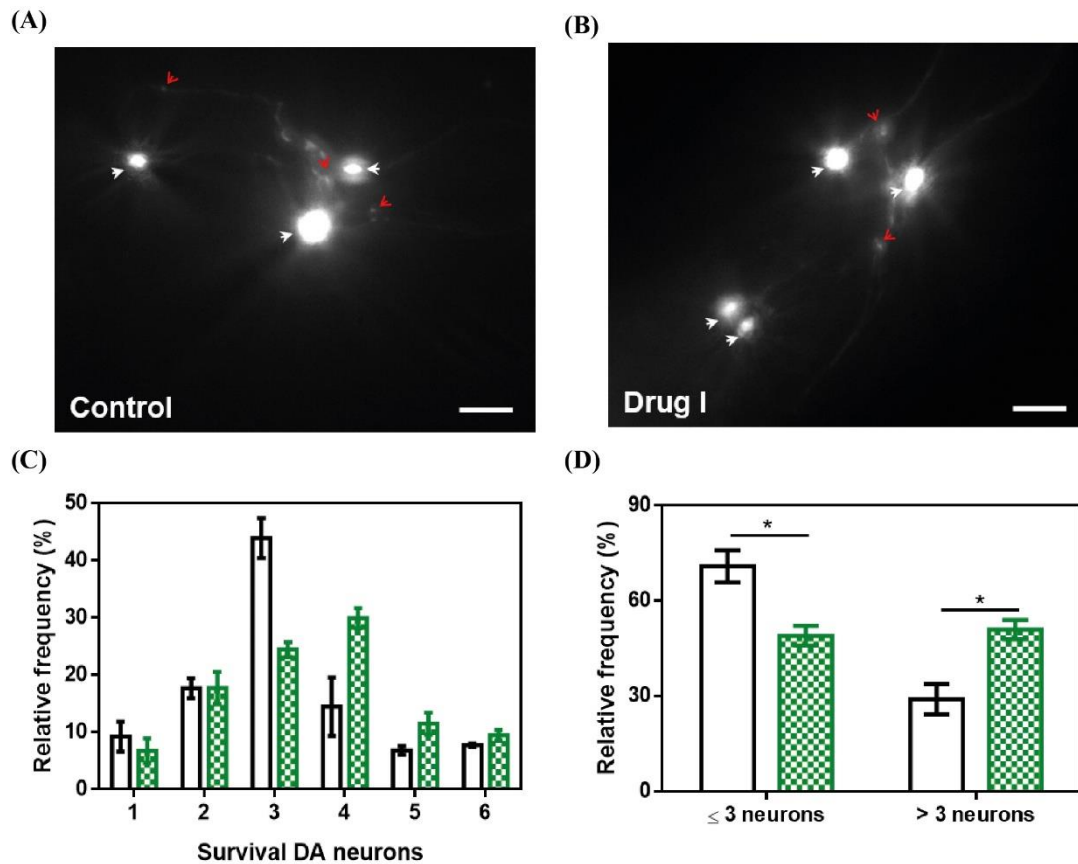
Figure 9. Neuroprotective activity of ZPD-2 in a *Caenorhabditis elegans* model of PD. Representative images of GFP and α-Syn expressing anterior DA neurons in worms treated without (A) and with ZPD-2 (B) for 7 d after L4. Healthy neurons are labelled with white arrows. (C) Distribution of DA surviving neurons in the anterior region of worms. (D) Percentage of worms containing 3 neurons functional or less and more than three functional neurons after 7 days post-hatching. White bars indicate control samples while the green ones correspond to treated samples. * indicates *p < 0.05*.

Discussion

Protein aggregation is tightly connected with neurodegenerative disorders such as Alzheimer's and Parkinson's diseases. Immediately after the identification of α-Syn as the main fibrillar component in LBs and LNs [5; 44] it became evident that targeting the aggregation of this protein might hold therapeutic potential [23].

Nevertheless, the absence of a defined three-dimensional structure for the functional state of α-Syn due to its intrinsically disordered nature makes the rational design of effective inhibitors that stabilize α-Syn and thus prevent or delay its aggregation, as it has been successfully done for globular proteins like transthyretin [45; 46], difficult. In this scenario, evaluation of large chemical libraries appears as one of the few strategies we have to discover an effective inhibitor of α-Syn deposition and, indeed, this approach has already rendered promising molecules [20; 22; 23; 24; 25]. In the present work, we describe the discovery of ZPD-2, a small molecule able to prevent up to 90 % the in vitro aggregation of WT α-Syn and of familial mutants of the protein when used in a 0.7:1 (protein:ZPD-2) ratio, delaying also significantly the completion of the reaction. Its inhibitory capacity was confirmed by orthogonal techniques such as light scattering and TEM.

Further analysis demonstrated that ZPD-2 was able to prevent the aggregation in a concentration-dependent manner, with ~50% inhibition at a 7:1 protein:compound ratio. This, together with solution NMR measurements indicate that ZPD-2 does not interact significantly with soluble monomeric α-Syn, which suggests that it will not interfere with the functional state of the protein. In addition, the inhibitory potential of ZPD is time-dependent, being more significant at early (0-8 h) stages, in fair contrast with SC-D, a compound we identified in the same screening campaign, whose activity was time-independent, being able to target late species [25]. The largest affinity of ZPD-2 for early aggregating species is also inferred from the fact that it mainly impacts the nucleation constant, reducing it by three-fold. This might also explain why, at a 0.7:1 ratio, the molecule works well for the A30P (96 % inhibition) and H50Q (94 % inhibition) familial variants, provided that both mutations facilitate oligomerization, H50Q favouring also fibrillation [47].

ZPD-2 is able to inhibit the aggregation of α-Syn under different solution conditions. This ability opens a possibility for its use in different synucleinopathies, where different α-Syn

strains might occur [33; 34]. Importantly, ZPD-2 is one of a few small molecules shown to inhibit efficiently α-Syn seeded aggregation, where the lag phase of the reaction is shortened or abrogated because the soluble protein can be directly incorporated on top of the preformed fibrillar fragments. This seeding-blocking activity explains why ZPD-2 is so effective preventing the formation of PK-resistant/Th-T-positive species in PMCA assays, which promote both templated seeding and aggregates amplification. This effect might respond to the ability of the compound to either destabilize small aggregates or to prevent their elongation, a property that can be very relevant to prevent the cell-to-cell spreading of misfolded α-Syn.

ZPD-2 had not detectable toxic effect for neuronal cells at 10 μM, a concentration at which it reduces the presence of α-Syn inclusions in a C. elegans model of PD expressing human α-Syn in body wall muscle cells and extends lifespan. Not surprisingly, this anti-aggregational activity translates in reduced DA neurons degeneration in a C. elegans model that over-expresses human α-Syn exclusively in these cells, increasing significantly the proportion of animals that keep > 50 % of their anterior part DA neurons intact.

In conclusion, ZPD-2 properties make this molecule a promising hit for the sake of developing leads able to tackle α-Syn aggregation and seeds propagation in PD and, potentially, other synucleinopathies.

Materials and methods

Protein purification

Protein expression and purification of WT α-Syn and its variants (H50Q and A30P) were carried out as previously described [26] and the resulting purified protein was lyophilised and kept at -80oC until its use.

In vitro aggregation of α-Syn

α-Syn was resuspended in sterile PBS and filtered through 0,22 μm membranes to remove small aggregates. Aggregation was performed in a sealed 96-well plate, containing 70 μM α-Syn (WT, A30P or H50Q), 40 μM Th-T in PBS 1X, a 1/8" diameter Teflon polyball (Polysciences Europe GmbH, Eppelheim, Germany) and 100 μM ZPD-2 or DMSO (in control samples) in a total volume of 150 μL per well. The plate was incubated at 100 rpm and 37 oC after having been fixed in an orbital culture shaker Max-Q 4000 (ThermoScientific, Waltham, Massachusetts, USA). Measurements of Th-T fluorescence were done every 2 h in a Victor3.0 Multilabel Reader (PerkinElmer, Waltham, Massachusetts, USA), exciting through a 430–450 nm filter and collecting the emission signal with a 480–510 filter. Each assay was done in triplicate. The values of the aggregation kinetics were fitted to the following equation (1) [31]:

$$\text{Equation (1)} \qquad \propto = 1 - \frac{1}{k_b(e^{k_a t} - 1) + 1}$$

where kb and ka constitute the homogeneous nucleation rate constant and the secondary rate constant (fibril elongation and secondary nucleation), respectively [31].

Titration assays were done by applying different ZPD-2 concentrations (200, 150, 100, 75, 50, 25 and 10 μM). Time-dependent assays were developed by adding 100 μM of ZPD-2 at different time points after the beginning of the reaction (4, 8, 12, 16, 20 and 24 h). In all cases a fixed concentration of α-Syn at 70 μM was maintained.

Strains were generated as previously described [32; 33; 34]. Briefly, lyophilised α-Syn was resuspended in PBS 1X and dialysed for 24h in a 1:1000 (v/v) ratio with buffer B (50 mM Tris-HCl pH 7.0), or buffer C (50 mM Tris-HCl pH 7.0 supplemented with 150 mM NaCl). Then, the protein was filtered through 0,22 μm membrane and incubated at 70 μM in presence or absence of 100 μM ZPD-2 in a 96-well plate as described above. For the seeding assays, α-Syn pre-formed fibrils were sonicated for 5 minutes and then

added to the aggregation reaction at ratios of 1 % (v/v) for each condition. The plate was then incubated, and Th-T fluorescence measured as previously indicated.

The soluble fraction was obtained for subsequent quantification by centrifuging 300 μL of aggregated sample at 16900 g for 90 min. The supernatant was then recovered and loaded into a Tricine-SDS-PAGE gel. Gels were stained with Blue safe. Finally, the density of the α-Syn bands was calculated using Quantity One software (Bio-Rad, Hercules, California; USA). Experiments were done at least in triplicate.

Transmission Electron Microscopy (TEM)

End-point α-Syn aggregates incubated for 32 h were collected, diluted 1:10 with PBS 1X and sonicated for 5 min. 5 μL of these sonicated samples were placed rapidly on a carbon-coated copper grid and incubated for 5 min. The grids were dried with a filter paper to withdraw the excess of sample and immediately washed twice with miliQ water. Finally, 5 μL of 2 % (w/v) uranyl acetate were added to the top of the grid and incubated for 2 min. The excess of uranyl acetate was removed with a filter paper and grids were left to air-dry for 10 min. Images were obtained using a Transmission Electron Microscopy Jeol 1400 (Peabody, Massachusetts, USA) operating at an accelerating voltage of 120 kV. A minimum of 30 fields were screened per sample, in order to collect representative images.

Light scattering

End-point α-Syn aggregates were collected, placed into a quartz cuvette and analysed in a Cary Eclipse Fluorescence Spectrophotometer (Agilent, Santa Clara, California, USA). The sample was excited at 300 nm and the subsequent scattering at 90° monitored between 280 and 320 nm.

Protein misfolding cyclic amplification (PMCA)

The PMCA assay was carried out as previously described [35]. Briefly, α-Syn was resuspended to a final concentration of 90 μM in Conversion Buffer (PBS 1X, 1% Triton X-100, 150 mM NaCl), supplemented with Complete Protease Inhibitor Mixture (Roche Applied Science, Penzberg, Germany). 60 μL of this α-Syn solution were added into 200-μL PCR tubes containing 1.0 mm silica beads (Biospec Products, Bartlesville, OK, USA). Samples were exposed to 24-hour cycles of 30 s sonication and 30 min incubation at 37 oC, using a Misonix 4000 sonicator, at 70 % power. After every 24 h-cycle, 1 μL of the incubated sample was added to a new PCR-tube containing fresh α-Syn. This process was

repeated for 5 days. In the case of treated samples, ZPD-2 was added in each cycle to the fresh non-sonicated sample to a final concentration of 128 μM, which corresponds to the 0.7:1 α-Syn:ZPD-2 ratio of the previous set of aggregation kinetics assays. Untreated samples were prepared adding the same concentration of DMSO (0.26%) present in the treated mixtures. All the reactions were made in triplicate.

At the end of each cycle, 10 μL of the incubated samples were diluted 1:10 with 90 μL of PBS 1X, 40 μM Th-T. Th-T fluorescence was measured in a Cary Eclipse Fluorescence Spectrophotometer (Agilent, Santa Clara, California, USA), exciting at 445 nm and collecting the emission signal between 460 and 600 nm.

Proteinase K Digestion

For protein digestion, 6 μL of Proteinase K (5 μg/mL final concentration) were added to 18 μL of PMCA aggregated samples and incubated for 30 min at 37 oC. After the incubation, 8 μL of loading buffer containing 1% β-mercaptoethanol were added and the enzyme was thermally inactivated at 95 ºC for 10 min. Finally, 7 μL of the incubated and stained samples were loaded into a Tricine-SDS-PAGE gel together with unstained Protein Standard markers (ThermoFisher Scientific, Waltham, Massachusetts, USA). Gels were stained with Blue safe.

Nuclear Magnetic Resonance (NMR)

Expression of 15N-labelled human WT α-Syn was carried out in E. coli BL21 DE3 strain. First, cells were grown in LB medium until an OD600 of 0.6. The culture was then centrifuged at 3000 rpm for 15 min and the pellets collected and resuspended in 1L minimal medium, composed of: 768 mL of miliQ water with 1 mL of ampicillin 100 mg/mL, 100 μL CaCl2 1M, 2 mL MgSO4 2 M, 20 mL glucose 20%, 10 mL vitamins 100x (Sigma-Aldrich, Darmstadt, Germany), 200 mL salts M9 and 1 g 15NH4 (Cambridge Isotope Laboratories, Inc., Tewksbury, Massachusetts, USA). Cells were incubated for 1 h at 37ºC and 250 rpm. After that, protein expression was induced for 4 h with 1 mM IPTG. Protein was purified as previously described [26].

1H-15N HSQC spectra were obtained at 20 oC on a Bruker 600 MHz NMR spectrometer equipped with a cryoprobe. in a mixture containing 70 μM 15N-labeled α-Syn, PBS buffer (pH 7.4), 2.5% d6-DMSO and 10 % D2O in the absence or in the presence of 100 μM ZPD-2.

Toxicity assays

Neuroblastoma cells were incubated 24 h in DMEM medium in a 96 wells-plate before the addition of different concentrations of ZPD-2 (from 1 μM to 1 mM). Cells were incubated for 48 h at 37 oC and PrestoBlue® reagent (ThermoFisher Scientific, Waltham, Massachusetts, USA) was added to analyse cell death. Treated and untreated cells were incubated with PrestoBlue® for 10 min at 37 oC. Finally, fluorescence emission was measured by exciting at 560 nm and collecting at 590 nm.

Caenorhabditis elegans assays

Maintenance

Animals synchronization was carried out by bleaching and overnight hatching in M9 (3 g/L KH2PO4, 6 g/L Na2HPO4, 5 g/L NaCl, 1 M MgSO4) buffer. Thus, nematodes were cultured at 20 oC on growth media plates (NGM) containing 1 mM CaCl2, 1 mM MgSO4, 5 μg/mL cholesterol, 250 M KH2PO4 pH 6.0, 17 g/L Agar, 3 g/L NaCl. Plates were previously seeded with E. coli OP50 strain. Nematodes were maintained using standard protocols [36].

Strains

Strain NL5901, unc-119(ed3) III; pkIs2386 [Punc-54::α-SYN::YFP; unc-119(+)] was obtained from the C. elegans Genetic Center (CGC). For the α-Syn induced dopaminergic degeneration analysis, strain UA196 [30], gifted generously by the laboratory of Dr Guy Caldwell (Department of Biological Science, The University of Alabama, Tuscaloosa, USA), was used; (sid-1(pk3321); baIn33 [Pdat-1::sid-1, Pmyo-2::mCherry]; baIn11 [Pdat-1::α-SYN; Pdat-1::GFP]). In the main text, this strain was named Pdat-1::GFP; Pdat-1::α-SYN.

ZPD-2 administration

After cooled, the autoclaved NGM agar medium (1 mM CaCl2, 1 mM MgSO4, 5 μg/mL cholesterol, 250 M KH2PO4 pH 6.0, 17 g/L Agar, 3 g/L NaCl) was enriched with 100 μM of a stock of ZPD-2 in 0.2% DMSO to a final concentration of 10 μM. After 2 days, plates were seeded with 250 μL of E. coli OP50 with 10 μM of ZPD-2. Nematodes were placed on the plates at larval stages L4 and exposed either to ZPD-2 or DMSO (controls) for 7 days. Daily transfer was done to avoid cross progeny.

Aggregate quantification

The number of cellular inclusions was quantified as previously described [29; 37]. Briefly, NL5901 (Punc-54::α-SYN::YFP) worms were age-synchronized and left overnight to hatch. Nematodes in phase L1 were cultured and grown into individual NGM plates seeded with E. coli OP50. When animals reached L4 developmental stage, they were transferred onto either ZPD-2 treated plates or DMSO treated plates (negative control). Every day, animals were transferred into a new plate to avoid cross contamination. At stage L4+7, the aggregates in the anterior part of every single animal were counted. For each experiment, thirty 7-days old nematodes per treatment were analysed using a Nikon Eclipse E800 epifluorescence microscope equipped with an Endow GFP HYQ filter cube (Chroma Technology Corp, Bellows Falls, Vermont USA) and each experiment was carried out in triplicate. Inclusions could be described as discrete bright structures, with edges distinguishable from surrounding fluorescence. ImageJ software was used for measuring the number of cellular aggregates considering the area dimensions. For the quantification of α-syn aggregates in C. elegans one single image was taken from each animal. Every image contained among 30-45 stacks (1 μm) that allowed to detect aggregates at different animal positions. At least 30 animals were imaged for each assayed condition.

C. elegans lifespan analysis

L4-stage synchronized C. elegans were exposed to 10 uM of ZPD-2 or DMSO (controls) during lifespan analysis. The worms were classified as alive, dead, or censored every two days by determining their movement and response to nose and tail tap. The numbers of alive and dead worms were recorded until all worms perished. The data were plotted as a Kaplan-Meier survival curve and groups compared using a Wilcoxon-test.

C. elegans neurodegeneration assays

Worms were analysed for α-Syn–induced dopaminergic neurodegeneration as described previously [30]. Briefly, 20-30 L4 staged animals were transferred to ZPD-2 - NGM plates and make them grow up to seven days (L4 + 7 days of development) after which the dopaminergic cell death induced by the over-expression of α-Syn was analysed by fluorescence. Plates containing only 0.2 % DMSO, without ZPD-2, were used as control. Worms were transferred daily to avoid cross contamination.

The six anterior DA neurons (four CEP and two ADE DA neurons) were scored for neurodegeneration according to previously described criteria [30; 38]. Worms were considered normal when all six anterior DA neurons (four CEP, cephalic, and two ADE, anterior deirid) were present without any visible signs of degeneration. If a worm displayed degeneration in at least one of the six neurons, it was scored as exhibiting degeneration. For each independent experiment, thirty worms of each treatment were examined under a Nikon Eclipse E800 epifluorescence microscope equipped with an Endow GFP HYQ filter cube (Chroma Technology Corp, Bellows Falls, Vermont USA).

Microscopy and imaging

Animals were placed in a 1 mM solution of sodium azide and mounted with a coverslip on a 4% agarose pad. Animals were visualized with a Nikon Eclipse E800 epifluorescence microscope. The system acquires a series of frames at specific Z-axis position (focal plane) using a Z-axis motor device. Animals were examined at $100\times$ magnification to examine α-Syn induced DA cell death and at $40\times$ to examine α-Syn apparent aggregate.

Statistical analysis

All graphs were generated with GraphPad Prism 6.0 software (GraphPad Software Inc, La Jolla, California, USA). Data were analysed by two-way ANOVA Tukey test using SPSS software version 20.0 (IBM Analytics, Armonk, NY, United States) and t-test using GraphPad software version 6.0 (GraphPad Software Inc, La Jolla, California, USA). All data are shown as means and standard error of mean (SEM). $p < 0.05$ was considered statistically significant. In the graphs *, ** and *** indicate $p < 0.05$, $p < 0.01$ and $p < 0.001$, respectively.

Supplementary Information for:

ZPD-2, a small compound that inhibits α-synuclein amyloid aggregation and its seeded polymerization

Samuel Peña-Díaz[a,b,‡], Jordi Pujols[a,b,‡], María Conde-Giménez[c], Anita Čarija[a,b], Esther Dalfo[e,f], Jesús García[d], Susanna Navarro[a,b], Francisca Pinheiro[a,b], Jaime Santos[a,b], Xavier Salvatella[d], Javier Sancho[c] and Salvador Ventura[a,b,*]

[a] Institut de Biotecnologia i Biomedicina. Universitat Autonoma de Barcelona, 08193-Bellaterra, Spain.

[b] Departament de Bioquimica i Biologia Molecular. Universitat Autonoma de Barcelona, 08193-Bellaterra, Spain.

[c] Department of Biochemistry and Molecular and Cell Biology, Institute for Biocomputation and Physics of Complex Systems (BIFI), University of Zaragoza, 50018 Zaragoza, Spain

[d] Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Baldiri Reixac 10, 08028, Barcelona, Spain

[e] Medicine, M2, Universitat Autonoma de Barcelona (UAB), Bellaterra Campus, Cerdanyola del Valles, Barcelona, Spain. ORCID: 0000-0003-4677-8515

[f] Faculty of Medicine. University of Vic-Central University of Catalonia (UVic-UCC)

[*] Correspondence: salvador.ventura@uab.cat

[‡] These authors contributed equally to this work

Figure S1. α-synuclein soluble fraction at the end of the aggregation. (A) Soluble fraction of α-Syn when incubated in absence (black) or presence (green) of ZPD-2 in PBS solution. (B) Soluble fraction of strain B and C at final point of the aggregation when incubated in presence (green) or absence (black) of ZPD-2.
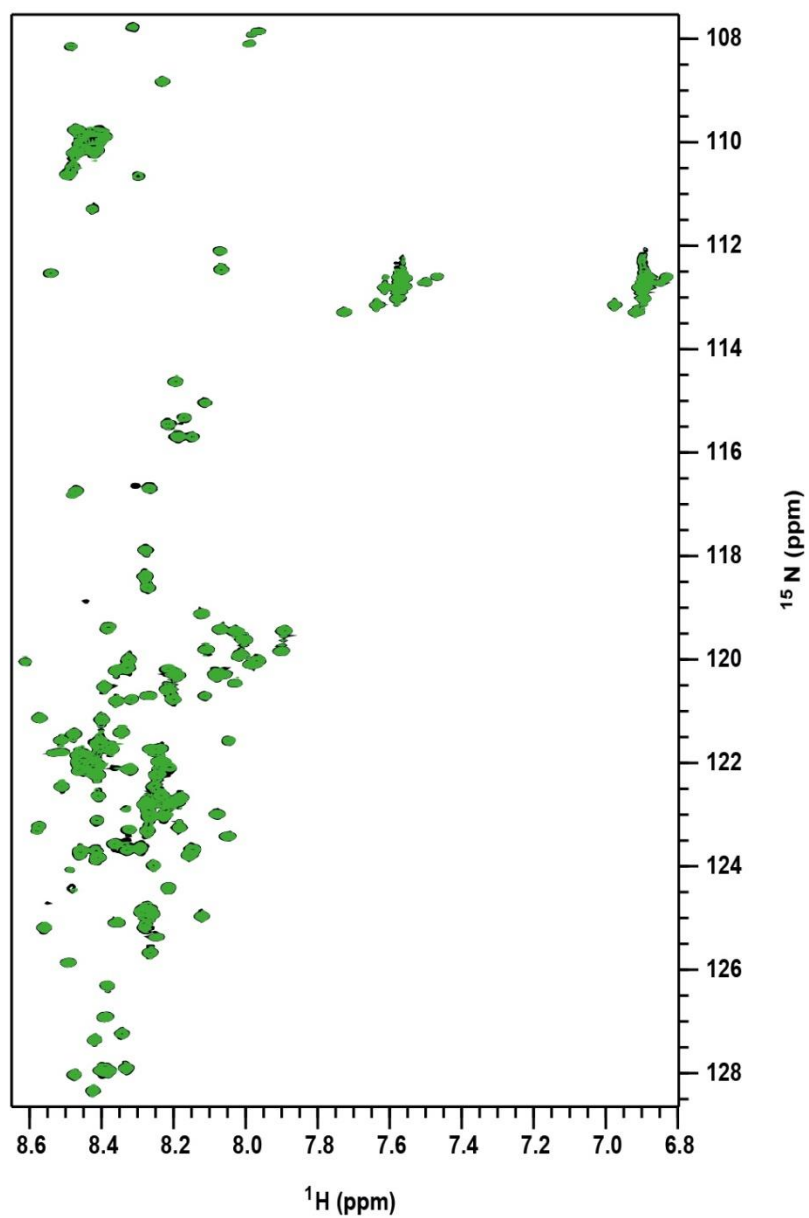
Figure S2. Lack of interaction between monomeric α-synuclein and ZPD-2 assessed by NMR.
Superposition of the 1H-15N HSQC NMR spectra of 15N-labeled α-Syn (70 μM) in absence (black) and presence (green) of 100 μM of ZPD-2.
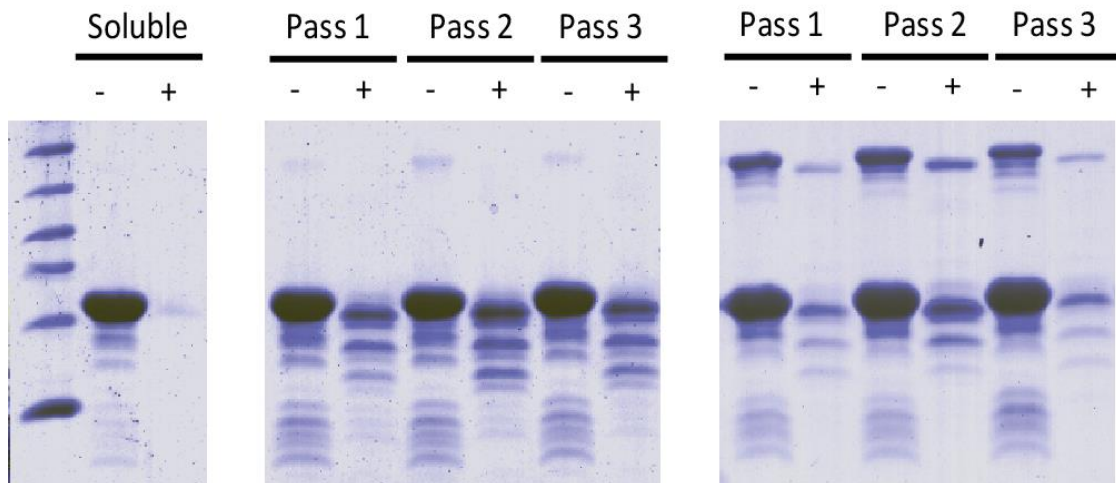
Figure S3. PMCA assay at early stages. Tricine-SDS-PAGE gels of untreated (middle) and ZPD-2 treated (right) PMCA samples before (-) and after (+) being digested with proteinase K. Soluble α-Syn and PMCA steps 1 to 3 are shown.
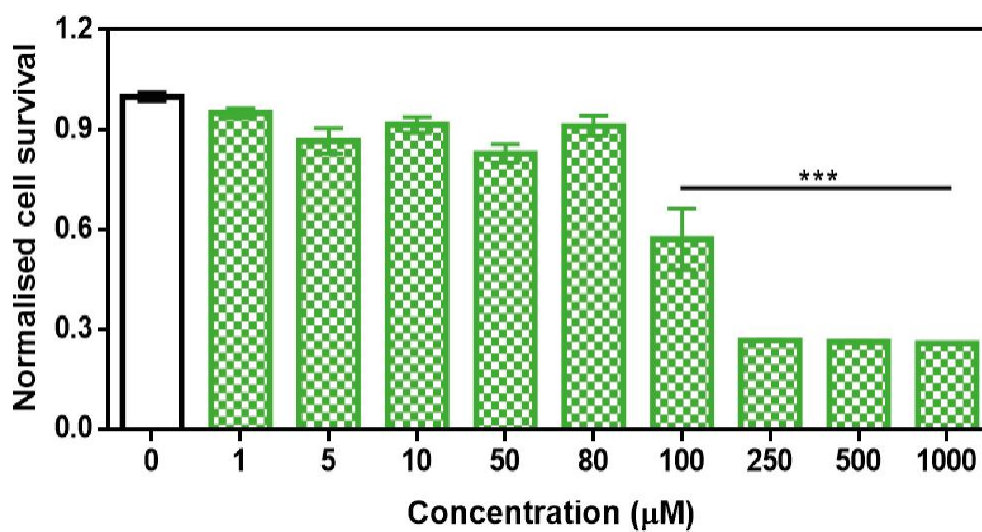
Figure S4. Toxicity assays. Analysis of neuronal cells culture survival in presence of different concentration of ZPD-2. Survival is potted as normalized means. Error bars are shown as standard error of means values, where *p < 0.001* was indicated by ***.
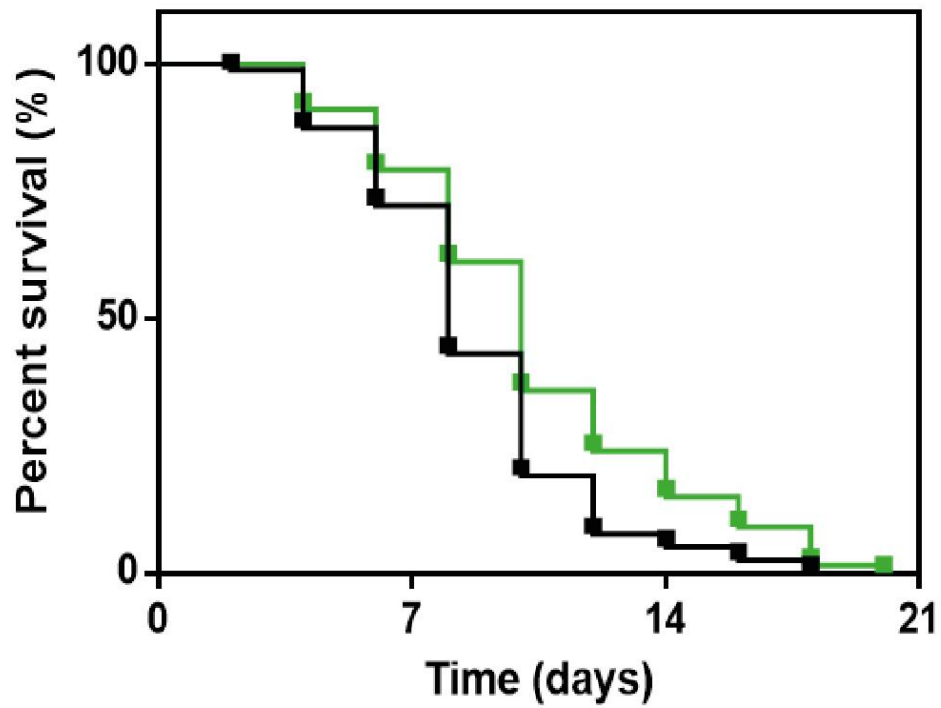
Figure S5. C. elegans lifespan analysis. Effect of ZPD-2 treatment (green) on the survival of PD model animals, in comparison with untreated PD worms (black). The data represent the survival ratio (approximately 60–80 animals per group).
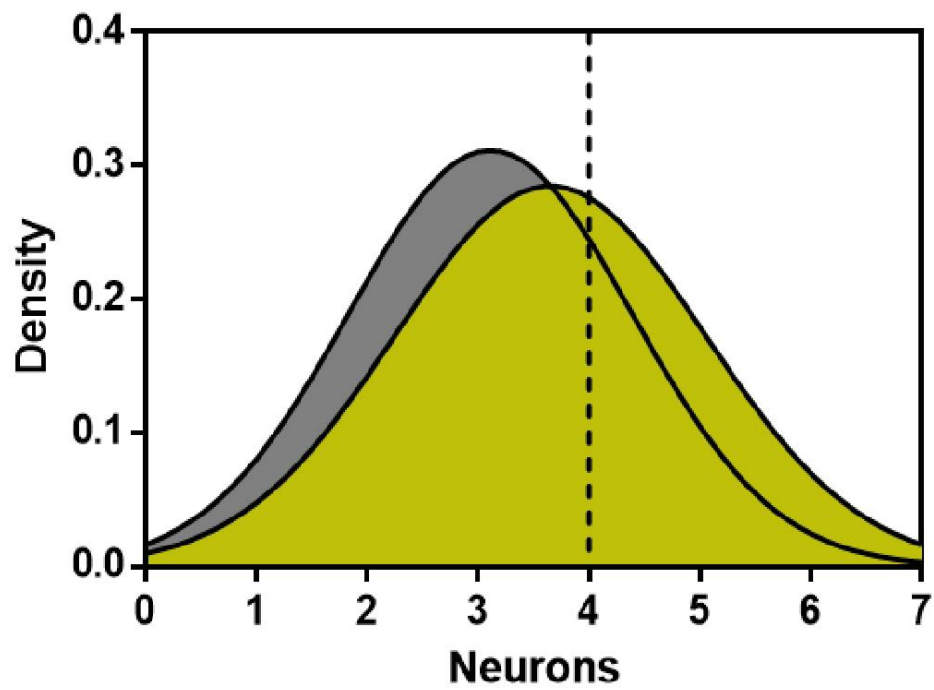
Figure S6. Distribution of functional neurons in the *C. elegans* dopaminergic model. Normal distribution of the remaining functional dopaminergic (DA) neurons in transgenic animals when treated with ZPD-2 (green) or vehicle (grey). The dashed line delimits animals having four or more functional DA neurons.

Work 7

Inhibition of α-Synuclein aggregation

and mature fibril disassembling

with a minimalistic compound, ZPDm

Inhibition of α-Synuclein aggregation and mature fibril disassembling with a minimalistic compound, ZPDm

Samuel Peña-Díaz[a,b,‡], Jordi Pujols[a,b,‡], Francisca Pinheiro[a,b], Jaime Santos[a,b], Irantzu Pallarés[a,b], Susanna Navarro[a,b], María Conde-Jimenez[c], Jesús García[d], Xavier Salvatella[d], Esther Dalfó[e,f], Javier Sancho[c], and Salvador Ventura[a,b,*]

[a] Institut de Biotecnologia i Biomedicina. Universitat Autonoma de Barcelona, 08193-Bellaterra, Spain.

[b] Departament de Bioquimica i Biologia Molecular. Universitat Autonoma de Barcelona, 08193-Bellaterra, Spain.

[c] Department of Biochemistry and Molecular and Cell Biology, Institute for Biocomputation and Physics of Complex Systems (BIFI), University of Zaragoza, 50018 Zaragoza, Spain

[d] Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Baldiri Reixac 10, 08028, Barcelona, Spain

[e] Medicine, M2, Universitat Autonoma de Barcelona (UAB), Bellaterra Campus, Cerdanyola del Valles, Barcelona, Spain. ORCID: 0000-0003-4677-8515

[f] Faculty of Medicine. University of Vic-Central University of Catalonia (UVic-UCC)

[*] Correspondence: salvador.ventura@uab.cat
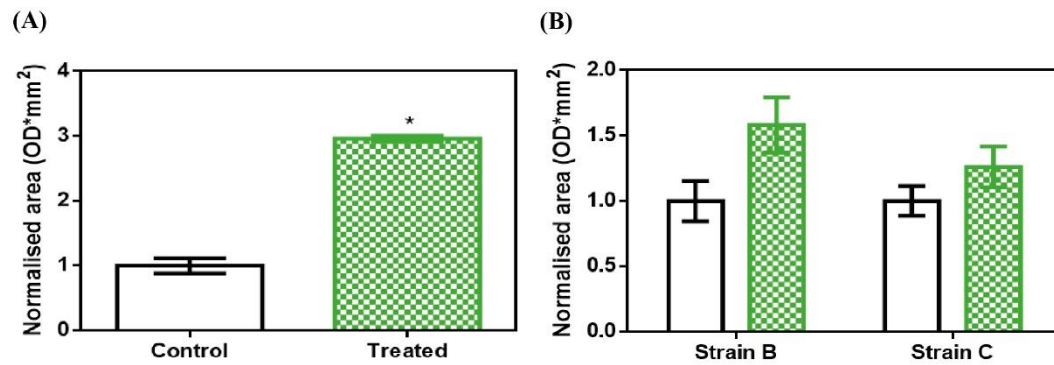
[‡] These authors contributed equally to this work

Abstract

Synucleinopathies are a group of disorders characterized by the accumulation of α-Synuclein amyloid inclusions in the brain. Preventing α-Synuclein aggregation is challenging because of the disordered nature of the protein and the stochastic nature of fibrillogenesis, but, at the same time, it is a promising approach for therapeutic intervention in these pathologies. A high-throughput screening initiative allowed us to discover ZPDm, the smallest active molecule in a library of more than 14.000 compounds. Although the ZPDm structure is highly related to that of the previously described ZPD-2 aggregation inhibitor, we show here that their mechanisms of action are entirely different. ZPDm inhibits the aggregation of wild-type, A30P, and H50Q α-Synuclein variants *in vitro* and interferes with α-Synuclein seeded aggregation in protein misfolding cyclic amplification assays. However, ZPDm distinctive feature is its strong potency to dismantle preformed α-Synuclein amyloid fibrils. Studies in a *Caenorhabditis elegans* model of Parkinson's Disease, prove that these *in vitro* properties are translated into a significant reduction in the accumulation of α-Synuclein inclusions in ZPDm treated animals. Together with previous data, the present work illustrates how different chemical groups on top of a common molecular scaffold can result in divergent but complementary anti-amyloid activities.

Keywords: α-Synuclein, protein aggregation, amyloid inhibitor, Parkinson's Disease, synucleinopathies, small molecules

Introduction

Parkinson's disease (PD) is an incurable disorder that affects around 0.3 % of the population and more than 1 % of people over 60 years of age (4 % over 80 years), being the second most prevalent neurodegenerative disease worldwide. (Nussbaum and Ellis, 2003; Dexter and Jenner, 2013; Kalia and Lang, 2015). Together with Dementia with Lewy Bodies (DLB) and Multiple System Atrophy (MSA), PD is part of a group of human disorders known as synucleinopathies (Spillantini et al., 1998a; Spillantini et al., 1998b; Fanciulli and Wenning, 2015). Intracellular proteinaceous inclusions constitute the main culprit of neuronal damage and disease progression in the synucleinopathies, although the aggregates accumulate in different cell types and affect distinct brain regions depending on the disease (Fellner et al., 2011; Luk et al., 2012). These abnormal protein deposits are mostly composed of aggregated α-synuclein (α-Syn). In the particular case of PD, α-Syn aggregation occurs in the dopaminergic neurons of *substantia nigra pars compacta*. As a consequence, PD suffering patients display reduced dopamine levels, which results in the archetypic motor and non-motor symptoms of the disease (Spillantini et al., 1997). Indeed, single-point mutations and multiplications of the gene that encodes for α-Syn (*SNCA*) (Singleton et al., 2003; Ibanez et al., 2004) have been related to familial cases of PD with early-onset (Polymeropoulos et al., 1997), thus reinforcing the connection between α-Syn and PD.

α-Syn is an intrinsically disordered protein highly expressed in the brain and associated with vesicle trafficking in healthy conditions (Bendor et al., 2013). In pathological situations, α-Syn aggregates into oligomers and amyloid fibrils that compromise cellular homeostasis, exert toxicity and ultimately led to neuronal death (Serpell et al., 2000). Remarkably, diffusible aggregated species can be internalized by healthy neighbouring neurons, where they seed the aggregation of soluble α-Syn molecules, a mechanism that has been compared with the templated conformational conversion occurring in prion diseases (Hansen et al., 2011). As it occurs in prions, α-Syn assemblies can present diverse structural arrangements, forming strains (Li et al., 2018) that differ in their aggregation properties (Bousset et al., 2013), and target distinct brain regions and cell types (Lau et al., 2020).

Many of the current therapeutic approaches for PD aim to reduce the neuronal load of aggregated α-Syn, either by targeting the α-Syn polypeptide directly or through indirect

approaches such as the stimulation of degradation pathways (Spencer et al., 2009; Decressac et al., 2013; Xilouri et al., 2013) and gene silencing (Faustini et al., 2018; Kantor et al., 2018; Lassot et al., 2018; Zharikov et al., 2019). Among them, the identification of small compounds that might act as chemical chaperones blocking the aggregation and propagation of α-Syn or, in the best-case scenario, dismantling α-Syn mature aggregates into non-toxic species is receiving increasing attention. However, the disordered nature of α-Syn, together with the multiplicity of conformationally different species that populate the aggregation process, imposes significant difficulties for the rational design of effective α-Syn binders. For this reason, high-throughput screening (HTS) of large chemical libraries has become a significant focus of research in the hunt for disease-modifying lead compounds (Silva B, 2011).

To analyse the inhibitory potential of a chemical library with more than 14000 chemically diverse structures, we optimized a robust HTS screening protocol (Pujols et al., 2017) based on Thioflavin-T (Th-T) fluorescence, light-scattering measurements and Transmission Electron Microscopy (TEM). This pipeline was used to discover and characterize small molecules that act as potent inhibitors of α-Syn amyloid formation, such as SynuClean-D (SC-D) (Pujols et al., 2018) and ZPD-2 (Peña-Díaz et al., 2019). In the present work, we present and characterize ZPDm (**Figure 1**), a novel molecule identified in the primary HTS screen. ZPDm was the smallest compound in the library displaying a significant inhibitory potency at a substoichiometric concentration. ZPDm prevents the aggregation of *wild type* α-Syn and the familial A30P and H50Q variants. The molecule acts preferentially at the late stages of the polymerization reaction, suggesting that it targets preferentially ordered aggregates. Indeed, ZPDm is highly effective at disaggregating the mature α-Syn fibrils of different strains. To the best of our knowledge, it constitutes the minimal synthetic molecule that *co*njugates inhibitory and α-Syn fibril disrupting activity in the same chemical scaffold. These *in vitro* activities are translated into the ability to reduce significantly α-Syn aggregation in a well-established *Caenorhabditis elegans* (*C. elegans*) model of PD.

Results

ZPDm inhibits α-Synuclein aggregation in vitro

ZPDm is a molecule that was initially identified as a positive hit in the HTS performed by our lab back in 2017 on top of the HitFinderTM chemical library from Maybridge (Pujols et al., 2017). Although they were found independently, structurally, ZPDm is a minimalistic version of ZPD-2 (Peña-Díaz et al., 2019) with a reduced MW, LogP, and TPSA, which, in principle, would increase its drug-likeness (Fig. 1 and Table 1).
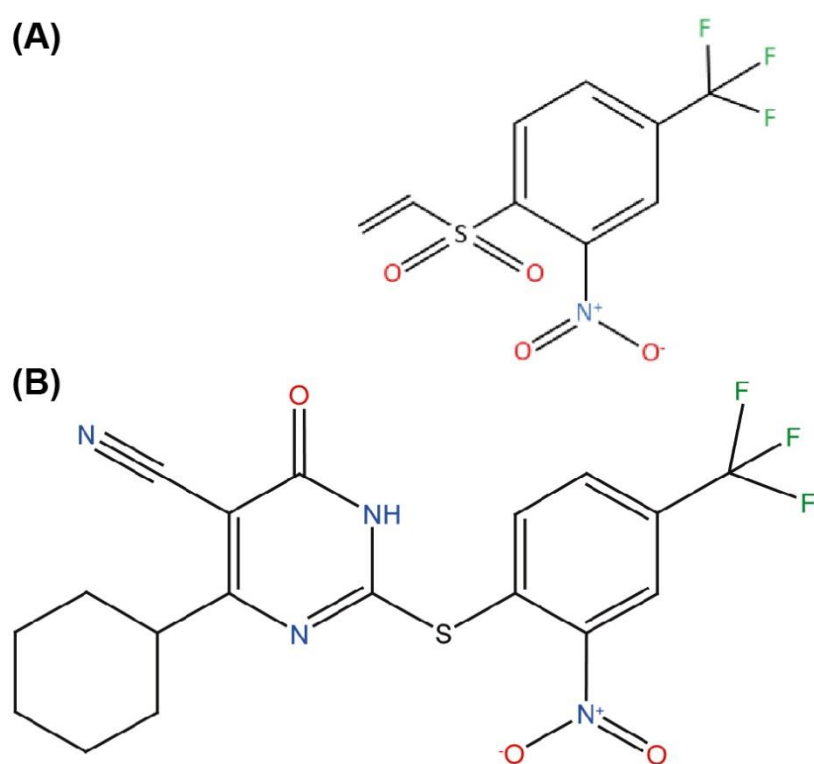
Figure 1. ZPDm, a ZPD-2 minimal structure. Chemical structures of (A) 2-nitro-4-(trifluoromethyl)phenyl vinyl sulfone, named ZPDm, and (B) ZPD-2, both constituted by a hydrophobic core formed by aromatic rings and polar projections.

Still, ZPDm implements the generic physicochemical properties common to most of the published inhibitors of α-Syn aggregation, namely, a planar hydrophobic core formed by aromatic rings that interact with apolar exposed regions in α-syn assemblies. This core is frequently coated with polar projections that interfere with hydrophobic packing and disrupt intermolecular hydrogen bonds; they difficult elongation and, eventually, might promote fibril disassembly (Pujols et al., 2020). A benzene ring constitutes the

228

hydrophobic moiety of ZPDm, connected to three polar groups – vinyl sulfone, nitro, and trifluoromethyl – in carbon positions 1, 2, and 4 respectively. Remarkably, the ZPD-2 chemical structure displays the same functional groups in carbons 1, 2, and 4 of the primary phenyl ring. However, the vinyl sulfone is substituted by a sulfide group that extends the molecule to incorporate additional aromatic moieties, which we previously assumed to be critical for its activity (Pena-Diaz et al., 2020). Thus, ZPDm can be considered as a building block for the synthesis of the more complex ZPD-2 molecule. Considering these structural differences, we performed a set of orthogonal experiments to contrast if, as ZPD-2, ZPDm might turn to be a lead compound with significant α-Syn anti-aggregational activity.

Table 1. SwissADME predicted properties of ZPDm and ZPD-2.

	ZPDm	ZPD-2
Molecular weight (g/mol)	281,21	424,4
Heavy atoms	18	29
Aromatic heavy atoms	6	12
H-bond acceptors	7	8
H-bond donors	0	1
TPSA (Å^2)	88,34	140,66
Log P$_{O/W}$	2,42	3,9
Solubility (mg/mL)	9,28E-02	1,94E-03
GI absorption	High	Low
Drug-like (Lipinski)	0 violations	0 violations
Drug-like (Veber)	0 violations	1 violations
Drug-like (Egan)	0 violations	2 violations
Leadlikeness	0 violations	2 violations

The incubation of 70 µM α-Syn with 100 µM ZPDm inhibited the protein aggregation, decreasing the final Th-T fluorescence signal by 60 % compared to control untreated samples (Fig. 2A). Fitting of the kinetic data to a typical sigmoidal nucleation-polymerization reaction revealed that ZPDm diminishes the primary nucleation rate constant by eightfold (kb= 0,0034), relative to the control reaction (kb= 0,0275), at the expenses of a slightly higher autocatalytic rate constant, with ka of 0.449 h-1 and 0.323 h-1 for ZPDm treated and untreated samples, respectively. This results in t0 and t1/2 being increased by 4 h and 2 h, respectively, in the compound's presence. Light-scattering measurements at 300 nm and 340 nm at the end of the reaction reported a decrease of 81 % and 78 % in the dispersed light in the presence of ZPDm, respectively, consistent with a reduction of the total aggregated material. The visual inspection of α-Syn samples by TEM corroborated a reduction in the number of amyloid fibrils per field in the presence of ZPDm (Fig. 2F), compared to untreated samples (Fig. 2E). A titration assay in which we incubated 70 µM α-Syn in the presence of decreasing amounts of ZPDm indicated a dose-dependent inhibition, with a statistically significant activity at a substoichiometric concentration of 50 µM, at which ZPDm still reduces the Th-T signal at the end of the reaction by 35 % (Fig. 2D).

We further examined if ZPDm was able to prevent the aggregation of two mutants of α-Syn, H50Q, and A30P, which have been associated with familial PD (Kruger et al., 1998; Appel-Cresswell et al., 2013). The incubation of these α-Syn variants with ZPDm reduced Th-T fluorescence at the end of the reaction by 81% and 71 % for H50Q and A30P, respectively (Fig. 2C).

ZPDm, ZPD-2, and SC-D and other positive hits in the library, share a common property, the presence of a trifluoromethyl group connected to an aromatic ring. We hypothesized that perhaps we were in front of the minimal inhibitory unit, which might be very useful for future Structure Activity Relationship (SAR) studies. Therefore, we synthesized the (trifluoromethyl)benzene moiety and assessed its anti-aggregational potential (Supplementary Figure S1). Both kinetic data using Th-T and light scattering measurements converged to indicate that this molecule is devoid of any activity, suggesting that it might be necessary, but not sufficient to endorse ZPDm with the above-described anti-aggregation properties.
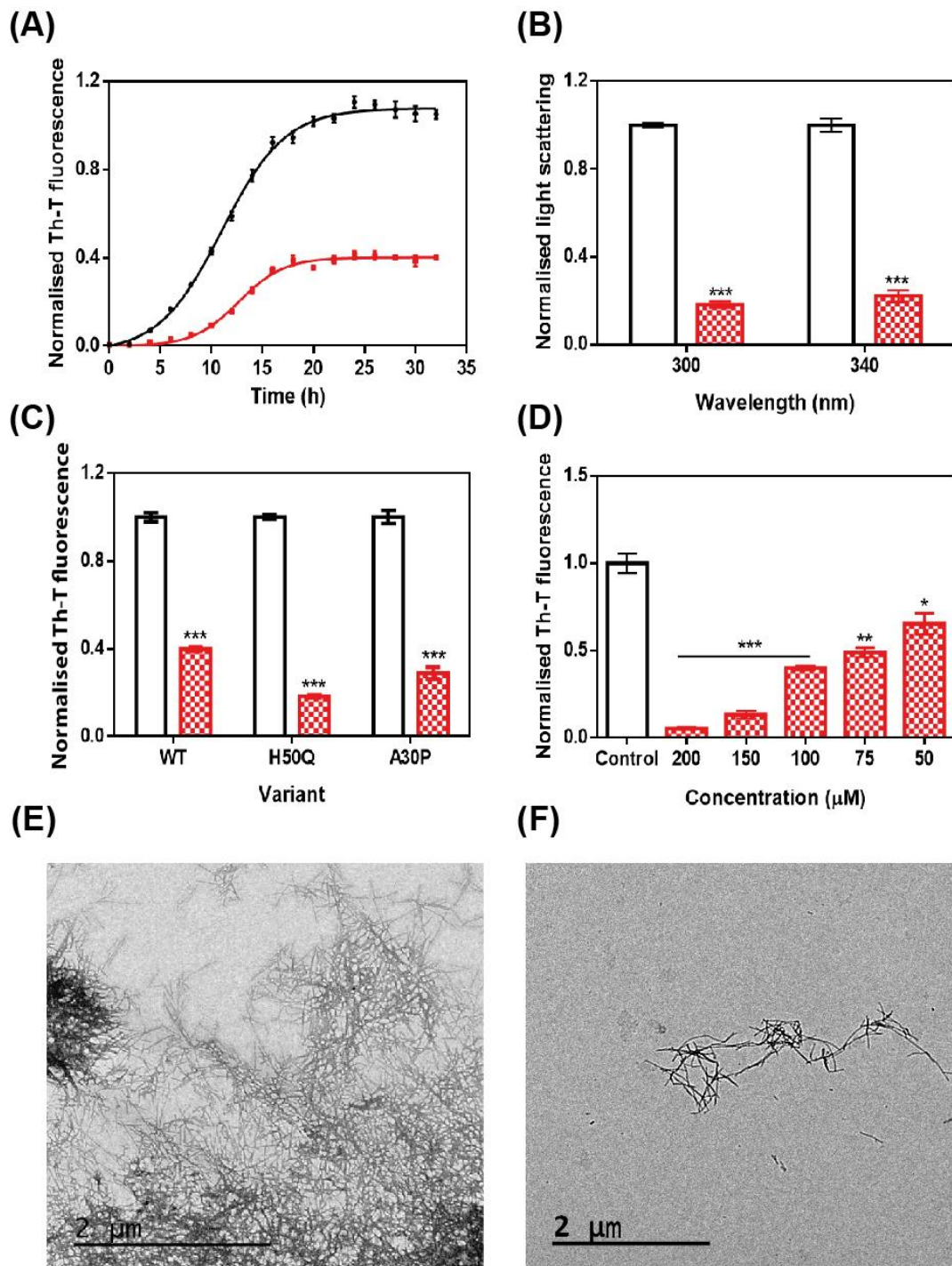
Figure 2. In vitro analysis of the capacity of ZPDm to inhibit α-Syn aggregation. (A) α-Syn aggregation kinetics in the absence (black) and presence (red) of 100 μM of ZPDm followed by Th-T fluorescence. (B) Light-scattering measurements at 300 and 340 nm, in the absence (white) and presence (red) of ZPDm. (C) H50Q and A30P α-Syn variants aggregation in the absence (white) and presence (blue) of ZPDm. (D) Inhibition of α-Syn aggregation with different concentrations of ZPDm. (E and F) Representative TEM images in the absence (E) and presence (F) of ZPDm. Th-T fluorescence is plotted as normalized means. Final points were obtained at 48 h. Error bars are represented as SE of mean values; * p <0,05, ** p < 0.01 and *** p < 0.001. ZPDm prevents the aggregation of WT, A30P and H50Q α-Syn variants in vitro, even at substoichiometric ratios.

ZPDm Prevents α-Syn Aggregation in Protein Misfolding Cyclic Amplification Assays

We used protein-misfolding cyclic amplification (PMCA) to test the inhibitory capacity of ZPDm under continuous seeding conditions. Based on the nucleation-dependent polymerization model for prion replication, PMCA is a technique that forces aggregation to happen by seeding soluble α-Syn with preformed fibrils (Jung et al., 2017). After the first round of fibril elongation, aggregates are sonicated and used as seeds for the second cycle of PMCA. A representative sample from each cycle is then treated with proteinase-K (PK) and analyzed by SDS-PAGE, to evidence fibril formation, since in contrast to soluble α-Syn, the fibrils are significantly resistant to proteolysis. Using this protocol, PK-resistant species could be observed already in the 1st cycle of PMCA in untreated samples, with a maximum of PK resistance at the 4th cycle (Fig. 3A). Th-T fluorescence measurements of the same samples indicated that this protection correlates with the presence of amyloid-like assemblies (Fig. 3C). In contrast, in the presence of ZPDm, PK-resistant species are absent until the 3rd cycle, and they never rich the levels of the control samples, although certain adaptation of misfolded α-Syn to ZPDm seems to occur in cycles 4th and 5th (Fig. 3B). The Th-T fluorescence signal is negligible in the two first PMCA cycles and significantly lower than that of control samples at any considered cycle (Fig. 3C).
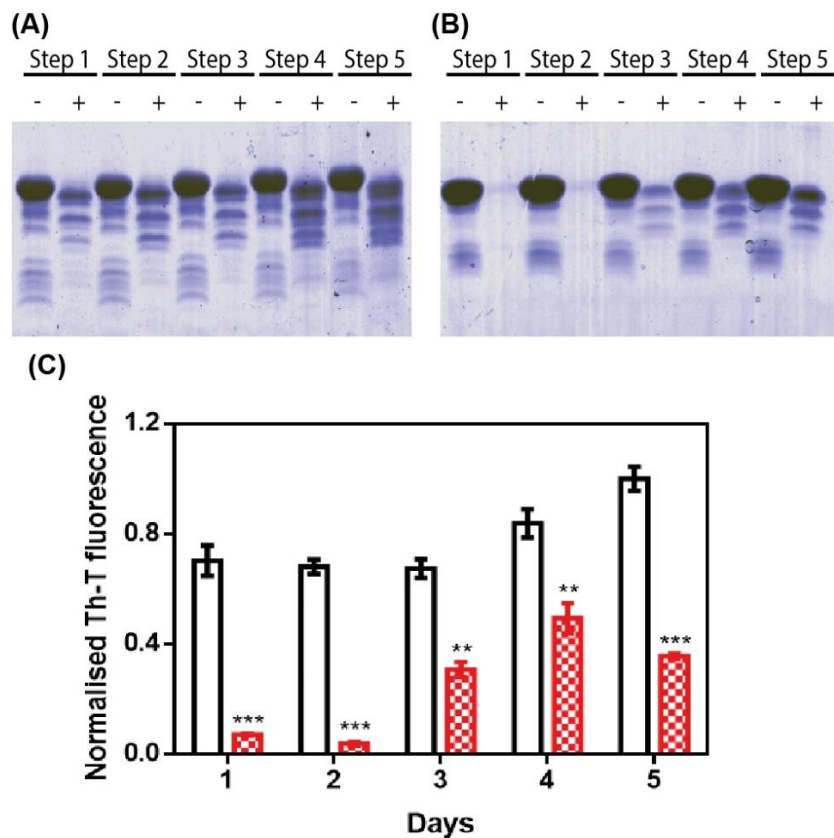
Figure 3. Inhibitory activity of ZPDm in PMCA assays. (A and B) Bis/Tris SDS-PAGE gels of PMCA samples in the absence (A) and presence (B) of ZPDm, before (-) and after (+) PK digestion. (C) Th-T fluorescence of different PMCA cycles in samples treated (red) and untreated (white) with ZPDm. Th-T fluorescence is plotted as normalized means. Error bars are represented as SE of mean values. ZPDm anti-aggregation activity results in delayed formation of PK-resistant and Th-T positive amyloid structures.

ZPDm exhibits amyloid disaggregation activity in vitro

ZPD-2 does not interact with soluble and monomeric α-Syn and, therefore, is not expected to interfere with the protein's functional state. Nuclear Magnetic Resonance 1H-15N HSQC spectra of N15 labeled α-Syn in the presence and absence of ZPDm, indicates that this is also the case for this smaller molecule since we could not identify any perturbations in chemical shifts or peak intensities in the spectra (Supplementary Figure S2)

To address the time window in which ZPDm remains active, we set up an experiment in which a constant amount of ZPDm was added to different aggregation reactions at different time intervals after the reaction has begun (Fig. 4A). To our surprise, the respective Th-T signals indicated that the anti-amyloid activity increased as the reaction progressed, which is in stark contrast with the behaviour of ZPD-2, which was mostly active when added at the early stages of the reaction and inactive when added at the

plateau phase (Pena-Diaz et al., 2019). The time-dependent activity profile of ZPDm can only be explained if this compound recognizes the Th-T positive aggregated species and exerts an intense fibril disruption activity. To confirm this extent, α-Syn mature fibrils were incubated for 24 h with ZPDm. The Th-T fluorescence analysis revealed that treated samples suffered a signal reduction of 74 % (Fig. 4B). This data was supported by TEM images, which illustrated the disruption of large fibrillar clusters into shorter fibrils or amorphous aggregates (Fig. 4C-D).
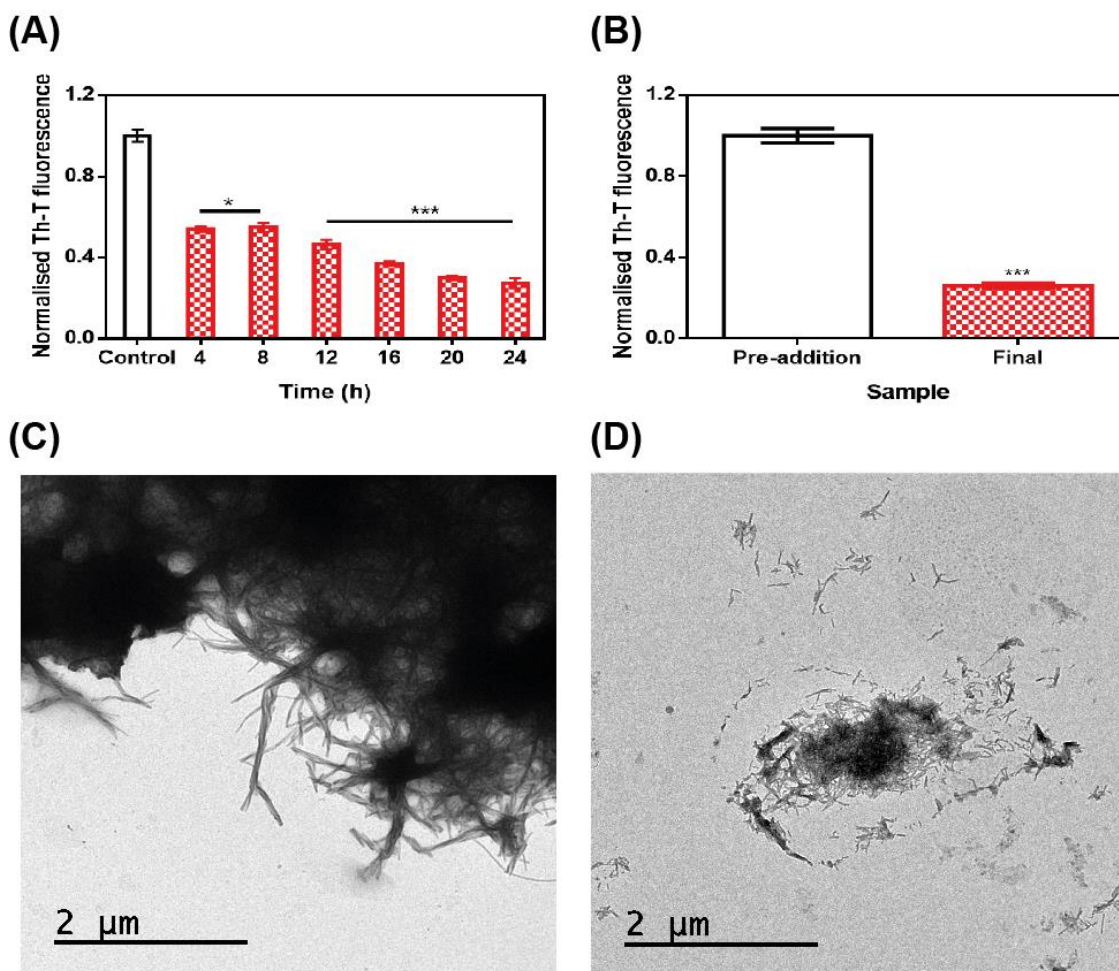
Figure 4. Disaggregational activity of ZPDm. (A) Th-T fluorescence of □-Syn end-point aggregates after the addition of ZPDm at different time points during the aggregation kinetics. (B) Th-T fluorescence assay before and 24 h after the addition of ZPDm to mature α-Syn fibrils. (C and D) Representative TEM images in the absence (C) and presence (D) of ZPDm. Th-T fluorescence is plotted as normalized means. Error bars are represented as SE of mean values; * $p < 0.05$ and *** $p < 0.001$. ZPDm inhibitory capacity increases with the reaction progress, indicating that the compound may interact with aggregated structures and disentangle them.

We assessed if this amyloid-disrupting activity was independent of the conformational properties of the mature fibrils, by aggregating α-Syn in 50 mM Tris–HCl pH 7.0 in the

absence or presence of 150 mM NaCl, which generates two different strains, known as strain B and C, respectively (Bousset et al., 2013; Carija et al., 2019). As shown in Figure 4, the addition of ZPDm to the mature fibrils of these strains promoted a significant decrease in the amount of amyloid-like material as monitored both by Th-T fluorescence and TEM 24 h after the addition of the molecule (Fig. 5). These data suggest that in contrast to ZPD-2, insensitive to preformed amyloid fibrils, ZPDm is endorsed with a generic and potent disaggregation activity.
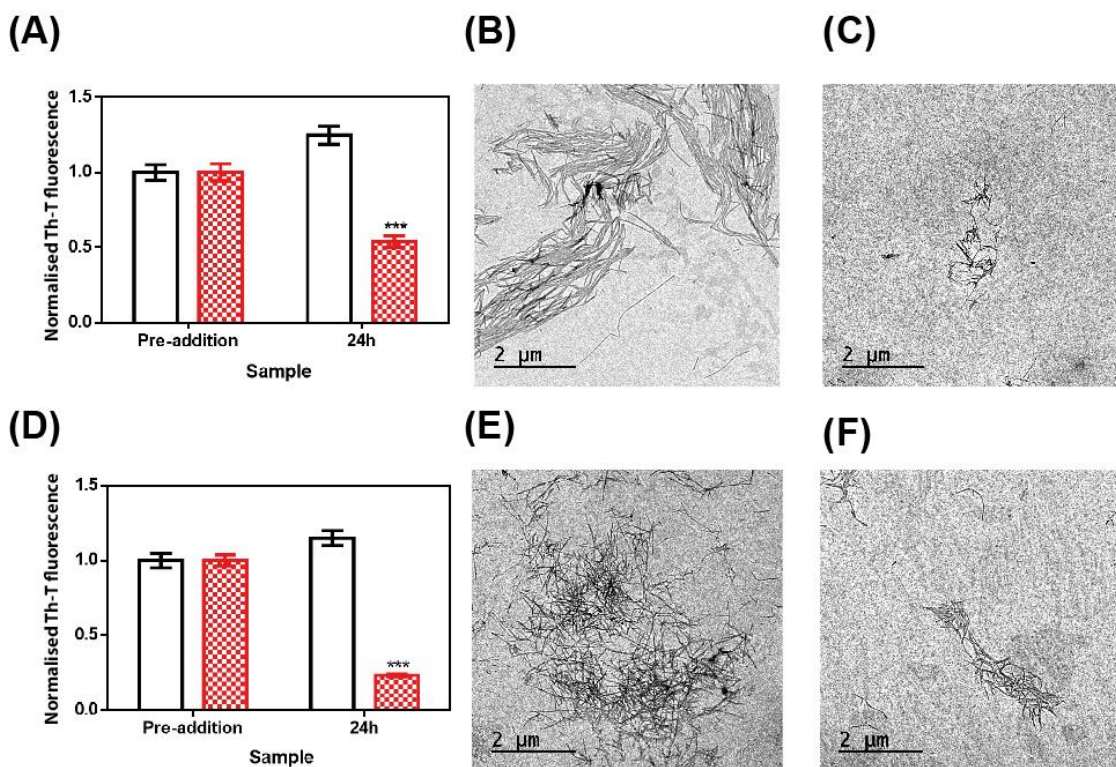
Figure 5. Disaggregational effect of ZPDm in preformed fibrils of two different strains. (A) Strain B aggregates disaggregation in the presence (red) and absence (white) of ZPDm as monitored by Th-T fluorescence. (C and B) Representative TEM images of untreated (B) and ZPDm treated (B) samples. (D) Strain C aggregates disaggregation in presence (red) and absence (white) of ZPDm as monitored by Th-T fluorescence. (E and F) Representative TEM images of untreated (E) and ZPDm treated (E) samples. Data are shown as means, and error bars are shown as the SE of means; *** $p < 0.001$. The disaggregational ability of ZPDm is also observed in two morphologically different α-Syn strains.

ZPDm decreases the formation of α-Syn aggregates in a C. elegans model of PD

We decided to test if the ZPDm in vitro activity can be translated in vivo to a simple animal model of PD. To do so, we employed the well-described strain NL5901 of C. elegans. In this strain, α-Syn is fused to Yellow Fluorescence Protein (YFP) and expressed under the control of the unc-54 promoter, transgene pkIs2386 [Punc-54::α-

SYN::YFP] (Hamamichi et al., 2008; van Ham et al., 2008), (Hamamichi et al., 2008; van Ham et al., 2008), generating protein inclusions in body wall muscle cells. ZPDm was administered in the food at 10 □M final concentration to animals at the L4 stage, and they were examined 7 days later, 9 days after hatching (L4+7). These aged worms are intended to mimic aged PD patients. We used epifluorescence microscopy to visualize the fluorescent aggregates. The images demonstrated that ZPDm reduced the formation of muscular inclusions by 43 % (Fig. 6A), with an average of 20.2 ± 2.13 apparent α-Syn aggregates in treated worms (Fig. 6C) compared with the 35.2 ± 3.04 observed in control samples (Fig. 6B).
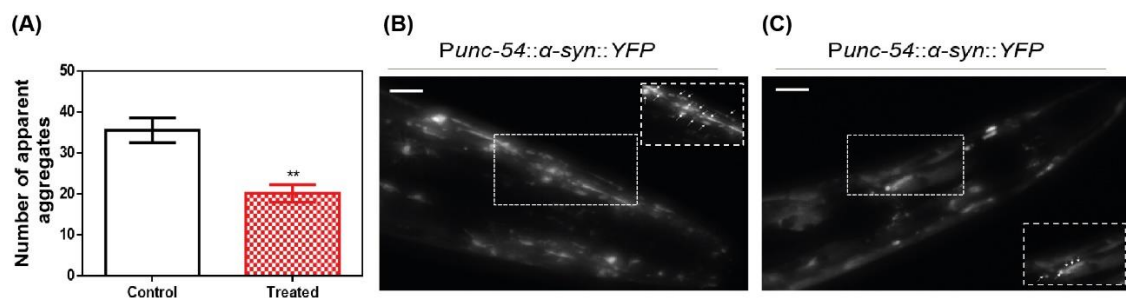
Figure 6. Inhibition of α-Syn inclusions formation in a C. elegans model. (A) Quantification of α-Syn muscle inclusions per area in NL5901 worms in the absence (white) and presence of ZPDm (red). (B and C) Representative images of apparent α-Syn muscle aggregates obtained by epifluorescence microscopy of NL5901 worms treated without (B) and with ZPDm (C). (Scale bars, 10 μm). Between 40 and 50 animals were analysed per condition. Aggregates are indicated by white arrows. Data are shown as means, and error bars are shown as the SE of means; ** $p < 0.01$ and *** $p < 0.001$. NL5901 C. elegans strain forms visible accumulations of aggregated α-Syn that are reduced when ZPDm is administered.

Discussion

The identification of small compounds that may abrogate the process of protein aggregation in neurodegenerative disorders is attracting increasing interest, both in academia and industry (Pujols et al., 2020).

The lack of structural information about the intermediate species that populate the reaction, and the intrinsically disordered nature of many of the proteins behind these diseases, has made it challenging to use of structure-guided drug design for amyloid inhibitors. Only recently, the high-resolution structures of the fibrils formed by proteins connected to different amyloidosis have allowed the rational design of peptides that interfere with the growth or seeding of the fibrils (Seidler et al., 2018; Saelices et al., 2019; Sangwan et al., 2020). However, because peptides usually display poor pharmacokinetics, which should be significantly optimized before they become drugs, small molecules are still the preferred option for the treatment of the diseases caused by the aggregation of proteins within the brain.

The screening of large chemical libraries in the search for α-Syn aggregation inhibitors has provided potent molecules like anle138b (Wagner et al., 2013), BIOD303 (Moree et al., 2015), SynuClean-D (Pujols et al., 2018), 582032 (Toth et al., 2019) or the collection of compounds recently reported by Otzen and co-workers (Kurnik et al., 2018). All these molecules display two or more aromatic rings in their structures, a property that is shared by active polyphenols like curcumin (Pandey et al., 2008), EGCG (Bieschke et al., 2010), and baicalein (Jiang et al., 2010), repurposed inhibitors like Fasudil (Tatenhorst et al., 2016) and LMTM (Schwab et al., 2017) or compounds generated by rational design like NPT100-18A (Wrasidlo et al., 2016). Usually, the aromatic rings form a planar hydrophobic core that is thought to interact with apolar exposed regions in α-Syn or its assemblies. However, despite the presence of multiple aromatic groups is recurrent in natural α-Syn aggregation inhibitors and many of the reported screening efforts result in the identification of this kind of molecules, several natural compounds exhibiting a single aromatic ring have been shown to act as α-Syn aggregation modulators (**Figure S3**), including scyllo-inositol, gallic acid, dopamine, safranal and caffeic acid (Herrera et al., 2008; Di Giovanni et al., 2010; Liu et al., 2014; Ibrahim and McLaurin, 2016; Save et al., 2019).

We recently used a robust high-throughput screening pipeline to uncover molecules able to modulate α-Syn fibrillation (Pujols et al., 2017). Among the active compounds, we searched for a small compound bearing a single aromatic ring. We identified ZPDm, which, interestingly enough, is a minimal version of ZPD-2, a potent inhibitor identified in the same library (Pena-Diaz et al., 2019), with half of its heavy aromatic atoms. This opened an opportunity to approach a comparative SAR for these molecules.

ZPDm reduces the *in vitro* aggregation of WT α-Syn and the protein's A30P and H50Q familial variants in a 60 %, or higher, at a 0.7:1 (protein: ZPDm) ratio. This activity was orthogonally confirmed by light-scattering and TEM. Moreover, the inhibitory activity of ZPDm reduced the number of PK-resistant and Th-T positive species in PMCA assays, thus interfering with α-Syn templated seeding and/or aggregates amplification (Herva et al., 2014). It should be explored whether adaptation of α-Syn to ZPDm at late PMCA stages might translate in some resistance to the molecule during aggregates propagation.

Solution NMR measurements indicated that ZPDm was not interacting with soluble α-Syn monomers, and, therefore, it is not expected to impact the physiological function of the protein. Moreover, the addition of ZPDm at different time points of the aggregation reaction suggested that ZPDm is mainly active at the latest stages of the aggregation and indeed, further analysis demonstrated that the molecule is capable of disassembling mature α-Syn amyloid fibrils generated under different solution conditions, conceptually similar to the α-Syn strains observed in different synucleinopathies (Bousset et al., 2013; Peelaerts et al., 2015). Importantly, these features translate into a significant reduction in the number of apparent aggregates in the muscular cell wall of a *C. elegans* model of PD when the compound is added in the food at a concentration of 10 uM; whether this i*n vivo* anti-aggregational effect results in animal phenotypic benefits should be further explored.

The above-described results illustrate how despite ZPD-2 and ZPDm share a significant part of their chemical structure and both are effective α-Syn aggregation inhibitors, their mechanism of action differ significantly, with ZPD-2 acting preferentially at the early stages of the fibrillation and becoming inactive once the polymerization has advanced significantly, being devoid of detectable fibril disrupting activity.

In contrast, ZPDm is more effective at later stages and behaves as a robust disaggregating agent. The fact that (trifluoromethyl)benzene is an inactive molecule indicates that the bulk of the structure shared by ZPD-2 and ZPDm acts as a building block and that the particular chemistry and spatial disposition of the groups that decorate this moiety are

responsible for the different mode of action of these compounds. For instance, the (trifluoromethyl)benzene contains the aromatic ring common to the vast majority of small active compounds. However, it lacks a strong hydrogen bond donor/acceptor, which is another characteristic common to many of these molecules (Figure S3). Therefore, our data suggest that these are the minimum requirements for an active α-Syn aggregation inhibitor. The aromatic rings would allow interactions with hydrophobic regions, and the polar groups might disrupt the abundant short inter-strand hydrogen bonds that contribute to the amyloid structure's sustainment. Indeed, despite their different size, the number of hydrogen bonds acceptors in ZPD-2 and ZPDm is fairly similar (Table 1). Despite speculative, the preferential affinity for early-stage aggregates exhibited by ZPD-2 could be explained by its extended aromatic core and higher Log $P_{O/W}$, which might facilitate interactions with exposed hydrophobic patches in oligomers and small aggregates. In contrast, the compact structure of ZPDm might allow targeting defined binding pockets at the ends of amyloid fibrils, interfering with fibril elongation, and eventually disrupting pre-formed non-covalent interactions.

From a pharmacokinetic point of view, ZPDm is predicted to be more soluble than ZPD-2, to exhibit a higher gastrointestinal absorption and better drug-likeness (Table 1). ZPDm is also predicted to be a better lead compound from a medicinal chemistry perspective than ZPD-2 (Table 1).

To the best of our knowledge, the only other active molecule with a single aromatic ring derived from the screening of a large chemical library is the compound 576755 (Toth et al., 2019), which in addition to a benzene ring, displays the expected hydrogen bonds acceptors/donors (**Figure S3**). 576755 is a potent aggregation inhibitor that acts at the oligomerization stage both *in vitro* and *in cells*. However, it was identified in a screening for compounds that interact with monomeric α-Syn, and therefore, is not expected to have fibril disrupting activity. Indeed, it did not impact fibril transmission (Toth et al., 2019), consistent with its activity and being complementary to that of ZPDm, a fibril anti-propagating agent in PMCA assays.

Overall, here we describe a new small molecule with the potential to be converted into a lead compound and, perhaps more importantly, together with previous data, envision a way to design minimal aromatic molecules with different α-Syn anti-aggregational activities rationally. Because ZPD-2 and ZPDm function on the same target but have

complementary activity, it will be interesting to test if a combination of them can have a synergic effect that overpasses the individual molecules' potential.

Materials and Methods

Protein expression and purification

WT α-Syn and its variants (H50Q and A30P) were expressed and purified as previously described (Pujols et al., 2017); the obtained protein was lyophilised and kept at -80 ºC until its use.

In vitro aggregation of α-Syn

Lyophilised α-Syn was carefully resuspended in sterile PBS 1X and filtered through 0,22 µm membrane to discard small aggregates. Aggregation was performed at 70 µM of α-Syn (WT, H50Q or A30P) in a sealed 96-well plate, in a total volume of 150 µL. 40 µM Th-T in PBS 1X, a 1/8" diameter Teflon polyball (Polysciences Europe GmbH, Eppelheim, Germany) and 100 µM ZPDm, (trifluoromethyl)benzene or DMSO (in control samples) were also added to each well. The plate was incubated at 37 ºC and 100 rpm fixed in an orbital shaker Max-Q 4000 (ThermoScientific, Waltham, Massachusetts, USA). Every 2h, Th-T fluorescence was measured in a Victor3.0 Multilabel Reader (PerkinElmer, Waltham, Massachusetts, USA), exciting through a 430–450 nm filter and collecting emission signal with a 480–510 filter. Each assay was done in triplicate and the values of the kinetic fitted according to the following equation:

Equation (1) $\qquad \propto = 1 - \dfrac{1}{k_b(e^{k_a t}-1)+1}$

where k_b and k_a constitute the homogeneous nucleation rate constant and the autocatalytic rate constant, respectively (Crespo et al., 2016).

ZPDm was added at different concentrations in the titration assays (200, 150, 100, 75 and 50 µM). In time-dependent assays, 7 independent aggregation reactions were prepared simultaneously and incubated as aforementioned in a 96-well plate as triplicates. 100 µM of ZPDm were added at different time points after the reaction had begun (4, 8, 12, 16, 20 and 24 h). In all cases α-Syn concentration was constant at 70 µM. An equivalent volume of DMSO was added to the control sample at the beginning of the reaction.

For the study of the disaggregation assays, α-Syn 70 µM was incubated in a 96-well plate as previously described for 2 days and Th-T fluorescence measured. Then, ZPDm was added to a final concentration of 100 µM. The plate was incubated for an additional 24 h and Th-T fluorescence was measured.

Strains were generated as previously described (Bousset et al., 2013; Peelaerts et al., 2015; Li et al., 2018; Carija et al., 2019). Briefly, α-Syn was resuspended in PBS 1X and dialysed for 24 h in a 1:1000 (v/v) ratio with either buffer B (50 mM Tris-HCl pH 7.0) or buffer C (50 mM Tris-HCl pH 7.0, 150 mM NaCl). Dialysed samples were filtered through 0,22 μm membranes and incubated at 70 μM for 2 days as described above. ZPDm was added to a final concentration of 100 μM and plates incubated for additional 24 h. Then, Th-T fluorescence was measured.

Light-scattering

80 μL of end-point aggregates were collected, placed into a quartz cuvette and analysed in a Cary Eclipse Fluorescence Spectrophotometer (Agilent, Santa Clara, California, USA) by exciting at 300 and 340 nm and 90º collecting between 280 and 360 nm.

Transmission Electron Microscopy (TEM)

End-point α-Syn aggregates were collected and diluted 1.10 in PBS 1X. Diluted samples were gently sonicated for 5 min and 5 μL of the resultant sample were placed on a carbon-coated copper grid for 5 min. Using a filter paper, the grids were dried to remove the excess of sample and washed twice with miliQ water. Finally, 5 μL of 2 % (w/v) uranyl acetate were added and left incubate for 2 min. As previously indicated, the excess of uranyl acetate was removed, and grids were left to air-dry for 10 min. Images were obtained using a Transmission Electron Microscopy Jeol 1400 (Peabody, Massachusetts, USA) operating at an accelerating voltage of 120 kV. A minimum of 30 fields were screened per sample, in order to collect representative images.

Protein Misfolding Cyclic Amplification (PMCA)

PMCA protocol was performed as previously described (Herva et al., 2014). Briefly, α-Syn was resuspended in Conversion Buffer (PBS 1X, 1% Triton X-100, 150 mM NaCl) to a final concentration of 90 μM and supplemented with Complete Protease Inhibitor Mixture (Roche Applied Science, Penzberg, Germany). 60 μL of this solution were loaded into 200 μL PCR tubes containing 1.0 mm silica beads (Biospec Products, Bartlesville, OK, USA). A-Syn was then exposed to 24-hour cycles of 30 s sonication and 30 min of incubation at 37 ºC, using a Misonix 4000 sonicator, at 70 % power. The incubated sample was recovered after each 24-hour cycle and 1 μL was added to a new PCR tube containing fresh α-Syn at 90 μM. In the case of ZPDm treated samples, the

compound was added to the fresh sample in each step to a final concentration of 128 μM, which corresponds to the 0.7:1 α-Syn:ZPDm ratio of the previous aggregation assays. Untreated samples were prepared adding the same concentration of DMSO (0.26%) present in the treated mixtures. This process was repeated for 5 days. All the reactions were made in triplicate.

10 μL of aggregated samples at the end of each cycle were diluted 1:10 with 90 μL of PBS 1X, 40 μM Th-T. Th-T fluorescence emission was measured in a Cary Eclipse Fluorescence Spectrophotometer (Agilent, Santa Clara, California, USA), by exciting the samples at 445 nm and collecting the emission signal between 460 and 600 nm.

Proteinase K digestion

18 μL of PMCA-aggregated α-Syn were incubated with 6 μL of Proteinase K (5 μg/mL as final concentration) for 30 min at 37 °C. Then, 8 μL of loading buffer containing 1% β-mercaptoethanol was added and the sample was incubated 10 min at 95 °C for PK inactivation. Finally, 7 μL of the samples were loaded into a Tricine-SDS-PAGE gel. Unstained Protein Standard markers (ThermoFisher Scientific, Waltham, Massachusetts, USA) were used as a reference. Gels were stained with Blue safe.

Nuclear Magnetic Resonance (NMR)

^{15}N-labelled human WT α-Syn was expressed in E. coli BL21 DE3 strains. Cells were grown in LB medium until the optical density (OD) at 600 nm reached a level of 0.6. Cultures were then centrifuged at 3000 rpm for 15 min and the obtained pellets resuspended in 1L of minimal medium: : 768 mL of miliQ water with 1 mL of ampicillin 100 mg/mL, 100 μL $CaCl_2$ 1M, 2 mL $MgSO_4$ 2 M, 20 mL glucose 20 %, 10 mL vitamins 100x (Sigma-Aldrich, Darmstadt, Germany), 200 mL salts M9 and 1 g $^{15}NH_4$ (Cambridge Isotope Laboratories, Inc., Tewksbury, Massachusetts, USA). Cells were incubated for 1h at 250 rpm and 37 °C. Finally, 1 mM IPTG was added to induce protein expression for 4 h. Protein was purified as previously described (Pujols et al., 2017).

Caenorhabditis elegans assays

Maintenance

Animals synchronization was carried out by bleaching and overnight hatching in M9 (3 g/L KH$_2$PO$_4$, 6 g/L Na$_2$HPO$_4$, 5 g/L NaCl, 1 M MgSO$_4$) buffer. Thus, nematodes were cultured at 20 oC on growth media plates (NGM) containing 1 mM CaCl$_2$, 1 mM MgSO$_4$, 5 μg/mL cholesterol, 250 M KH$_2$PO$_4$ pH 6.0, 17 g/L Agar, 3 g/L NaCl. Plates were previously seeded with E. coli OP50 strain. Nematodes were maintained using standard protocols [36].

Strains

Strain NL5901, unc-119(ed3) III; pkIs2386 [Punc-54::α-SYN::YFP; unc-119(+)] was obtained from the C. elegans Genetic Center (CGC). For the α-Syn induced dopaminergic degeneration analysis, strain UA196 [30], gifted generously by the laboratory of Dr Guy Caldwell (Department of Biological Science, The University of Alabama, Tuscaloosa, USA), was used; (sid-1(pk3321); baIn33 [Pdat-1::sid-1, Pmyo-2::mCherry]; baIn11 [Pdat-1::α-SYN; Pdat-1::GFP]). In the main text, this strain was named Pdat-1::GFP; Pdat-1::α-SYN.

ZPDm administration

After cooling, the autoclaved NGM agar medium (1 mM CaCl$_2$, 1 mM MgSO$_4$, 5 μg/mL cholesterol, 250 M KH$_2$PO$_4$ pH 6.0, 17 g/L Agar, 3 g/L NaCl) was enriched with 100 μM of a stock solution of 4 mM ZPDm in 0.2 % DMSO to a final concentration of 10 μM. After 2 days, plates were seeded with 250 μL of E. coli OP50 with 10 μM of ZPDm. Nematodes were placed on the plates at larval stages L4 and exposed either to ZPDm or DMSO (controls) for 7 days. Daily transfer was done to avoid cross progeny.

Aggregate quantification

The number of cellular inclusions was quantified as previously described [29; 37]. Briefly, NL5901 (Punc-54::α-SYN::YFP) worms were age-synchronized and left overnight to hatch. Nematodes in phase L1 were cultured and grown into individual NGM plates seeded with E. coli OP50. When animals reached L4 developmental stage, they were transferred onto either ZPDm treated plates or DMSO treated plates (negative control). Every day, animals were transferred into a new plate to avoid cross contamination. At stage L4+7, the aggregates in the anterior part of every single animal were counted. For each experiment, thirty 7-days old nematodes per treatment were analysed using a Nikon Eclipse E800 epifluorescence microscope equipped with an

Endow GFP HYQ filter cube (Chroma Technology Corp, Bellows Falls, Vermont USA) and each experiment was carried out in triplicate. Inclusions could be described as discrete bright structures, with edges distinguishable from surrounding fluorescence. ImageJ software was used for measuring the number of cellular aggregates considering the area dimensions. For the quantification of α-syn aggregates in C. elegans one single image was taken from each animal. However, every image contained among 30-45 stacks (1 µm) that allowed to detect aggregates that are at different positions.

Microscopy and imaging

Animals were placed in a 1 mM solution of sodium azide and mounted with a coverslip on a 4% agarose pad. Animals were visualized with a Nikon Eclipse E800 epifluorescence microscope. The system acquires a series of frames at specific Z-axis position (focal plane) using a Z-axis motor device. Animals were examined at 100× magnification to examine α-Syn induced DA cell death and at 40× to examine α-Syn apparent aggregate.

Statistical analysis

All graphs were generated with GraphPad Prism 6.0 software (GraphPad Software Inc, La Jolla, California, USA). Data were analysed by two-way ANOVA Tukey test using SPSS software version 20.0 (IBM Analytics, Armonk, NY, United States) and t-test using GraphPad software version 6.0 (GraphPad Software Inc, La Jolla, California, USA). All data are shown as means and standard error of mean (SEM). $P < 0.05$ was considered statistically significant. In the graphs *, ** and *** indicate $p < 0.05$, $p < 0.01$ and $p < 0.001$, respectively.

Supplementary Information for:

Inhibition of α-Synuclein aggregation and mature fibril disassembling with a minimalistic compound, ZPDm

Samuel Peña-Díaz[a,b,‡], Jordi Pujols[a,b,‡], Francisca Pinheiro[a,b], Jaime Santos[a,b], Irantzu Pallarés[a,b], Susanna Navarro[a,b], María Conde-Jimenez[c], Jesús García[d], Xavier Salvatella[d], Esther Dalfó[e,f], Javier Sancho[c], and Salvador Ventura[a,b,*]

[a] Institut de Biotecnologia i Biomedicina. Universitat Autonoma de Barcelona, 08193-Bellaterra, Spain.

[b] Departament de Bioquimica i Biologia Molecular. Universitat Autonoma de Barcelona, 08193-Bellaterra, Spain.

[c] Department of Biochemistry and Molecular and Cell Biology, Institute for Biocomputation and Physics of Complex Systems (BIFI), University of Zaragoza, 50018 Zaragoza, Spain

[d] Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Baldiri Reixac 10, 08028, Barcelona, Spain

[e] Medicine, M2, Universitat Autonoma de Barcelona (UAB), Bellaterra Campus, Cerdanyola del Valles, Barcelona, Spain. ORCID: 0000-0003-4677-8515

[f] Faculty of Medicine. University of Vic-Central University of Catalonia (UVic-UCC)

[*] Correspondence: salvador.ventura@uab.cat

[‡] These authors contributed equally to this work

Supplementary 1. Inhibitory capacity of (trifluoromethyl)benzene. (A) Chemical structures of (trifluoromethyl)benzene, (left) and ZPDm (right). (B) α-Syn aggregation kinetics in the absence (black) and presence (green) of 100 μM of (trifluoromethyl)benzene followed by Th-T fluorescence. (C) Light-scattering measurements at 300 and 340 nm, in the absence (white) and presence (green) of (trifluoromethyl)benzene.
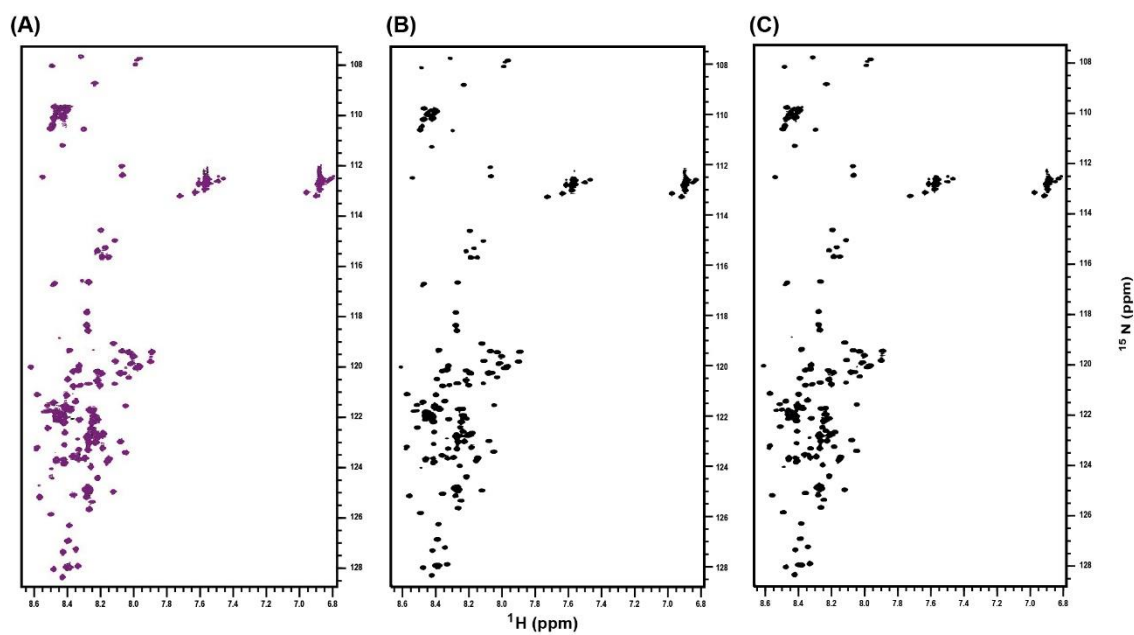
Supplementary 2. NMR analysis of ZPDm lack of interaction with monomeric α-Syn. ^{1}H-^{15}N HSQC NMR spectra of ^{15}N-labeled α-Syn (70 µM) in the presence (A) and in the absence (B) of ZPDm (100 µM). The superposition of the two NMR spectra is shown in (C).
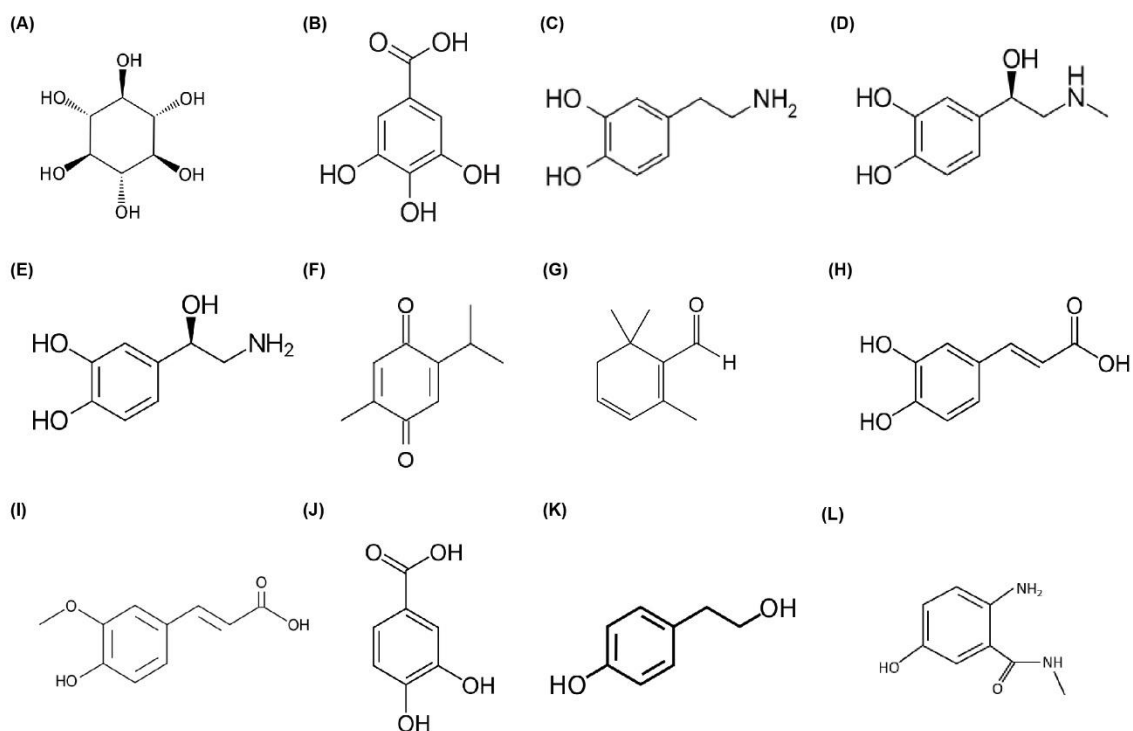
Supplementary 3. Chemical structures of different α-Syn aggregation inhibitors with a single aromatic ring. Chemical structures of (A) scyllo-inositol, (B) gallic acid, (C) dopamine, (D) epinephrine, (E) norepinephrine, (F) thymoquinone, (G) safranal, (H) caffeic acid, (I) ferulic acid, (J) protocathecuic acid, (K) tyrosol, and (L) 576755.

CHAPTER 6

Conclusions

Conclusions

Chapter 3. Conditional folding of disulfide-rich proteins

Work 1: Disulfide driven folding for a conditionally disordered protein

- Reduced COX17 is mostly disordered. In the absence of the structural disulfides, COX17 lacks secondary structural elements and conformational compaction.

- The disordered nature of reduced COX17 is in agreement with the intrinsic disorder predictions performed *in silico*.

- COX17 air-oxidizes and spontaneously folds into the native state. The native disulfides' formation follows a sequential oxidation, being the inner disulfide the first to form and the rate-limiting step of the folding reaction.

- The formation of the inner disulfide bond suffices to trigger the disorder-to-order transition towards a compacted native-like state, in which COX17 already possesses most of its helical content. This structural transition precedes the formation of the outer disulfide bond.

- MIA40 promotes the disorder-to-order transition by catalyzing the specific oxidation of the inner disulfide.

- The unspecific oxidation of COX17 with redox agents accelerates the folding rate, keeping the inner disulfide as the predominant intermediate species, indicating that its formation is conformationally favored.

- Induction of α-helices has a negative impact on COX17 folding. In contrast, COX17 oxidation in the presence of denaturing agents accelerates the folding rates. Therefore, the formation of secondary structural elements plays a minor role in guiding COX17 folding pathway.

Work 2: MIA40 bypasses the folding constraints imposed by TRIAP1 function

- Reduced TRIAP1 is partially folded. In the absence of the structural disulfides, it possesses a high content of helical structure and conformational packing. This state is

dynamic and exposes hydrophobic clusters to solvent. Altogether, reduced TRIAP1 behaves as a molten globule.

- The molten globule arrangement of reduced TRIAP1 is in agreement with its inherent low disorder tendency as reported by intrinsic disorder predictions performed *in silico*.

- TRIAP1 spontaneously attains its folded state by the air-oxidation of its cysteines and disulfide bridging. The oxidative folding involves two intermediates and is slow.

- The molten globule's conformational bias determines the folding pathway of TRIAP1 by facilitating the accumulation of an early, kinetically trapped and off-pathway intermediate.

- MIA40 funnels the oxidative folding of TRIAP1. In the presence of MIA40, TRIAP1 folding proceeds fast and bypasses the kinetically trapped intermediate.

- The unsual molten globule arrangement of reduced TRIAP1 responds to the high hydrophobic load of TRIAP1, required in its native state to perform its function as a molecular chaperone of the PRELI-like protein family.

Chapter 4. Disorder and amyloidogenicity

Work 3: The disordered C-terminus of yeast Hsf1 contains a cryptic low-complexity

amyloidogenic region

- We used a series of bioinformatics analysis to identify a low-complexity region, biased towards asparagine and glutamine amino acids, in the yeast's C-terminal disordered region Hsf1 protein. This theoretically soluble region contains an 18-residues region displaying cryptic amyloidogenicity.

- Hsf1 peptide self-assemblies spontaneously into β-sheet enriched aggregates in a concentration-dependent manner, despite the polar nature of the sequence. Hsf1 peptide aggregates are positive to amyloid-specific probes such as binding to Congo red, Thioflavin-T, and Thioflavin-S dyes. Transmission Electron Microscopy confirmed the amyloid aggregation of Hsf1 peptide.

- The amyloid core of Hsf1 overlaps with a predicted interaction region, emphasizing the overlap between function and aggregation determinants not only in globular proteins but also in disordered regions and proteins.

Chapter 5. Chemical chaperones as novel drugs for Parkinson's Disease

Work 4: High-throughput screening methodology to identify α-synuclein aggregation inhibitors

- We set up a reproducible system to monitor α-synuclein aggregation in 96-well plates by combining continuous mild agitation at 37ºC in an orbital shaker with the presence of Teflon beads in the reaction solution. With this strategy, we reduce the amount of pure protein used per assay while achieving complete reactions after only 24 hours, significantly faster than other methodologies.

- The method is suitable to screen modulators of α-synuclein aggregation in a fast and reproducible fashion. Therefore, we successfully screened a chemical library of more than 14.000 compounds.

- The analysis of the derived kinetics allowed discriminating bona fide modulators of the aggregation reaction.

Work 5: A novel small molecule inhibits α-synuclein aggregation, disrupts amyloid fibrils, and prevents degeneration of dopaminergic neurons

- Synuclean-D inhibits α-synuclein aggregation in vitro in a concentration-dependent manner, reaching significant activity even at substoichiometric ratios.

- The H50Q and A30P versions of α-synuclein, associated with Parkinson's Disease's familiar variants, are also sensitive to Synuclean-D.

- Synuclean-D targets late-stages of α-synuclein aggregation and disrupts preformed mature amyloid fibrils.

- Synuclean-D does not bind to soluble monomers of α-synuclein. Instead, induced fit docking simulations reported that Synuclean-D could interact with the mature fibril's core via van der Waals contacts. This interaction might promote fibril disassembling.

- Synuclean-D inhibited the accumulation of intracellular aggregates in neuroglioma cells (H4) that overexpressed α-synuclein.

254

- The anti-aggregational activity of Synuclean-D was recapitulated in a simple animal model of Parkinson's Disease. Synuclean-D reduced the number of α-synuclein inclusions in C. elegans overexpressing α-synuclein in muscular cells, resulting in motor amelioration, compared to non-treated worms.

- Synuclean-D has a neuroprotective effect when administered to C. elegans that overexpress α-synuclein in dopaminergic neurons, increasing the number of functional neurons compared to non-treated worms.

Work 6: ZPD-2, a small compound that inhibits α-synuclein amyloid aggregation and its seeded polymerization

- ZPD-2 is a potent inhibitor of α-synuclein aggregation. It reduces the formation of Th-T positive aggregates by 80% at equimolar ratios and maintains significant inhibition at substoichiometric equivalents.

- ZPD-2 significantly delays the t50 and reduces the nucleation rate constant of the aggregation reaction, which differs from the Synuclean-D mode of action and suggests that it impacts the early stages. We hypothesize that different molecular scaffolds might affect different stages of the reaction by interacting with different assembled species.

- Similar to Synuclean-D, ZPD-2 does not interact with the monomeric species of α-synuclein, keeping the native function of the protein.

- The H50Q and A30P versions of α-synuclein are also sensitive to ZPD-2.

- ZPD-2 inhibits the aggregation of different α-synuclein conformations, even in the presence of preformed seeds.

- Food supplementation with ZPD-2 reduces the number of apparent aggregates and promotes neuroprotection in two C. elegans models of Parkinson's Disease, overexpressing α-synuclein in muscular cells and dopaminergic neurons, respectively.

Work 7: Inhibition of α-Synuclein aggregation and mature fibril disassembling with a minimalistic compound, ZPDm

- ZPDm is a minimalistic version of ZPD-2 that possesses inhibition activity against the aggregation of wild type and two familiar versions of α-synuclein, H50Q, and A30P. As far as we know, is the only non-natural compound composed by a single aromatic ring that possess anti-aggregation activity against α-synuclein aggregation.

- The (trifluoromethyl)benzene scaffold of ZPDm, also present in ZPD-2 and other positive hits of the initial screening, is devoid of anti-aggregational activity. This suggests that the (trifluoromethyl)benzene is not sufficient to promote the inhibition of α-synuclein and reinforces the role of the other chemical groups of ZPDm.

- In line with ZPD-2 behavior, ZPDm impacts the nucleation rate of α-synuclein aggregation. However, ZPDm also disassembles different preformed fibril polymophs, a property not shared by ZPD-2.

- T*he anti-aggregational activity of ZPDm is recapitulated in the* C. elegans *model of Parkinson's Disease, in which the administration of the compound reduces the number of apparent aggregates in muscular cells overexpressing α-synuclein.*

Overall conclusions of Chapter 5:

- The disordered nature α-synuclein precludes using rational strategies to develop active molecules targeted against specific regions of the protein structure. Therefore, blind high-throughput screening of large chemical libraries stands as a valuable strategy to identify α-synuclein aggregation inhibitors.

- The active molecules reported in the present thesis share some structural features, although they impact different reaction stages. This standard structure can be outlined as a hydrophobic scaffold of aromatic rings coated with polar substitutions. We hypothesize that this disposition allows the molecule to interact with apolar stretches and destabilize hydrogen bonding in amyloid assemblies.

- The differences in size, hydrophobicity, and chemical substituents of Synuclean-D, ZPD-2, and ZPDm might explain their remarkably different mechanisms of action.

- We found no correlation between *in vitro* potency and performance *in vivo*, meaning that additional properties of the molecules, such as administration, distribution, metabolism and degradation, also play an important role in their effectiveness. Therefore, their pharmacological profiles must be assessed before progressing to proof of concept assays in mammals or pre-clinical stages.

References

1.   Kendrew, J. C. *et al.* A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* **181**, 662–666 (1958).

2.   Perutz, M. F. *et al.* Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5-A. resolution, obtained by X-ray analysis. *Nature* **185**, 416–422 (1960).

3.   Wu, H. Studies on denaturation of proteins. XIII. A theory of denaturation. *Adv Protein Chem* **46**, 6–26; discussion 1-5 (1931).

4.   Kauzmann, W. Structural factors in protein denaturation. *J Cell Physiol Suppl* **47**, 113–131 (1956).

5.   Blake C, C. Structure of hen egg-white lysozyme. A three-dimensional Fourier synthesis at 2 Angstrom resolution - PubMed. *Nature* **185**, 416–422 (1960).

6.   Matthews, B. W., Sigler, P. B., Henderson, R. & Blow, D. M. Three-dimensional structure of tosyl-alpha-chymotrypsin. *Nature* **214**, 652–656 (1967).

7.   Wyckoff, H. W. *et al.* The structure of ribonuclease-S at 3.5 A resolution. *J Biol Chem* **242**, 3984–3988 (1967).

8.   Lipscomb, W, N. The structure of carboxypeptidase A. IX. The x-ray diffraction results in the light of the chemical sequence - PubMed. *Proc Natl Acad Sci USA* **64**, 28–35 (1969).

9.   Arnone, A. *et al.* A high resolution structure of an inhibitor complex of the extracellular nuclease of Staphylococcus aureus. I. Experimental procedures and chain tracing. *J Biol Chem* **246**, 2302–2316 (1971).

10.   Anfinsen, C. B., Haber, E., Sela, M. & White, F. H. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc Natl Acad Sci U S A* **47**, 1309–1314 (1961).

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

12.   Berman, H. M. The Protein Data Bank. *Nucleic Acids Res* **28**, 235–242 (2000).

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

14.   Radivojac, P. *et al.* Intrinsic Disorder and Functional Proteomics. *Biophysical Journal* **92**, 1439–1456 (2007).

15.   Habchi, J., Tompa, P., Longhi, S. & Uversky, V. N. Introducing protein intrinsic disorder. *Chem Rev* **114**, 6561–6588 (2014).

16.   Uversky, V. N. & Dunker, A. K. Understanding protein non-folding. *Biochim Biophys Acta* **1804**, 1231–1264 (2010).

17.   Uversky, V. N., Gillespie, J. R. & Fink, A. L. Why are 'natively unfolded' proteins unstructured under physiologic conditions? *Proteins* **41**, 415–427 (2000).

18.   Theillet, F.-X. *et al.* The alphabet of intrinsic disorder: I. Act like a Pro: On the abundance and roles of proline residues in intrinsically disordered proteins. *Intrinsically Disord Proteins* **1**, e24360 (2013).

19.   Shoulders, M. D. & Raines, R. T. Collagen structure and stability. *Annu Rev Biochem* **78**, 929–958 (2009).

20.   Kumar, P. & Bansal, M. Structural and functional analyses of PolyProline-II helices in globular proteins. *J Struct Biol* **196**, 414–425 (2016).

21.   Das, R. K., Ruff, K. M. & Pappu, R. V. Relating sequence encoded information to form and function of intrinsically disordered proteins. *Curr Opin Struct Biol* **32**, 102–112 (2015).

22.   Holehouse, A. S. & Pappu, R. V. Collapse Transitions of Proteins and the Interplay Among Backbone, Sidechain, and Solvent Interactions. *Annu. Rev. Biophys.* **47**, 19–39 (2018).

23.   van der Lee, R. *et al.* Classification of intrinsically disordered regions and proteins. *Chem Rev* **114**, 6589–6631 (2014).

24.   Uversky, V. N. Protein intrinsic disorder and structure-function continuum. in *Progress in Molecular Biology and Translational Science* vol. 166 1–17 (Elsevier, 2019).

25.   Meng, F., Uversky, V. N. & Kurgan, L. Comprehensive review of methods for prediction of intrinsic disorder and its molecular functions. *Cell Mol Life Sci* **74**, 3069–3090 (2017).

26.   Ferron, F., Longhi, S., Canard, B. & Karlin, D. A practical overview of protein disorder prediction methods. *Proteins* **65**, 1–14 (2006).

27.   Ward, J. J., Sodhi, J. S., McGuffin, L. J., Buxton, B. F. & Jones, D. T. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J Mol Biol* **337**, 635–645 (2004).

28.   Uversky, V. N. A decade and a half of protein intrinsic disorder: biology still waits for physics. *Protein Sci* **22**, 693–724 (2013).

29.   Xie, H. *et al.* Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions. *J Proteome Res* **6**, 1882–1898 (2007).

30.   Dunker, A. K., Cortese, M. S., Romero, P., Iakoucheva, L. M. & Uversky, V. N. Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS J* **272**, 5129–5148 (2005).

31.   Kim, P. M., Sboner, A., Xia, Y. & Gerstein, M. The role of disorder in interaction networks: a structural analysis. *Mol Syst Biol* **4**, 179 (2008).

32.   Frye, J. J. *et al.* Electron microscopy structure of human APC/C(CDH1)-EMI1 reveals multimodal mechanism of E3 ligase shutdown. *Nat Struct Mol Biol* **20**, 827–835 (2013).

33.   Xie, H. *et al.* Functional Anthology of Intrinsic Disorder. III. Ligands, Postranslational Modifications and Diseases Associated with Intrinsically Disordered Proteins. *J Proteome Res* **6**, 1917–1932 (2007).

34.   Sirota, F. L., Maurer-Stroh, S., Eisenhaber, B. & Eisenhaber, F. Single-residue posttranslational modification sites at the N-terminus, C-terminus or in-between: To be or not to be exposed for enzyme access. *Proteomics* **15**, 2525–2546 (2015).

35.   Darling, A. L. & Uversky, V. N. Intrinsic Disorder and Posttranslational Modifications: The Darker Side of the Biological Dark Matter. *Front Genet* **9**, (2018).

36.   Wright, P. E. & Dyson, H. J. Linking Folding and Binding. *Curr Opin Struct Biol* **19**, 31–38 (2009).

37.   Smith, M. D. & Jelokhani-Niaraki, M. pH-Induced Changes in Intrinsically Disordered Proteins. in *Intrinsically Disordered Protein Analysis* (eds. Uversky, V. N. & Dunker, A. K.) vol. 896 223–231 (Springer New York, 2012).

38.   Huihui, J., Firman, T. & Ghosh, K. Modulating charge patterning and ionic strength as a strategy to induce conformational changes in intrinsically disordered proteins. *J Chem Phys* **149**, 085101 (2018).

39.   Bonetti, D., Troilo, F., Brunori, M., Longhi, S. & Gianni, S. How Robust Is the Mechanism of Folding-Upon-Binding for an Intrinsically Disordered Protein? *Biophys J* **114**, 1889–1894 (2018).

40.   Dogan, J., Gianni, S. & Jemth, P. The binding mechanisms of intrinsically disordered proteins. *Phys. Chem. Chem. Phys.* **16**, 6323–6331 (2014).

41.   Reichmann, D. & Jakob, U. The roles of conditional disorder in redox proteins. *Curr Opin Struct Biol* **23**, 436–442 (2013).

42.   Bhopatkar, A. A., Uversky, V. N. & Rangachari, V. Disorder and cysteines in proteins: A design for orchestration of conformational see-saw and modulatory functions. in *Progress in Molecular Biology and Translational Science* vol. 174 331–373 (Elsevier, 2020).

43.   Mayer, M. P. The unfolding story of a redox chaperone. *Cell* **148**, 843–844 (2012).

44.   Chiti, F. & Dobson, C. M. Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade. *Annu Rev Biochem* **86**, 27–68 (2017).

45.   Jakob, U., Kriwacki, R. & Uversky, V. N. Conditionally and transiently disordered proteins: awakening cryptic disorder to regulate protein function. *Chem Rev* **114**, 6779–6805 (2014).

46.   Scollo, F. & La Rosa, C. Amyloidogenic Intrinsically Disordered Proteins: New Insights into Their Self-Assembly and Their Interaction with Membranes. *Life (Basel)* **10**, (2020).

47.   Mukhopadhyay, S. The Dynamism of Intrinsically Disordered Proteins: Binding-Induced Folding, Amyloid Formation, and Phase Separation. *J Phys Chem B* **124**, 11541–11560 (2020).

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

49.   Toyama, B. H. & Weissman, J. S. Amyloid Structure: Conformational Diversity and Consequences. *Annu. Rev. Biochem.* **80**, 557–585 (2011).

50.   Jarrett, J. T. & Lansbury, P. T. Seeding 'one-dimensional crystallization' of amyloid: a pathogenic mechanism in Alzheimer's disease and scrapie? *Cell* **73**, 1055–1058 (1993).

51.   Harper, J. D. & Lansbury, P. T. MODELS OF AMYLOID SEEDING IN ALZHEIMER'S DISEASE AND SCRAPIE: Mechanistic Truths and Physiological Consequences of the Time-Dependent Solubility of Amyloid Proteins. *Annual Review of Biochemistry* **66**, 385–407 (1997).

52.   Carulla, N. *et al.* Molecular recycling within amyloid fibrils. *Nature* **436**, 554–558 (2005).

53.   Conchillo-Solé, O. *et al.* AGGRESCAN: a server for the prediction and evaluation of 'hot spots' of aggregation in polypeptides. *BMC Bioinformatics* **8**, 65 (2007).

54.   López de la Paz, M. & Serrano, L. Sequence determinants of amyloid fibril formation. *Proc Natl Acad Sci U S A* **101**, 87–92 (2004).

55.   Pastor, M. T., Esteras-Chopo, A. & Serrano, L. Hacking the code of amyloid formation: the amyloid stretch hypothesis. *Prion* **1**, 9–14 (2007).

56.   Fitzpatrick, A. W. P. *et al.* Atomic structure and hierarchical assembly of a cross-β amyloid fibril. *Proc Natl Acad Sci U S A* **110**, 5468–5473 (2013).

57.   Gianni, S. *et al.* Understanding the frustration arising from the competition between function, misfolding, and aggregation in a globular protein. *Proc Natl Acad Sci U S A* **111**, 14141–14146 (2014).

58.   Aguzzi, A. & Calella, A. M. Prions: protein aggregation and infectious diseases. *Physiol Rev* **89**, 1105–1152 (2009).

59.   Kraus, A., Groveman, B. R. & Caughey, B. Prions and the potential transmissibility of protein misfolding diseases. *Annu Rev Microbiol* **67**, 543–564 (2013).

60.   Prusiner, S. B. Molecular biology of prion diseases. *Science* **252**, 1515–1522 (1991).

61.   Halfmann, R., Alberti, S. & Lindquist, S. Prions, protein homeostasis, and phenotypic diversity. *Trends Cell Biol* **20**, 125–133 (2010).

62.   Newby, G. A. & Lindquist, S. Blessings in disguise: biological benefits of prion-like mechanisms. *Trends Cell Biol* **23**, 251–259 (2013).

63.   Uptain, S. M. & Lindquist, S. Prions as protein-based genetic elements. *Annu Rev Microbiol* **56**, 703–741 (2002).

64.   Toombs, J. A. *et al.* De novo design of synthetic prion domains. *Proc Natl Acad Sci U S A* **109**, 6519–6524 (2012).

65.   Batlle, C., Iglesias, V., Navarro, S. & Ventura, S. Prion-like proteins and their computational identification in proteomes. *Expert Rev Proteomics* **14**, 335–350 (2017).

66.   Espinosa Angarica, V., Ventura, S. & Sancho, J. Discovering putative prion sequences in complete proteomes using probabilistic representations of Q/N-rich domains. *BMC Genomics* **14**, 316 (2013).

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

68.   Alberti, S., Halfmann, R., King, O., Kapila, A. & Lindquist, S. A systematic survey identifies prions and illuminates sequence features of prionogenic proteins. *Cell* **137**, 146–158 (2009).

69.   Stefani, M. Structural features and cytotoxicity of amyloid oligomers: Implications in Alzheimer's disease and other diseases with amyloid deposits. *Progress in Neurobiology* **99**, 226–245 (2012).

70.   Poole, L. B. The basics of thiols and cysteines in redox biology and chemistry. *Free Radical Biology and Medicine* **80**, 148–157 (2015).

71.   Arolas, J. L., Aviles, F. X., Chang, J.-Y. & Ventura, S. Folding of small disulfide-rich proteins: clarifying the puzzle. *Trends Biochem Sci* **31**, 292–301 (2006).

72.   Tu, B. P., Ho-Schleyer, S. C., Travers, K. J. & Weissman, J. S. Biochemical basis of oxidative protein folding in the endoplasmic reticulum. *Science* **290**, 1571–1574 (2000).

73.   Ellgaard, L. & Ruddock, L. W. The human protein disulphide isomerase family: substrate interactions and functional properties. *EMBO Rep* **6**, 28–32 (2005).

74.   Soares Moretti, A. I. & Martins Laurindo, F. R. Protein disulfide isomerases: Redox connections in and out of the endoplasmic reticulum. *Arch Biochem Biophys* **617**, 106–119 (2017).

75.   Mordas, A. & Tokatlidis, K. The MIA pathway: a key regulator of mitochondrial oxidative protein folding and biogenesis. *Acc Chem Res* **48**, 2191–2199 (2015).

76.   Fraga, H., Graña-Montes, R., Illa, R., Covaleda, G. & Ventura, S. Association Between Foldability and Aggregation Propensity in Small Disulfide-Rich Proteins. *Antioxid Redox Signal* **21**, 368–383 (2014).

77.   Herrmann, J. M. & Riemer, J. The intermembrane space of mitochondria. *Antioxid Redox Signal* **13**, 1341–1358 (2010).

78.   Modjtahedi, N., Tokatlidis, K., Dessen, P. & Kroemer, G. Mitochondrial Proteins Containing Coiled-Coil-Helix-Coiled-Coil-Helix (CHCH) Domains in Health and Disease. *Trends Biochem Sci* **41**, 245–260 (2016).

79.   Pfanner, N., Warscheid, B. & Wiedemann, N. Mitochondrial proteins: from biogenesis to functional networks. *Nat Rev Mol Cell Biol* **20**, 267–284 (2019).

80.   Mesecke, N. *et al.* A Disulfide Relay System in the Intermembrane Space of Mitochondria that Mediates Protein Import. *Cell* **121**, 1059–1069 (2005).

81.   Fischer, M. & Riemer, J. The Mitochondrial Disulfide Relay System: Roles in Oxidative Protein Folding and Beyond. *Int J Cell Biol* **2013**, 12 (2013).

82.   Chacinska, A. *et al.* Essential role of Mia40 in import and assembly of mitochondrial intermembrane space proteins. *EMBO J* **23**, 3735–3746 (2004).

83.   Naoé, M. *et al.* Identification of Tim40 that mediates protein sorting to the mitochondrial intermembrane space. *J Biol Chem* **279**, 47815–47821 (2004).

84.   Chatzi, A. & Tokatlidis, K. The Mitochondrial Intermembrane Space: A Hub for Oxidative Folding Linked to Protein Biogenesis. *Antioxid Redox Signal* **19**, 54–62 (2013).

85.   Longen, S. *et al.* Systematic Analysis of the Twin Cx9C Protein Family. *J Mol Biol* **393**, 356–368 (2009).

86.   Habich, M. *et al.* Vectorial Import via a Metastable Disulfide-Linked Complex Allows for a Quality Control Step and Import by the Mitochondrial Disulfide Relay. *Cell Rep* **26**, 759-774.e5 (2019).

87.   Müller, J. M., Milenkovic, D., Guiard, B., Pfanner, N. & Chacinska, A. Precursor Oxidation by Mia40 and Erv1 Promotes Vectorial Transport of Proteins into the Mitochondrial Intermembrane Space. *Mol Biol Cell* **19**, 226–236 (2008).

88.   Herrmann, J. M. & Köhl, R. Catch me if you can! Oxidative protein trapping in the intermembrane space of mitochondria. *J Cell Biol* **176**, 559–563 (2007).

89.   Banci, L. *et al.* MIA40 is an oxidoreductase that catalyzes oxidative protein folding in mitochondria. *Nat Struct Mol Biol* **16**, 198–206 (2009).

90.   Chacinska, A. *et al.* Mitochondrial Biogenesis, Switching the Sorting Pathway of the Intermembrane Space Receptor Mia40. *J Biol Chem* **283**, 29723–29729 (2008).

91.   Hofmann, S. *et al.* Functional and Mutational Characterization of Human MIA40 Acting During Import into the Mitochondrial Intermembrane Space. *Journal of Molecular Biology* **353**, 517–528 (2005).

92.   Banci, L. *et al.* Molecular chaperone function of Mia40 triggers consecutive induced folding steps of the substrate in mitochondrial protein import. *PNAS* **107**, 20190–20195 (2010).

93.   Koch, J. R. & Schmid, F. X. Mia40 targets cysteines in a hydrophobic environment to direct oxidative protein folding in the mitochondria. *Nat Commun* **5**, 3041 (2014).

94.   Sideris, D. P. *et al.* A novel intermembrane space–targeting signal docks cysteines onto Mia40 during mitochondrial oxidative folding. *J Cell Biol* **187**, 1007–1022 (2009).

95.   Milenkovic, D. *et al.* Identification of the signal directing Tim9 and Tim10 into the intermembrane space of mitochondria. *Mol Biol Cell* **20**, 2530–2539 (2009).

96.   Milenkovic, D. *et al.* Biogenesis of the Essential Tim9–Tim10 Chaperone Complex of Mitochondria: SITE-SPECIFIC RECOGNITION OF CYSTEINE RESIDUES BY THE INTERMEMBRANE SPACE RECEPTOR Mia40. *J Biol Chem* **282**, 22472–22480 (2007).

97.   Riemer, J., Bulleid, N. & Herrmann, J. M. Disulfide Formation in the ER and Mitochondria: Two Solutions to a Common Process. *Science* **324**, 1284–1287 (2009).

98.   Banci, L., Barbieri, L., Luchinat, E. & Secci, E. Visualization of Redox-Controlled Protein Fold in Living Cells. *Chemistry & Biology* **20**, 747–752 (2013).

99. Fischer, M. *et al.* Protein import and oxidative folding in the mitochondrial intermembrane space of intact mammalian cells. *Mol Biol Cell* **24**, 2160–70 (2013).

100. Morgan, B. & Lu, H. Oxidative folding competes with mitochondrial import of the small Tim proteins. *Biochem J* **411**, 115–122 (2008).

101. Araiso, Y. *et al.* Structure of the mitochondrial import gate reveals distinct preprotein paths. *Nature* **575**, 395–401 (2019).

102. Backes, S. & Herrmann, J. M. Protein Translocation into the Intermembrane Space and Matrix of Mitochondria: Mechanisms and Driving Forces. *Front Mol Biosci* **4**, 83 (2017).

103. Wrobel, L. *et al.* Mistargeted mitochondrial proteins activate a proteostatic response in the cytosol. *Nature* **524**, 485–488 (2015).

104. Topf, U., Wrobel, L. & Chacinska, A. Chatty Mitochondria: Keeping Balance in Cellular Protein Homeostasis. *Trends Cell Biol* **26**, 577–586 (2016).

105. Mohanraj, K. *et al.* Inhibition of proteasome rescues a pathogenic variant of respiratory chain assembly factor COA7. *EMBO Mol Med* **11**, (2019).

106. Bragoszewski, P., Gornicka, A., Sztolsztener, M. E. & Chacinska, A. The Ubiquitin-Proteasome System Regulates Mitochondrial Intermembrane Space Proteins. *Molecular and Cellular Biology* **33**, 2136–2148 (2013).

107. Bien, M. *et al.* Mitochondrial disulfide bond formation is driven by intersubunit electron transfer in Erv1 and proofread by glutathione. *Mol Cell* **37**, 516–528 (2010).

108. Bragoszewski, P. *et al.* Retro-translocation of mitochondrial intermembrane space proteins. *PNAS* **112**, 7713–7718 (2015).

109. Friederich, M. W. *et al.* Mutations in the accessory subunit NDUFB10 result in isolated complex I deficiency and illustrate the critical role of intermembrane space import for complex I holoenzyme assembly. *Hum Mol Genet* **26**, 702–716 (2017).

110. Spillantini MG *et al.* Alpha-synuclein in Lewy bodies. *Nature* **388**, 839–840 (1997).

111. Wakabayashi, K. *et al.* Accumulation of α-synuclein/NACP is a cytopathological feature common to Lewy body disease and multiple system atrophy. *Acta Neuropathol* **96**, 445–452 (1998).

112.   Sorrentino, Z. A., Giasson, B. I. & Chakrabarty, P. α-Synuclein and astrocytes: tracing the pathways from homeostasis to neurodegeneration in Lewy body disease. *Acta Neuropathol* e1–e21 (2019) doi:10.1007/s00401-019-01977-2.

113.   Hogarth, P. Neurodegeneration with Brain Iron Accumulation: Diagnosis and Management. *J Mov Disord* **8**, 1–13 (2015).

114.   Halliday, G. M., Holton, J. L., Revesz, T. & Dickson, D. W. Neuropathology underlying clinical variability in patients with synucleinopathies. *Acta Neuropathol* **122**, 187–204 (2011).

115.   Lashuel, H. A., Overk, C. R., Oueslati, A. & Masliah, E. The many faces of α-synuclein: from structure and toxicity to therapeutic target. *Nat Rev Neurosci* **14**, 38–48 (2013).

116.   Jankovic, J. Parkinson's disease: clinical features and diagnosis. *J Neurol Neurosurg Psychiatry* **79**, 368–376 (2008).

117.   Fereshtehnejad, S.-M. *et al.* Evolution of prodromal Parkinson's disease and dementia with Lewy bodies: a prospective study. *Brain* **Epub**, e1–e17 (2019).

118.   Goedert, M., Spillantini, M. G., Del Tredici, K. & Braak, H. 100 years of Lewy pathology. *Nat Rev Neurol* **9**, 13–24 (2013).

119.   Lynch, T., Farrer, M., Hutton, M. & Hardy, J. Genetics of Parkinson's Disease. *Science* **278**, 1209–1213 (1997).

120.   Krüger R *et al.* Ala30Pro mutation in the gene encoding alpha-synuclein in Parkinson's disease. *Nature* **18**, 106–108 (1998).

121.   Singleton AB *et al.* a-Synuclein Locus Triplication Causes Parkinson's Disease. *Science* **302**, 841–841 (2003).

122.   Chartier-Harlin, M.-C. *et al.* a-synuclein locus duplication as a cause of familial. *Lancet* **364**, 167–169 (2004).

123.   Ko, W. K. D. & Bezard, E. Experimental animal models of Parkinson's disease: A transition from assessing symptomatology to α-synuclein targeted disease modification. *Exp Neurol* **298**, 172–179 (2017).

124.   Burré, J. The Synaptic Function of α-Synuclein. *J Parkinsons Dis* **5**, 699–713 (2015).

125.   Bartels, T., Choi, J. G. & Selkoe, D. J. α-Synuclein occurs physiologically as a helically folded tetramer that resists aggregation. *Nature* **477**, 107–110 (2011).

126.   Corbett, G. T. & Kordower, J. H. The native form of α-Synuclein: Monomer, tetramer, or a combination in equilibrium: Hot Topics. *Mov Disord* **30**, 1870–1870 (2015).

127.   Vilar, M. *et al.* The fold of a-synuclein fibrils. *Proc Natl Acad Sci USA* **105**, 8637–8642 (2008).

128.   Jao, C. C., Hegde, B. G., Chen, J., Haworth, I. S. & Langen, R. Structure of membrane-bound a-synuclein from site-directed spin labeling and computational refinement. *Proc Natl Acad Sci USA* **105**, 19666–19671 (2008).

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

130.   Ueda, K. *et al.* Molecular cloning of cDNA encoding an unrecognized component of amyloid in Alzheimer disease. *Proc Natl Acad Sci USA* **90**, 11282–11286 (1993).

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

132.   Izawa, Y. *et al.* Role of C-terminal negative charges and tyrosine residues in fibril formation of α-synuclein. *Brain Behav* **2**, 595–605 (2012).

133.   Baba M *et al.* Aggregation of alpha-synuclein in Lewy bodies of sporadic Parkinson's disease and dementia with Lewy bodies. *Am J Pathol* **152**, 879–884 (1998).

134.   Fusco, G. *et al.* Structural basis of membrane disruption and cellular toxicity by a-synuclein oligomers. *Science* **358**, 1440–1443 (2017).

135.   Peelaerts, W. *et al.* α-Synuclein strains cause distinct synucleinopathies after local and systemic administration. *Nature* **522**, 340–344 (2015).

136.   Peng, C. *et al.* Cellular milieu imparts distinct pathological α-synuclein strains in α-synucleinopathies. *Nature* **557**, 558–563 (2018).

137.   Braak, H. *et al.* Staging of brain pathology related to sporadic Parkinson's disease. *Neurobiol Aging* **24**, 197–211 (2003).

138. Braak, H., R�b, U., Gai, W. P. & Del Tredici, K. 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**, 517–536 (2003).

139. Stokholm, M. G., Danielsen, E. H., Hamilton-Dutoit, S. J. & Borghammer, P. Pathological α-synuclein in gastrointestinal tissues from prodromal Parkinson disease patients: α-Synuclein in Prodromal PD. *Ann Neurol.* **79**, 940–949 (2016).

140. Li, B. *et al.* Cryo-EM of full-length α-synuclein reveals fibril polymorphs with a common structural kernel. *Nature Communications* **9**, 3609 (2018).

141. Kramer, M. L. & Schulz-Schaeffer, W. J. Presynaptic -Synuclein Aggregates, Not Lewy Bodies, Cause Neurodegeneration in Dementia with Lewy Bodies. *J Neurosci* **27**, 1405–1410 (2007).

142. Dehay, B. *et al.* Targeting α-synuclein for treatment of Parkinson's disease: mechanistic and therapeutic considerations. *Lancet Neurol* **14**, 855–866 (2015).

143. Hauser, R. A. α-Synuclein in Parkinson's disease: getting to the core of the matter. *Lancet Neurol* **14**, 785–786 (2015).

144. Kingwell, K. Zeroing in on neurodegenerative α-synuclein. *Nat Rev Drug Discov* **16**, 371–373 (2017).

145. Mikitsh, J. L. & Chacko, A.-M. Pathways for Small Molecule Delivery to the Central Nervous System across the Blood-Brain Barrier. *Perspect Medicin Chem* **16**, 11–24 (2014).

146. Li, J., Zhu, M., Manning-Bog, A. B., Di Monte, D. A. & Fink, A. L. Dopamine and L-dopa disaggregate amyloid fibrils: implications for Parkinson's and Alzheimer's disease. *FASEB J* **18**, 962–964 (2004).

147. Norris, E. H. *et al.* Reversible Inhibition of α-Synuclein Fibrillization by Dopaminochrome-mediated Conformational Alterations. *JBC* **280**, 21212–21219 (2005).

148. Herrera, F. E. *et al.* Inhibition of a-Synuclein Fibrillization by Dopamine Is Mediated by Interactions with Five C-Terminal Residues and with E83 in the NAC Region. *PlosOne* **3**, e3394 (2008).

149. Mor, D. E. *et al.* Dopamine induces soluble α-synuclein oligomers and nigrostriatal degeneration. *Nature Neurosci* **20**, 1560–1568 (2017).

150.   Mor, D. E., Daniels, M. J. & Ischiropoulos, H. The usual suspects, dopamine and alpha-synuclein, conspire to cause neurodegeneration. *Mov. Disord.* **34**, 167–179 (2019).

151.   Ono, K. & Yamada, M. Antioxidant compounds have potent anti-fibrillogenic and fibril-destabilizing effects for alpha-synuclein fibrils in vitro. *J Neurochem* **97**, 105–115 (2006).

152.   Masuda, M. *et al.* Small Molecule Inhibitors of α-Synuclein Filament Assembly [†]. *Biochemistry* **45**, 6085–6094 (2006).

153.   Dhouafli, Z. *et al.* Inhibition of protein misfolding and aggregation by natural phenolic compounds. *Cell Mol Life Sci* **75**, 3521–3538 (2018).

154.   Kujawska, M. & Jodynis-Liebert, J. Polyphenols in Parkinson's Disease: A Systematic Review of In Vivo Studies. *Nutrients* **10**, 642 (2018).

155.   Oliveri, V. Toward the discovery and development of effective modulators of α-synuclein amyloid aggregation. *European Journal of Medicinal Chemistry* **167**, 10–36 (2019).

156.   Ehrnhoefer, D. E. *et al.* EGCG redirects amyloidogenic polypeptides into unstructured, off-pathway oligomers. *Nat Struct Mol Biol* **15**, 558–566 (2008).

157.   Ahmad, B. & Lapidus, L. J. Curcumin Prevents Aggregation in α-Synuclein by Increasing Reconfiguration Rate. *JBC* **287**, 9193–9199 (2012).

158.   Zhu, M. *et al.* The Flavonoid Baicalein Inhibits Fibrillation of α-Synuclein and Disaggregates Existing Fibrils. *JBC* **279**, 26846–26857 (2004).

159.   Lorenzen, N. *et al.* How Epigallocatechin Gallate Can Inhibit α-Synuclein Oligomer Toxicity in Vitro. *JBC* **289**, 21299–21310 (2014).

160.   Bieschke, J. *et al.* EGCG remodels mature α-synuclein and amyloid-β fibrils and reduces cellular toxicity. *Proc Natl Acad Sci USA* **107**, 7710–7715 (2010).

161.   Singh, P. K. *et al.* Curcumin Modulates α-Synuclein Aggregation and Toxicity. *ACS Chem Neurosci* **4**, 393–407 (2013).

162.   Jiang, M. *et al.* Baicalein reduces E46K α-synuclein aggregation in vitro and protects cells against E46K α-synuclein toxicity in cell models of familiar Parkinsonism: Baicalein reduces α-syn aggregation and toxicity. *J Neurochem* **114**, 419–429 (2010).

163.	Yang, J. E. *et al.* EGCG-mediated Protection of the Membrane Disruption and Cytotoxicity Caused by the 'Active Oligomer' of α-Synuclein. *Sci Rep* **7**, (2017).

164.	Liu, Z., Yu, Y., Li, X., Ross, C. A. & Smith, W. W. Curcumin protects against A53T alpha-synuclein-induced toxicity in a PC12 inducible cell model for Parkinsonism. *Pharmacol Res* **63**, 439–444 (2011).

165.	Wang, M. S., Boddapati, S., Emadi, S. & Sierks, M. R. Curcumin reduces α-synuclein induced cytotoxicity in Parkinson's disease cell model. *BMC Neurosci* **11**, 57 (2010).

166.	Hu, Q. *et al.* Baicalein inhibits α-synuclein oligomer formation and prevents progression of α-synuclein accumulation in a rotenone mouse model of Parkinson's disease. *Biochim Biophys Acta* **1862**, 1883–1890 (2016).

167.	Sharma, N. & Nehru, B. Curcumin affords neuroprotection and inhibits α-synuclein aggregation in lipopolysaccharide-induced Parkinson's disease model. *Inflammopharmacology* **26**, 349–360 (2018).

168.	Choi, J. H. *et al.* Baicalein protects HT22 murine hippocampal neuronal cells against endoplasmic reticulum stress-induced apoptosis through inhibition of reactive oxygen species production and CHOP induction. *Exp Mol Med* **42**, 811–822 (2010).

169.	Zhang, X., Yang, Y., Du, L., Zhang, W. & Du, G. Baicalein exerts anti-neuroinflammatory effects to protect against rotenone-induced brain injury in rats. *Int Immunopharmacol* **50**, 38–47 (2017).

170.	Spinelli, K. J., Osterberg, V. R., Meshul, C. K., Soumyanath, A. & Unni, V. K. Curcumin Treatment Improves Motor Behavior in α-Synuclein Transgenic Mice. *PLoS One* **10**, e0128510 (2015).

171.	Chen, M. *et al.* Tea polyphenols alleviate motor impairments, dopaminergic neuronal injury, and cerebral α-synuclein aggregation in MPTP-intoxicated parkinsonian monkeys. *Neuroscience* **286**, 383–392 (2015).

172.	Li, Y. Y. & Jones, S. J. Drug repositioning for personalized medicine. *Genome Med* **4**, e1–e14 (2012).

173.	Pushpakom, S. *et al.* Drug repurposing: progress, challenges and recommendations. *Nature Reviews Drug Discovery* **18**, 41–58 (2019).

174.   Feng, Y., LoGrasso, P. V., Defert, O. & Li, R. Rho Kinase (ROCK) Inhibitors and Their Therapeutic Potential. *J Med Chem* **59**, 2269–2300 (2016).

175.   Bond, L. M., Sellers, J. R. & McKerracher, L. Rho kinase as a target for cerebral vascular disorders. *Future Med Chem* **7**, 1039–1053 (2015).

176.   Defert, O. & Boland, S. Rho kinase inhibitors: a patent review (2014 – 2016). *Expert Opin Ther Pat* **27**, 507–515 (2017).

177.   Tonges, L. *et al.* Inhibition of rho kinase enhances survival of dopaminergic neurons and attenuates axonal loss in a mouse model of Parkinson's disease. *Brain* **135**, 3355–3370 (2012).

178.   Tatenhorst, L. *et al.* Fasudil attenuates aggregation of α-synuclein in models of Parkinson's disease. *Acta Neuropathol Commun* **4**, (2016).

179.   Moore, K. S. *et al.* Squalamine: an aminosterol antibiotic from the shark. *Proc Natl Acad Sci USA* **90**, 1354–1358 (1993).

180.   Rao, M. N. *et al.* Aminosterols from the Dogfish Shark Squalus acanthias. *J. Nat. Prod.* **63**, 631–635 (2000).

181.   Schlottmann, P. G. *et al.* New Treatment Modalities for Neovascular Age-Related Macular Degeneration. *Asia Pac J Ophthalmol (Phila)* **6**, 514–519 (2017).

182.   Yeung, T. *et al.* Membrane Phosphatidylserine Regulates Surface Charge and Protein Localization. *Science* **319**, 210–213 (2008).

183.   Sumioka, A., Yan, D. & Tomita, S. TARP phosphorylation regulates synaptic AMPA receptors through lipid bilayers. *Neuron* **66**, 755–767 (2010).

184.   Alexander, R. T. *et al.* Membrane surface charge dictates the structure and function of the epithelial Na+/H+ exchanger. *EMBO J.* **30**, 679–691 (2011).

185.   Fernández, C. O. *et al.* NMR of alpha-synuclein-polyamine complexes elucidates the mechanism and kinetics of induced aggregation. *EMBO J.* **23**, 2039–2046 (2004).

186.   Perni, M. *et al.* A natural product inhibits the initiation of α-synuclein aggregation and suppresses its toxicity. *Proc Natl Acad Sci USA* **114**, E1009–E1017 (2017).

187.   Perni, M. *et al.* Multistep Inhibition of α-Synuclein Aggregation and Toxicity in Vitro and in Vivo by Trodusquemine. *ACS Chem. Biol.* **13**, 2308–2319 (2018).

188. Harrington, C. R. *et al.* Cellular Models of Aggregation-dependent Template-directed Proteolysis to Characterize Tau Aggregation Inhibitors for Treatment of Alzheimer Disease. *J. Biol. Chem.* **290**, 10862–10875 (2015).

189. Melis, V. *et al.* Effects of oxidized and reduced forms of methylthioninium in two transgenic mouse tauopathy models. *Behav Pharmacol* **26**, 353–368 (2015).

190. Wischik, C. M., Schelter, B. O., Wischik, D. J., Storey, J. M. D. & Harrington, C. R. Modeling Prion-Like Processing of Tau Protein in Alzheimer's Disease for Pharmaceutical Development. *J. Alzheimers Dis.* **62**, 1287–1303 (2018).

191. Schwab, K. *et al.* A Protein Aggregation Inhibitor, Leuco-Methylthioninium Bis(Hydromethanesulfonate), Decreases α-Synuclein Inclusions in a Transgenic Mouse Model of Synucleinopathy. *Front Mol Neurosci* **10**, 447 (2017).

192. Congdon, E. E. *et al.* Methylthioninium chloride (methylene blue) induces autophagy and attenuates tauopathy in vitro and in vivo. *Autophagy* **8**, 609–622 (2012).

193. Mandal, S., Moudgil, M. & Mandal, S. K. Rational drug design. *Eur. J. Pharmacol.* **625**, 90–100 (2009).

194. Mavromoustakos, T. *et al.* Strategies in the rational drug design. *Curr. Med. Chem.* **18**, 2517–2530 (2011).

195. Macalino, S. J. Y., Gosu, V., Hong, S. & Choi, S. Role of computer-aided drug design in modern drug discovery. *Arch. Pharm. Res.* **38**, 1686–1701 (2015).

196. Sinha, S. *et al.* Lysine-specific molecular tweezers are broad-spectrum inhibitors of assembly and toxicity of amyloid proteins. *J. Am. Chem. Soc.* **133**, 16958–16969 (2011).

197. Attar, A. *et al.* Protection of primary neurons and mouse brain from Alzheimer's pathology by molecular tweezers. *Brain* **135**, 3735–3748 (2012).

198. Prabhudesai, S. *et al.* A novel 'molecular tweezer' inhibitor of α-synuclein neurotoxicity in vitro and in vivo. *Neurotherapeutics* **9**, 464–476 (2012).

199. Richter, F. *et al.* A Molecular Tweezer Ameliorates Motor Deficits in Mice Overexpressing α-Synuclein. *Neurotherapeutics* **14**, 1107–1119 (2017).

200. Attar, A., Chan, W.-T. C., Klärner, F.-G., Schrader, T. & Bitan, G. Safety and pharmacological characterization of the molecular tweezer CLR01 - a broad-spectrum inhibitor of amyloid proteins' toxicity. *BMC Pharmacol Toxicol* **15**, 23 (2014).

201. Wrasidlo, W. *et al.* A de novo compound targeting α-synuclein improves deficits in models of Parkinson's disease. *Brain* **139**, 3217–3236 (2016).

202. Price, D. L. *et al.* The small molecule alpha-synuclein misfolding inhibitor, NPT200-11, produces multiple benefits in an animal model of Parkinson's disease. *Sci Rep* **8**, (2018).

203. Uversky, V. N. A Protein-Chameleon: Conformational Plasticity of α-Synuclein, a Disordered Protein Involved in Neurodegenerative Disorders. *J Biomol Struct Dyn* **21**, 211–234 (2003).

204. Tsigelny, I. F., Sharikov, Y., Miller, M. A. & Masliah, E. Mechanism of alpha-synuclein oligomerization and membrane interaction: theoretical approach to unstructured proteins studies. *Nanomedicine* **4**, 350–357 (2008).

205. Tsigelny, I. F. *et al.* Role of α-synuclein penetration into the membrane in the mechanisms of oligomer pore formation: α-syn membrane penetration. *FEBS J* **279**, 1000–1013 (2012).

206. Giehm, L. & Otzen, D. E. Strategies to increase the reproducibility of protein fibrillization in plate reader assays. *Analytical Biochemistry* **400**, 270–281 (2010).

207. Pujols, J. *et al.* High-Throughput Screening Methodology to Identify Alpha-Synuclein Aggregation Inhibitors. *Int J Mol Sci* **18**, 478 (2017).

208. Kurnik, M. *et al.* Potent α-Synuclein Aggregation Inhibitors, Identified by High-Throughput Screening, Mainly Target the Monomeric State. *Cell Chem Biol* **25**, e1389–e1402 (2018).

209. Xu, M., Loa-Kum-Cheung, W., Zhang, H., Quinn, R. J. & Mellick, G. D. Identification of a New α-Synuclein Aggregation Inhibitor via Mass Spectrometry-based Screening. *ACS Chem Neurosci* (2019) doi:10.1021/acschemneuro.9b00092.

210. Wagner, J. *et al.* Anle138b: a novel oligomer modulator for disease-modifying therapy of neurodegenerative diseases such as prion and Parkinson's disease. *Acta Neuropathol.* **125**, 795–813 (2013).

211.   Deeg, A. A. *et al.* Anle138b and related compounds are aggregation specific fluorescence markers and reveal high affinity binding to α-synuclein aggregates. *Biochim. Biophys. Acta* **1850**, 1884–1890 (2015).

212.   Levin, J. *et al.* The oligomer modulator anle138b inhibits disease progression in a Parkinson mouse model even with treatment started after disease onset. *Acta Neuropathol.* **127**, 779–780 (2014).