



Universitat Autònoma de Barcelona

Departament de Bioquímica i Biologia Molecular

Institut de Biotecnologia i Biomedicina *Vicent Villar Palasí*

Characterization of intracellular protein aggregates

Anna Villar i Piqué

May 2013



Universitat Autònoma de Barcelona

Departament de Bioquímica i Biologia Molecular
Institut de Biotecnologia i Biomedicina *Vicent Villar Palasí*

Characterization of intracellular protein aggregates

Doctoral thesis presented by Anna Villar i Piqué for the degree of Ph.D in Biochemistry, Molecular Biology and Biomedicine from the Universitat Autònoma de Barcelona.

Thesis performed in the Department of Biochemistry and Molecular Biology and in the Institute of Biotechnology and Biomedicine, supervised by Prof. Salvador Ventura i Zamora.

Anna Villar i Piqué

Prof. Salvador Ventura i Zamora

Bellaterra, May 2013

*“No esperem el blat
sense haver sembrat,
no esperem que l’arbre doni fruits sense podar-lo;
l’hem de treballar,
l’hem d’anar a regar,
encara que l’ossada ens faci mal.*

Cal que neixin flors a cada instant

Lluís Llach

A la meva família...

Summary

During the last decades, protein aggregation has become a dynamic research topic extending across distinct investigation fields, including biochemistry, biotechnology, biomedicine and nanotechnology. On one side, the accumulation of proteins into insoluble amyloid deposits constitutes a common hallmark of many human disorders, known as conformational diseases. On the other side, from a biotechnological point of view, protein deposition is regarded as a usual hindrance in the production of recombinant proteins, which usually assemble into intracellular inclusion bodies. Although inclusion bodies were traditionally considered unstructured particles lacking of interest, the increasing number of evidences indicating that these aggregates contain amyloid-like structure pave the way for employing them in the study of amyloid deposition.

The thesis presented here recapitulates the work belonging to a set of publications concerning amyloid protein aggregation in the intracellular space. In three of these works, we use three distinct cellular models phylogenetically distant (bacteria, yeast and plants) to address the formation of protein deposits with the aim to characterize them and to analyze their impact in the cellular metabolism. In a complementary publication, we exploit bacterial aggregates to develop an *in vitro* screening assay for amyloid modulators. Finally, we also include two revision works about protein deposition in bacteria and its role as model in the study of amyloid aggregation.

The data collected from these studies indicate that aggregation into amyloid structures is a general property of polypeptides and a ubiquitous phenomenon in Nature. However, the apparition of these deposits in the cellular environment can be accompanied by a certain degree of toxicity. Here, we analyze the aging effect promoted by intracellular inclusion bodies in bacterial cells. In addition, we report how both prokaryotic and eukaryotic cells are endowed with a powerful protein quality machinery to challenge this damaging scenario. Moreover, the deep characterization of bacterial protein aggregates

and their formation process permits their use as a tool in the searching for amyloid aggregation inhibitors with putative biomedical and pharmaceutical interest.

Overall, this thesis delves into the study of amyloid protein aggregation and adds insights to employ simple organisms as relevant cellular models.

Contents

Summary	XIII
Table of Contents	XV
List of Publications	XVII
List of Figures	XIX
Abbreviations	XXI
1 Introduction	1
1.1 Proteins	1
1.1.1 Protein Folding	1
1.1.1.1 Acquisition of the native structure	1
1.1.1.2 Conformational Space	3
1.1.1.3 Folding control in bacteria	4
1.1.2 Intrinsically disordered proteins	5
1.2 Protein aggregation	7
1.2.1 Protein misfolding and conformational diseases	7
1.2.2 Determinants of protein aggregation	8
1.2.3 Prediction of protein aggregation	11
1.2.4 Natural selection against protein aggregation	12
1.2.5 Amyloid fibrils	14
1.3 Bacterial Inclusion bodies	15
1.3.1 Structure of inclusion bodies	15
1.3.2 Formation of Inclusion bodies	17
1.3.3 Segregation of IBs during bacterial division	18
1.3.4 Biotechnological exploitation of IBs	19
1.3.4.1 IBs contain active conformations	19
1.3.4.2 IBs in screening assays	20
1.4 Models used in the present thesis	22

1.4.1	Amyloid β -peptide	22
1.4.2	Cellular models	23
1.4.2.1	<i>Escherichia coli</i>	23
1.4.2.2	<i>Saccharomyces cerevisiae</i>	24
1.4.2.3	Plants	26
2	Aims	29
3	Publications	31
3.1	Publication I	31
3.2	Publication II	45
3.3	Publication III	59
3.4	Publication IV	61
3.5	Publication V	73
3.6	Publication VI	79
4	Discussion	119
4.1	Protein aggregates in unicellular organisms	119
4.2	Amyloid aggregation promotes aging	120
4.3	Cellular response to protein aggregation	121
4.4	Bacterial IBs in the screening of amyloid modulators	123
4.5	Intracellular amyloids in plants	124
5	Conclusions	127
	References	129
	Agraïments (Acknowledgments)	139
	Appendices	141

List of publications

The present thesis is based on the following publications:

- I **A. Villar-Piqué**, N.S. de Groot, R. Sabate, S.P. Acebrón, G. Celaya, X. Fernàndez-Busquets, A. Muga and S. Ventura. (2011) The effect of amyloidogenic peptides on bacterial aging correlates with their intrinsic aggregation propensity. *J. Mol. Biol.* 421(2-3): 2710-81.
- II **A. Villar-Piqué**, A. Esparagaró, R. Sabaté, N.S. de Groot and S. Ventura. (2012) Using bacterial inclusion bodies to screen for amyloid aggregation inhibitors. *Microb. Cell Fact.* 11(1): 55-65.
- III **A. Villar-Piqué** and S. Ventura. (2013) Linking intracellular protein aggregation propensity and protein quality control degradation. (manuscript for incoming submission)¹
- IV **A. Villar-Piqué**, R. Sabaté, O. Lopera, J. Gibert, J. M. Torne, M. Santos and S. Ventura. (2010) Amyloid-like protein inclusions in tobacco transgenic plants. *PLoS One.* 5(10): e13625.
- V **A. Villar-Piqué** and S. Ventura. (2012) Modeling amyloids in bacteria. *Microb. Cell Fact.* 11: 166-168.
- VI **A. Villar-Piqué** and S. Ventura. (2013) Inclusion bodies in the study of amyloid aggregation. Protein Aggregation in Bacteria: Functional and Structural Properties of Inclusion Bodies in Bacterial Cells. Editor: M. Lotti. *John Wiley and Sons, Inc.* Hoboken, New Jersey, USA. (in press)

¹ This publication is included in the thesis as annex.

XVIII List of Publications

Other works published by the author of the present thesis:

- **A. Villar-Piqué** and S. Ventura. (2012) Protein aggregation acts as strong constraint during evolution. *Evolutionary Biology, Mechanisms and Trends*: 103-120. Editor: P. Pontarotti. *Springer* Heidelberg, Germany.
- N.S. de Groot, M. Torrent, **A. Villar-Piqué**, B. Lang, S. Ventura, J. Gsponer and M.M. Babu. (2012) Evolutionary selection for protein aggregation. *Biochem. Soc. Trans.* 40(5): 1032-7.
- A. Espargaró, **A. Villar-Piqué**, R. Sabaté and S. Ventura. (2012) Yeast prions form infectious amyloid inclusion bodies in bacteria. *Microb. Cell Fact.* 11(1): 89.
- R. Sabaté, **A. Villar-Piqué**, A. Espargaró and S. Ventura. (2011) Temperature dependence of the aggregation kinetics of Sup35 and Ure2p yeast prions. *Biomacromolecules.* 13(2): 474-83.
- M.A. Anwar, S. Kralj, **A. Villar Piqué**, H. Leemhuis, M.J.E.C. van der Maarel and L. Dijkhuizen. (2010) Inulin and levan synthesis by probiotic *Lactobacillus gasseri* strains: Characterization of three novel fructansucrase enzymes and their fructan products. *Microbiology.* 156(Pt4): 1264-74.

List of Figures

1.1	Proteinogenic amino acids	2
1.2	Protein structure	3
1.3	Protein folding landscape	4
1.4	Molecular chaperones	6
1.5	Conformational diversity of disease related proteins	9
1.6	Amino acid relative frequencies in selected positions	10
1.7	Amyloid fibrils structure	14
1.8	Amyloid fibrils formation	16
1.9	Protein conformations during recombinant expression	18
1.10	Distribution of proteins in the IBs	19
1.11	Segregation of IBs during cell division	20
1.12	Example of an IBs-based screening assay for amyloid inhibitors	21
1.13	Histological samples from AD patients	22
1.14	APP processing	23
1.15	Yeast models of conformational diseases	27
4.1	Primary sequence as a determinant of aggregation	126

Abbreviations

In alphabetical order

Abbreviation	Meaning
A β 42	42 residues length amyloid- β peptide
AD	Alzheimer's disease
APP	Amyloid precursor protein
α -syn	α -synuclein
ATP	Adenosine-5'-triphosphate
Clp	Caseinolytic peptidase
DNA	Deoxyribonucleic acid
<i>E. coli</i>	<i>Escherichia coli</i>
GFP	Green fluorescent protein
GroEL	Large <i>groE</i> gene product
GroES	Small <i>groE</i> gene product
GrpE	<i>GroP</i> -like gene E product
HS	Hot spot
HD	Huntington's disease
Htt	Huntingtin
IB	Inclusion body
IDP	Intrinsically disordered protein
PD	Parkinson's disease
RNA	Ribonucleic acid
<i>S. cerevisiae</i>	<i>Saccharomyces cerevisiae</i>
TGZ	Transglutaminase
ThS	Thioflavin S
ThT	Thioflavin T
<i>wt</i>	<i>Wild-type</i>

Introduction

1.1 Proteins

Proteins are the main executors of a wide range of cellular functions and constitute indispensable structural elements, being thus essential players of life. Their tremendous diversity, functional versatility and plasticity are determined by the huge amount of spatial conformations they can adopt. The central dogma in biochemistry states that genetic information is stored in DNA, which, in turn, is transcribed to RNA molecules under a strict cellular control. Messenger RNA molecules act as information transporters that will be translated by ribosomes to give rise to polypeptide sequences, which are constituted by the linear combination of amino acids through peptide bonds [1]. There exist twenty natural amino acids, each of them with a different side chain that provides them with distinct physico-chemical properties (Fig. 1.1).

1.1.1 Protein Folding

1.1.1.1 Acquisition of the native structure

Protein folding is the process through a polypeptide sequence achieves its unique three dimensional conformation, known as the native state of a protein. Specific intramolecular contacts (disulfide bridges, hydrogen bonds or hydrophobic and ionic interactions) are codified in the primary sequence and determine and stabilize the spatial conformation of the protein [2]. Starting from the unfolded state, the number of different conformations that a protein can explore before adopting the native structure is tremendously high, making not feasible to obtain it in a biologically relevant timescale by a simple random search. Rather, during protein folding, there exist preferential kinetic pathways, where native interactions are more stable than non-native ones, thus limiting the number of accessible conformations [3].

In a polypeptide chain, local interactions form secondary structure elements, which can be regular such as α -helix and β -sheets, or irregular such as

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

Fig. 1.1. Table with the 20 natural proteinogenic amino acids. They are coloured depending on the physico-chemical properties of the side chain: hydrophobic are in blue, aromatic are in pink, negatively charged are in orange, positively charged are in green, polar uncharged are in yellow and special cases are in grey.

coiled coils. Spatial combination of these structures gives rise to the three dimensional conformation (the tertiary structure), glued by both short and long range interactions along the sequence. These molecular contacts permit local and tiny fluctuations usually indispensable for proteins to exert their biolog-

ical functions. Furthermore, structured polypeptides might establish native intermolecular contacts and assemble to form the quaternary structure of homomeric or multimeric proteins (Fig. 1.2) [4].

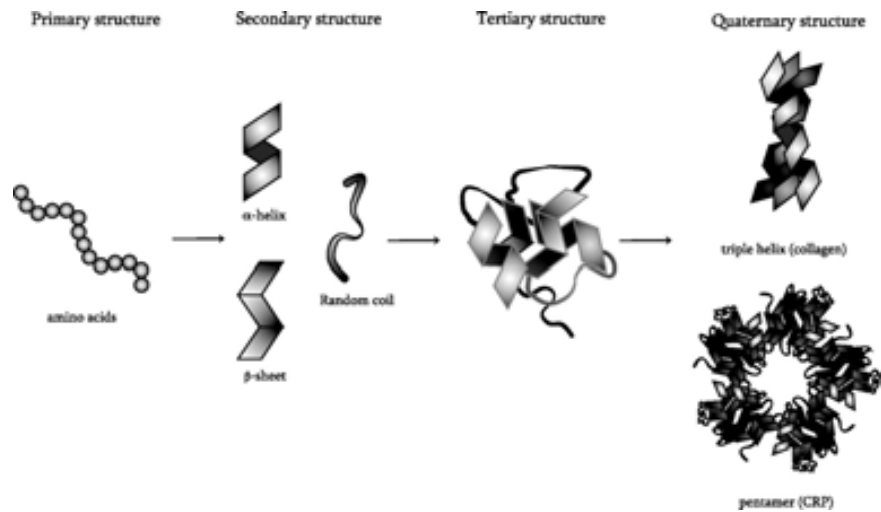


Fig. 1.2. Representation of the structural components of proteins, from the polypeptide chain to the formation of multimeric proteins [4].

1.1.1.2 Conformational Space

The conformational space of a protein includes all the possible conformations and routes that a polypeptide can sample during the folding process. This is a theoretical concept that can be visualized through *energy landscapes* (Fig. 1.3). These diagrams represent the different folding and intermediate states as a group of conformations and assume multiple and parallel folding pathways. They are normally simplified as bidimensional illustrations where the high energy space is widely populated by unfolded species and, in the low energy space, only few structures are present, basically native or native-like conformations and aggregated states depending on the type of interactions: intramolecular contacts lead to the native state whereas non-native intermolecular contacts lead to protein aggregation [5, 6]. The shape of the energy landscape depends on the polypeptide sequence and the environmental conditions. In this sense, large proteins tend to have more irregular and crooked landscapes representing the complexity of their folding pathways and the numerous partially folded conformations along the reaction [7].

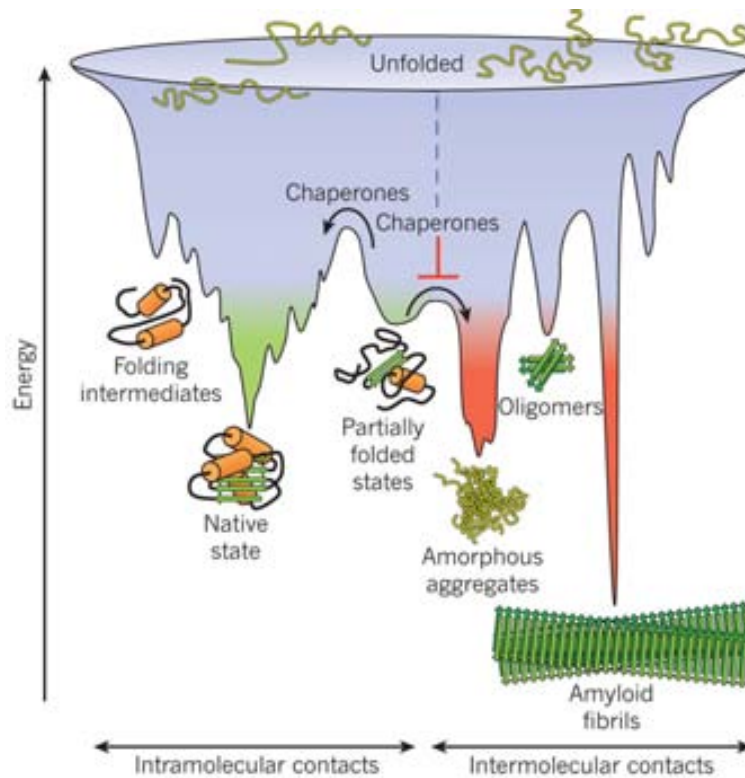


Fig. 1.3. Schematic example of a protein folding landscape where the above space is populated by high energy unfolded species. Intramolecular contacts lead to the native state (green), while non-native intermolecular ones generate the appearance of protein aggregates (red). Molecular chaperones assist partially folded species and folding intermediates to favour and accelerate the acquisition of the native state while preventing protein deposition [6].

1.1.1.3 Folding control in bacteria

Protein folding in the cellular environment is a finely tuned process, as even a minimum failure in the process can trigger aggregation through the appearance of misfolded or partially unfolded species. Hence, cells have been evolutionary provided with a powerful protein quality machinery, mainly composed by molecular chaperones and proteases that cope with this problem to maintain protein homeostasis.

On one side, chaperones are proteins able to specifically assist folding of non-native species and prevent their aggregation (Fig. 1.3 and 1.4). In bacteria, most of them are transcriptionally controlled by factor σ_{32} , whose synthesis, stability and activity are regulated by the ratio between chaperones and unfolded polypeptides [8]. Among molecular chaperones, *trigger*

factor binds to the ribosome and rapidly assists the folding of the nascent amino acid chain, being the first chaperone to interact with the polypeptide. Subsequently, two cytosolic chaperone complexes may act on misfolded or unfolded proteins: DnaK-DnaJ-GrpE and GroEL-GroES. Both complexes recognize unfolded chains through solvent-exposed hydrophobic regions and assist their folding through consecutive binding-releasing cycles until the protein achieves its native structure. Besides, the binding of the unfolded polypeptides to chaperones structurally blocks their aggregation. Although both complexes actively exert their function consuming ATP, GroEL-GroES can also passively assist folding of small proteins, which are encapsulated in the complex and protected from the external media and, therefore, the correct folding results thermodynamically favoured [6, 9].

On the other side, cytosolic proteases play a role on protein homeostasis by degrading off-pathway misfolded species unable to adopt the native state. One of the most important groups of bacterial proteases are Clp ATPases; they act as proteolytic machines, that similarly to the proteasome, function via a two-step process: first, the polypeptide is completely unfolded and afterwards, it is irreversibly degraded. Among this kind of enzymes, ClpB represents a relevant special case, whose binding to protein aggregates allows the extraction of assembled proteins with the posterior release of soluble species that might become targets for the folding machinery. Interestingly, the fact that ClpB activity is mediated by DnaK and that GroEL-GroES complex is also involved in proteolysis, suggests that both chaperones and proteases activities overlap to a certain extent in the cellular metabolism [9, 10].

1.1.2 Intrinsically disordered proteins

Although for a long time it was assumed that proteins need to fold into a three dimensional structure to be active, this *structure-function* dogma partially faded away with the discovery of a large set of proteins displaying a disordered structure in physiological conditions. These so-called *Intrinsically Disordered Proteins* (IDPs) are characterized by their inability to fold into a unique and stable conformation, rather they display a vast array of function-related structural organization. The term *disorder* also applies to regions and domains of larger proteins [11]. The amino acid composition of IDPs is substantially different from the primary sequence of ordered proteins. While the latter conceal hydrophobic stretches in the core of the tertiary structure, IDPs have been evolutionarily forced to increase their high net charge and decrease their overall hydrophobicity in order to diminish their risk of aggregation. Thus, they are particularly enriched in polar and charged residues as well as disorder-promoting ones (Asp, Met, Lys, Arg, Ser, Gln, Pro and Glu) [12].

IDPs exert an important number of crucial biological activities complementary to the functional repertoire of structured proteins, being principally involved in cell signaling, regulatory and recognition activities. Actually, their lack of a defined structure, together with their high flexibility, are important

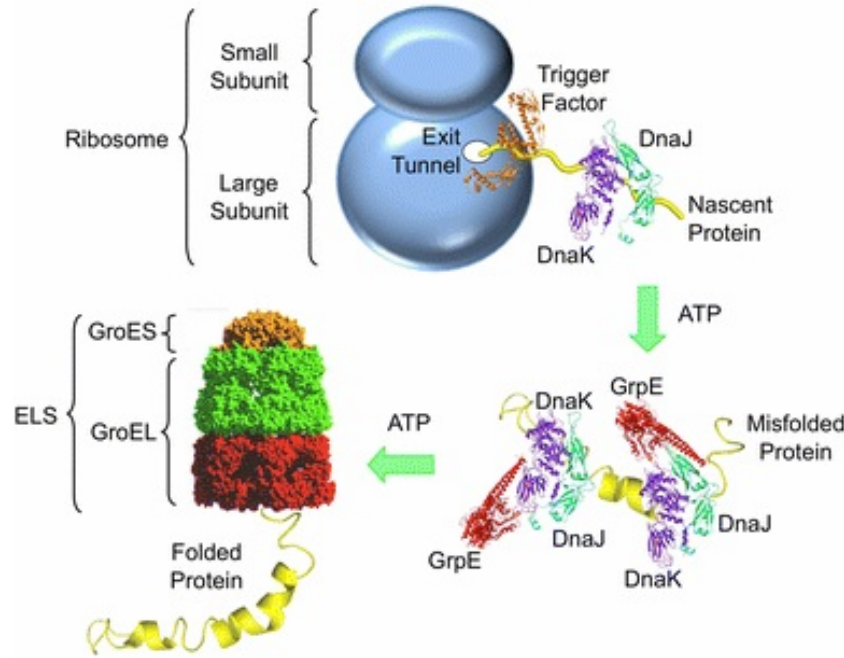


Fig. 1.4. Schematic representation of the protein folding assistance by molecular chaperones [9].

features for their functions providing conformational plasticity. These proteins can fold under a variety of conditions, including interactions with other proteins, nucleic acids, membranes and small molecules, as well as under environmental changes. Moreover, structural properties of IDPs permit interactions with multiple partners essential for organization, maintenance and control of complex protein-protein interaction networks [11, 13].

IDPs are present across the three domains of life, about 10-35 % of prokaryotic proteins and about 15-45 % of eukaryotic proteins contain disordered regions of at least 30 residues. The strong correlation between IDPs and signaling and regulatory activities is thought to be the cause that they are more present in eukaryotes, suggesting a possible co-evolution with these organisms [11, 14].

Finally, IDPs have gained a special relevance from a biomedical point of view, since they appear to be strongly linked to human disorders, including cancer, cardiovascular diseases and diabetes. Importantly, IDPs are especially associated to neurodegeneration, where over-expression and mutations in tau, α -synuclein (α -syn), β -amyloid peptide and huntingtin (htt) increase the risk of amyloid aggregation linked to pathological conditions [15]. Actually, this is one of the reasons thought to underlie the tight regulation of IDPs cellular levels. This regulation includes a finely tuned control of the transcription and

degradation that determines the half-life of IPDs, which is important since they are encoded by dosage-sensitive genes [16].

Due to the implication of IDPs in human diseases, they have emerged as a promising group of target proteins for therapeutic drug development. On one hand, as IDPs usually acquire a tertiary structure upon binding to their partner, drugs targeting the interaction surfaces would prevent the formation of pathological complexes [13]. On the other hand, alternative drug design strategies have been proposed based on targeting the regulatory enzymes and, therefore, affecting the availability of IDPs [16].

1.2 Protein aggregation

1.2.1 Protein misfolding and conformational diseases

The cellular cytoplasm is a crowded space where the total concentration of macromolecules can exceed 400 mg/mL, occupying between 5 % and 40 % of the total intracellular volume [17]. In this scenario, protein folding becomes a challenge where continuous collisions between molecules, favoured by the high concentration, result in an intense competition between intramolecular and intermolecular contacts [7], the latter leading to multiple forms of protein aggregates (Fig. 1.3). Hence, in a living cell, despite being endowed with a protein quality machinery (Section 1.1.1.3), the failure of a protein to adopt or maintain the native conformation provokes the appearance of misfolded or partially unfolded forms that might act as intermediates of aggregation. This altered condition has a pathological implication since several devastating human diseases are linked to protein aggregation.

The aberrant accumulation of protein deposits in different tissues and organs represents a common hallmark of a set of disorders, the so-called conformational diseases. Most of them involve the formation of highly ordered and insoluble deposits known as amyloid fibrils that might be located either in the extracellular space or forming intracellular inclusions. The number of conformational diseases is continuously increasing, including common disorders such as Alzheimer's disease (AD), Parkinson's disease (PD), Huntington's disease (HD) and type II diabetes. They can be grouped into: i) neurodegenerative conditions, in which aggregation occurs in the brain, ii) non-neuropathic localized amyloidosis, in which aggregation occurs in a single type of tissue except the brain, iii) and non-neuropathic systemic amyloidosis, in which aggregation occurs in multiple tissues. The origin of these diseases is sporadic in the 85 % of the cases, hereditary in the 10 % and transmissible in the 5 %, which is the case of transmissible spongiform encephalopathies, caused by prions [15].

The proteins involved in the formation of amyloid fibrils associated to human diseases are not related neither in sequence nor structure, according to the idea that amyloid fibrils are a low-energy state theoretically attainable by all the polypeptides (Fig. 1.3). The native structures of these disease-related

proteins range from IDPs, through α -helix rich polypeptides, to proteins containing β -sheet structure exclusively or in part. They also include monomeric proteins, small peptides and proteins found naturally as multimers (Fig. 1.5) [7].

1.2.2 Determinants of protein aggregation

Protein aggregation is mainly determined by the primary sequence, the conformational stability and external factors such as temperature, pH or protein concentration. About the sequential determinants of aggregation, mutagenesis experiments have been a traditional approach to evaluate the influence of the physico-chemical properties of the amino acid side chains over the aggregation propensity of the primary sequence. In this sense, it has been demonstrated that hydrophobicity, β -sheet propensity and low presence of charged residues are properties directly correlated with the tendency to aggregate. Among them, hydrophobicity has been considered the probably main sequential determinant of aggregation [18]. Because hydrophobic collapse is one of the most important forces involved in protein folding, hydrophobic residues are commonly present in polypeptide sequences providing certain aggregation propensity. Besides, the crucial role of the net charge of a polypeptide in its aggregation tendency might be easily explained since charged residues produce electrostatic repulsion forces that impair the establishment of intermolecular contacts. Accordingly, the increase in the net charge is negatively correlated with the aggregation propensity [19]. Importantly, the net charge of a protein is not only determined by the amino acid composition, but also depends on the environmental pH. The proximity between the pH and the isoelectric point of a protein decreases its number of charges favouring the non-specific interactions leading to self-assembly [20].

A relevant discovery has been that not all the polypeptide sequence is equally important for protein aggregation; rather there exist short amyloidogenic stretches that act as facilitators of protein self-assembly [21–23]. These regions, referred to as *Hot spots* (HSs), modulate the aggregation propensity of the entire sequence and are able to trigger the deposition process. They are enriched in hydrophobic residues (Ile, Leu and Val) and aromatic ones (Phe, Trp and Tyr), while charged residues (Arg, Lys, Asp and Glu) as well as Pro are poorly represented (Fig. 1.6) [24].

On one hand, in globular proteins, HSs are usually buried in the hydrophobic core of the three dimensional structure. This strategic position diminishes the risk of self-assembly while contributing to the network of native contacts that stabilizes the native state. Moreover, they are also present at interaction surfaces of multimeric proteins [25]. These observations corroborate that the acquisition of the native state competes against aberrant interactions in the cell, and that the former can be considered as an evolutive mechanism to prevent aggregation (Section 1.2.4). Therefore, the conformational stability of globular proteins represents an important determinant of aggregation.

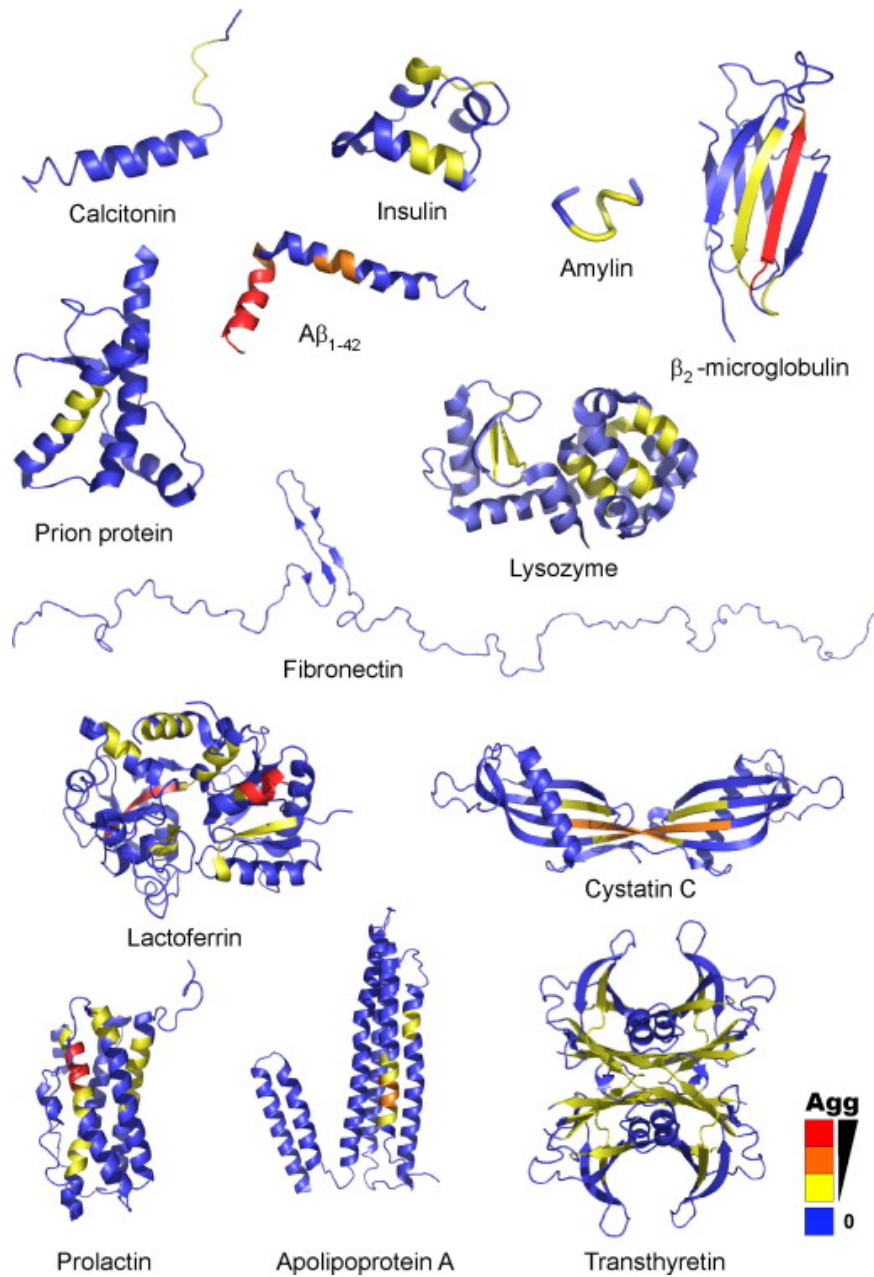


Fig. 1.5. Examples of proteins involved in conformational diseases. They are coloured according to their aggregation propensity: blue indicates no tendency to aggregation while warm colours represent the different propensities to self-assembly. Despite their conformational diversity, all of them are reported to form amyloid structures in pathological conditions [7].

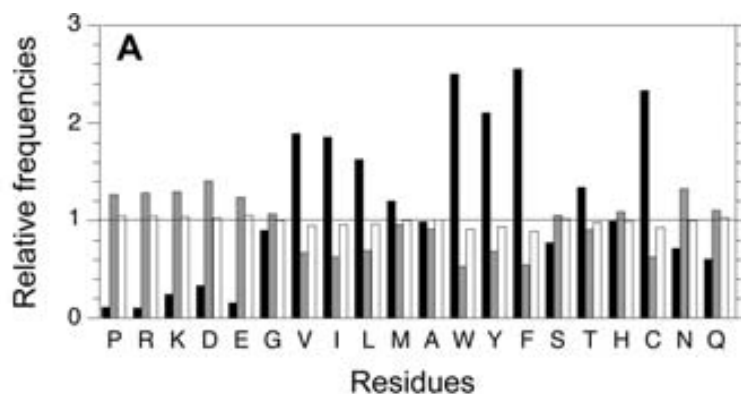


Fig. 1.6. Amino acid frequencies at different positions relative to their global frequencies in the human proteome. A relative frequency of 1 means that the residue occupies that position with the same frequency to that in the whole proteome. Black: inside HSs. Grey: at the HSs flanking positions (gatekeepers). White: outside and far from HSs. [24].

The establishment of intermolecular interactions requires a partial unfolding, such that HSs become unprotected and exposed to solvent [7]. Although folding intermediates appear during protein expression when the bulk of nascent polypeptides saturates the cellular protein quality machinery, local fluctuations of the native state are also responsible for the appearance of partially unfolded species. Tiny thermal conformational fluctuations provoke the transitory exposure of HSs, thus enabling the establishment of non-native contacts. Accordingly, the aggregation propensity of globular proteins correlates with their conformational stability [26, 27].

On the other hand, the aggregation propensity provided by HSs is commonly counteracted by the presence of *gatekeeper* residues, which are very short regions located at the flanking positions of HSs. Their main role is the reduction of the effect of HSs, thus preventing aberrant self-assembly. Gatekeepers are enriched in charged residues and proline, which plays a β -sheet disruptor role (Fig. 1.6) [28, 29]. Interestingly, unrelated chaperone families display substrate specificity for gatekeepers, representing a secondary and redundant mechanism by which these short stretches prevent protein deposition [30].

Finally, in the case of IDPs, the lack of a three dimensional structure turns the primary sequence into the principal determinant of aggregation. Accordingly, the analysis of the amino acid sequence of these proteins reveals a poor frequency of HSs [31]. However, the role of disordered regions in proteins as a determinant of aggregation remains an outstanding issue. On one side, their particular flexibility and the entropy penalty associated to the conformational change act against aggregation [32]. On the other side, in

prion proteins, prion domains overlap with disordered regions, suggesting a possible role of these stretches in promoting aggregation [15].

1.2.3 Prediction of protein aggregation

The broad knowledge achieved on the common determinants underlying protein deposition has been exploited to develop several computational methods to *in silico* predict the aggregation propensity. Up to date, approximately twenty algorithms have been published. They can be divided into: i) empirical or phenomenological methods, which make predictions by using physico-chemical properties (hydrophobicity, β -propensity and solubility) of the residues that compose the polypeptide chain, and ii) structure-based algorithms, which estimate the amyloid aggregation propensity by comparing sequences with the aggregation determinants in resolved amyloid structures [33].

The first predictive method totally based on *in vivo* empirical information was AGGRESCAN, which was developed by analysing the deposition properties of a library of point mutants of amyloid- β peptide fused to the green fluorescent protein (A β 42-GFP) in the *Escherichia coli* cytoplasm. After calculating their relative intracellular solubility, intrinsic aggregation propensity values were obtained for the 20 natural amino acids. AGGRESCAN was implemented as a web server that exploits this scale of amino acid aggregation propensities to analyse primary sequences in order to provide an aggregation propensity profile according to the relative amino acids position in the sequence. The predictor allows determining the overall aggregation propensity, the presence of HSs of aggregation or the effect of point mutations on the deposition properties [34].

Among other empirical algorithms, a pioneering one was TANGO, which estimates the aggregation propensity considering the tendency to acquire cross- β structure. The method scores the propensity of each residue in a polypeptide sequence to form different secondary structure elements and takes into account environmental variables, such as pH, temperature and ionic strength [35]. WALTZ is another bioinformatic tool developed from the combination between the empirical information of TANGO and structural studies of a large set of amyloid hexapeptides. A relevant advantage of WALTZ is the distinction between amyloid fibrils and amorphous β -sheet aggregates [36]. Another empirical predictor is SALSALSA, that attempts to detect regions with high propensity to form β -strands, assuming that this structure is strongly involved in the formation of fibrillar aggregates [37]. Finally, it is also worthy to mention ZYGREGATOR, which predicts the aggregation propensity taking into account the amino acid physico-chemical properties and environmental conditions where self-assemble reaction is expected to occur. A differential trait of this predictor is its ability to approximate the aggregation properties of globular proteins by considering the structural state of the protein besides its primary sequence [38].

On another side, methods classified as structure-based are very heterogeneous since they are based on a wide range of structural properties found in amyloid aggregates. One example is FoldAmyloid, which makes predictions from the steric elements involved in the formation of β -structures. Basically, it simultaneously considers two properties for each residue: the expected packing density and the probability of hydrogen bond formation (both the probability to act as donor and acceptor in the backbone-backbone hydrogen bond) [39]. On the contrary, BETASCAN is a structure-based algorithm derived from a selected set of structures of amphipathic β -sheets described in the bibliography. The method determines the preference of each pair of amino acids to be hydrogen bonded in a β -sheet, assuming that the likelihood of assembly of individual β -strands into ordered β -sheets is also determined by the primary sequence [40].

Although these tools are usually capable of accurately predict *in vitro* aggregation, the challenging question arises when studying aggregation in the complex cell environment [41, 42]. A comparative study was carried out with some of these predictors and showed good correlations with experimental data for some of them, indicating the usefulness of *in silico* aggregation prediction [42]. However, it is important to notice their tendency to overpredict in some cases [43]. A possible strategy to overcome this issue is to develop consensus methods based on the combination of previous algorithms [33]. An example of this new approach is AMYLPRED2, built up from the combination of eleven published predictors [44].

1.2.4 Natural selection against protein aggregation

Despite the deposition of insoluble protein aggregates is strongly linked to several human disorders, protein self-assembly is not restricted to a set of disease-related polypeptides, rather there are clear evidences indicating that the ability to establish non-native intermolecular contacts and achieve low energy aggregational states is a generic property of all polypeptides [7, 15]. Thus, it is likely that living systems have evolved mechanisms to prevent aggregation, maintaining proteins in their active conformations. In section 1.1.1.3, the main bacterial mechanisms of *in vivo* folding control, including molecular chaperones and proteases, have been introduced. However, polypeptides themselves have also been evolutionary optimized to avoid their self-assembly and, specifically, to prevent the formation of amyloid-like structures [45].

Since folding and aggregation compete in the cell, a primary way to avoid protein deposition is to favour the establishment of intramolecular contacts leading to the acquisition of the native conformation. Thus, in the evolution of proteins, foldability exerts a strong constraint, accelerating the folding kinetics and increasing the thermodynamic stability of the final state. On this regard, protein sequences have evolutionary reduced their folding time, which is reflected in the tendency of proteins to fold faster than their structural

ancestors [46]. This represents an important strategy to diminish the half-life of aggregation-prone folding intermediates. In this sense, shortening the polypeptide chain length has become a simple way to fold faster [46]. The presence of strategic disulfide bridges also increases the folding rate. These covalent links between cysteine residues crosslink regions that might be far away in the sequence, reducing the entropy of the unfolded species. Consequently, disulfide bridges accelerate the folding process, thus decreasing the aggregation rate. Furthermore, these bonds also prevent aggregation by means of the stabilization of the native state [47, 48]. Using a protein designed *de novo* without an evolutionary history, it has also been suggested that, in natural proteins, folding cooperatively into a unique globular state is the result of natural selection to escape aggregation [49].

Another mechanism to prevent aggregation is the formation of quaternary structures. The frequent overlapping between interaction surfaces and aggregation-prone regions indicates that both processes are regulated and promoted by similar physico-chemical properties. Therefore, the establishment of native interactions forming functional quaternary complexes exerts a protective role impeding the non-native self-assembly. Accordingly, destabilizing mutations at the interfaces usually result in the formation of protein deposits [25].

The pressure that natural selection exerts against protein aggregation has not only shaped protein structures, but also polypeptide sequences. In this line, some residues have been demonstrated to prevent aggregation. A deeply studied case is proline, whose particular side chain structure and stiffness endorses it with a β -sheet breaker ability. This property is extensively used by Nature to prevent the risk of self-assembly in β -structure. For example, selected proline residues are the cause that rat amylin is unable to form amyloid fibrils, in contrast to human amylin; or the high frequency of this residue in elastomeric proteins results essential to avoid aberrant aggregation [50, 51]. Glycine is another residue that has been evolutionary conserved in some sequences due to its role in preventing aggregation. Its conformational freedom generates a strict entropy penalty associated to the formation of β -structures [50, 52]. Furthermore, the positive selection for gatekeeper residues flanking amyloidogenic stretches represents yet another mechanism to escape from aggregation (Section 1.2.2) [28].

Finally, the regulation of protein abundance is also evolutionary constrained in order to diminish the likelihood of protein deposition. The concentration of a protein must be enough to perform its biological function but below a certain threshold to keep its native conformation without aggregating. In this sense, different studies have clearly concluded that highly abundant proteins are more soluble and have lower aggregation propensity [53–55]. Additionally, it has also been demonstrated a strong correlation between aggregation and protein turnover rates, where short-living proteins have experienced less evolutive pressure to diminish the aggregation risk than proteins with slow turnover rate [56].

1.2.5 Amyloid fibrils

A particular case of protein aggregates are amyloid fibrils. These deposits, firstly observed in different human tissues, were reported to have similar properties to cellulose in plants (which originated the word *amyloid*). Afterwards, they were found in histological samples from patients suffering from conformational diseases [57].

Amyloid fibrils are protein assemblies with a highly ordered and densely packed structure, whose main secondary structural element are β -sheets. They appear as straight unbranched fibrils with a diameter between 70 and 120 Å, generally composed of 2 to 6 smaller protofilaments, which might twist together to adopt a rope-like shape or be laterally connected to form long ribbons. Structural studies have revealed that polypeptide chains are embedded in the core of the fibrils in *cross- β motifs*. This element consists of arrays of β -strands arranged perpendicular to the longitudinal fibril axis (Fig. 1.7) [15, 58–60].

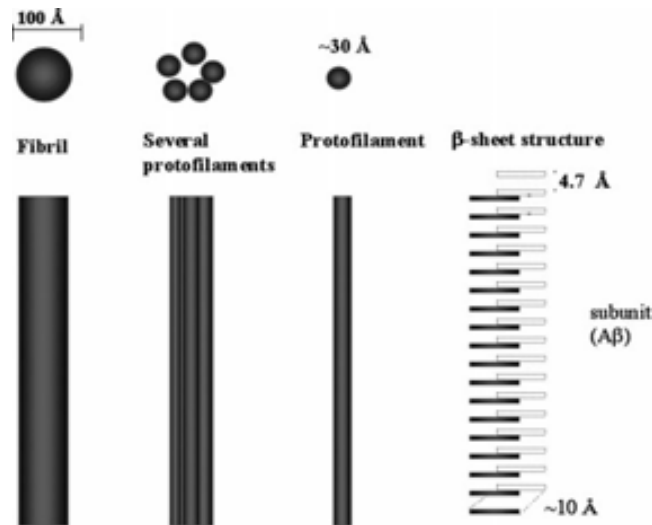


Fig. 1.7. Diagram of the hierarchical structures making up a mature amyloid fibril of A β peptide [60].

The formation of amyloid fibrils, a process known as amyloidogenesis, requires the initial exposure of HSs and, thus, in the case of globular proteins, a partial destabilization of the native state. Amyloid fibrils are formed in a nucleation-dependent fashion, where the transition from the soluble and monomeric state to the aggregated one is promoted by a pre-existing template or seed. This process is accompanied by a conformational change through which monomers assemblies adopt a fibrillar structure. Amyloidogenesis is a

concentration and time dependent reaction and exhibits three sequential steps: the lag or nucleation phase, the elongation or exponential phase and the stationary phase (Fig. 1.8) [61, 62]. During the lag phase, it takes place the formation of the nucleus through the association of soluble monomers. Due to a high entropic penalty, this process is thermodynamically disfavoured, being the rate-limiting step of the reaction. The lag phase might be abrogated by the addition of pre-formed aggregation nuclei of the same protein, which act as seeds for deposition (Fig. 1.8). Seeding experiments are usually performed to demonstrate the sequence specificity of amyloidogenesis. Later on, the population of these transient assemblies triggers the polymerization and fibril growth through monomer addition at the fibril ends, at the so-called elongation phase. This process occurs rapidly since it is highly favourable and gives rise to the mature fibrils [7]. As monomers are consumed to a certain level, the elongation phase slows down and the process comes into the stationary phase. Commonly, during this step the fibril adopts its mature form and coexists with a small amount of soluble monomers, which are in an assembly/disassembly equilibrium [61]. This equilibrium is a dynamic exchange mechanism that represents the recycling of molecules within the fibril population [63].

During amyloid fibril formation, intermediate structures can be isolated occurring as soluble oligomeric species formed by the assembly of a small number of monomers. They are normally flexible with high conformational heterogeneity, properties that prevent attaining accurate structural information. Oligomers have been commonly described as spherical particles exposing hydrophobic regions [65, 66]. Although they are compact, they possess a lower level of organization and lower β -structure content than the mature fibrils. Whether these oligomers are on-pathway or off-pathway intermediates is still a matter of debate [66]. Albeit mature fibrils might provoke deleterious cellular effects, increasing evidences suggest that pre-fibrillar oligomeric species are the main culprit of cytotoxicity [67, 68]. However, the mechanism of this oligomer inherent toxicity still remains unclear. It has been proposed that they might exert the damage by sequestering functional proteins or through the formation of pores on cell membranes [69, 70].

1.3 Bacterial Inclusion bodies

1.3.1 Structure of inclusion bodies

Recombinant protein expression in *E. coli* usually results in the formation of insoluble deposits known as inclusion bodies (IBs). They are mainly composed by the heterologous protein, whose content can be up to 90 % of the total embedded protein [71]. These deposits appear as rounded shaped, hydrated, porous and very dense particles, either in the periplasmic space or in the cytoplasm. Their size, from 0.2 to 1.2 μm , depends on the recombinant

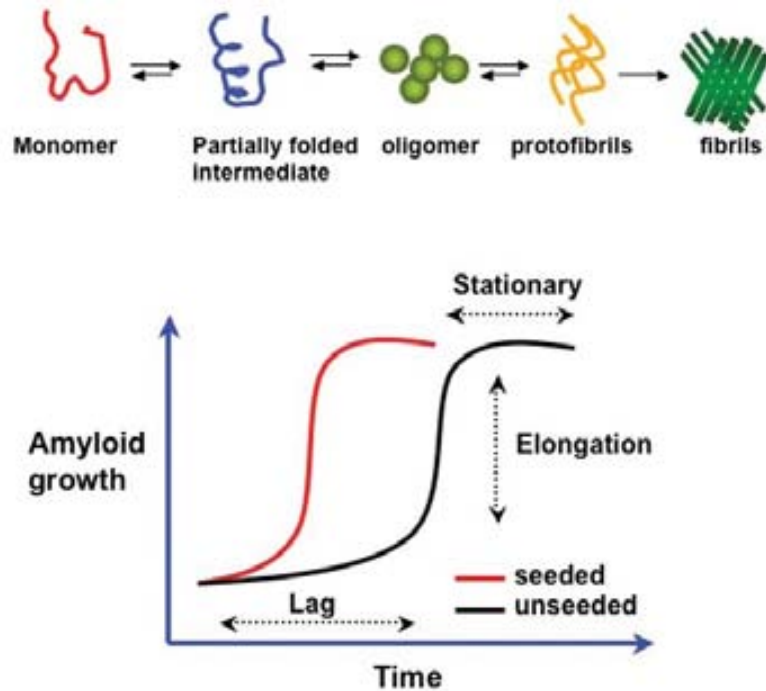


Fig. 1.8. Representation of the amyloidogenesis. Above: Protein structures populating each of the subsequent steps during amyloid fibrils formation. Below: Kinetic sigmoidal curves of the three phases of amyloidogenesis (lag, elongation and stationary phase) with and without seeding [64].

protein, the host cell and the expression conditions [72]. Traditionally, IBs were thought to be unstructured and inactive deposits of unfolded or misfolded species. The amorphous morphology visualized in microscopy images made consider IBs as *dust balls* with little interest.

Nonetheless, this classical vision has changed and nowadays, independent groups have demonstrated that these bacterial protein deposits have an amyloid-like nature [73–75], displaying structural and physico-chemical features resembling to those of amyloid fibrils. The set of methods allowing to characterize amyloid-like properties in IBs is continuously extending. Among them, the probably most used ones are the amyloid dyes thioflavin-T (ThT) and Congo Red, which undergo spectroscopical changes upon binding to amyloid β -sheets [73, 74]. Low resolution approaches, such as circular dichroism or infrared spectroscopy, are also commonly employed to corroborate the content in β secondary structure inside IBs [75, 76]. Moreover, high resolution techniques may result useful to study the structure of IBs at a molecular

level. This is the case of X-ray diffraction, that shows a characteristic pattern of cross- β structure when applied to IBs. Concretely, two reflections are displayed at 4.7 Å and at 10 Å, which are consistent with the spacing between amyloid strands and between adjacent β -sheets, respectively (Fig. 1.7). Finally, the ultimate proof that these bacterial aggregates are amyloid-like has been achieved by means of electronic microscopy. Although IBs have an amorphous external morphology, proteolytic digestion with proteinase K (highly active against most conformations except buried and densely packed β -strands) allows imaging fibrillar structures inside IBs [77].

Importantly, the demonstration that IBs are structured and ordered protein aggregates has been carried out using different proteins with completely distinct native folds, confirming that the amyloid nature of IBs is a general property of these bacterial aggregates regardless the protein expressed [78]. Therefore, despite their simplicity, the structural similarities between IBs and amyloid fibrils turn these bacterial aggregates into a simple but relevant model in the study of amyloid aggregation.

1.3.2 Formation of Inclusion bodies

Protein deposition, as explained in section 1.2, is the result of a kinetic competition between folding and aggregation which, in the case of bacteria, ends up in the formation of IBs. During recombinant production, the concentration of unfolded and partially folded conformers increases, amplifying the risk of aggregation through non-native intermolecular contacts. Native, partially folded as well as unfolded species can aggregate into IBs. The polypeptide chains embedded in these insoluble deposits are in dynamic equilibrium with soluble forms in the cytosol, which, in turn, can also be native or unfolded structures (Fig. 1.9) [9, 78].

The formation of IBs is a process that can be divided into two stages. Firstly, misfolded and partially folded species exposing HSs establish intermolecular contacts leading to aberrant self-assembly. This results in the appearance of small aggregates stochastically distributed in the cytoplasm. Afterwards, due to the nucleation nature of the protein aggregation process, these small deposition foci merge to give rise to a large aggregate, the IB, which is tethered to the cell pole [79]. The determinants of the polar localization of IBs are still unclear. Albeit there exist evidences pointing out that it is an energy dependent process [79], it has also been suggested that the localization at cell poles is circumstantial as it passively occurs in the DNA-free space [80]. In this case, the principle determinant of final IBs positioning is the nucleoid occlusion, usually taking place at the cell centre. Thereby, a bacterial cell contains one or, at the most, two IBs usually situated at the cell poles. However, under stress conditions, IBs might appear at other locations.

IBs assembly is not driven by generic hydrophobic contacts, as it was previously thought, but it is rather a highly specific process, where intermolecular contacts are established through selected regions of the polypeptides chains.

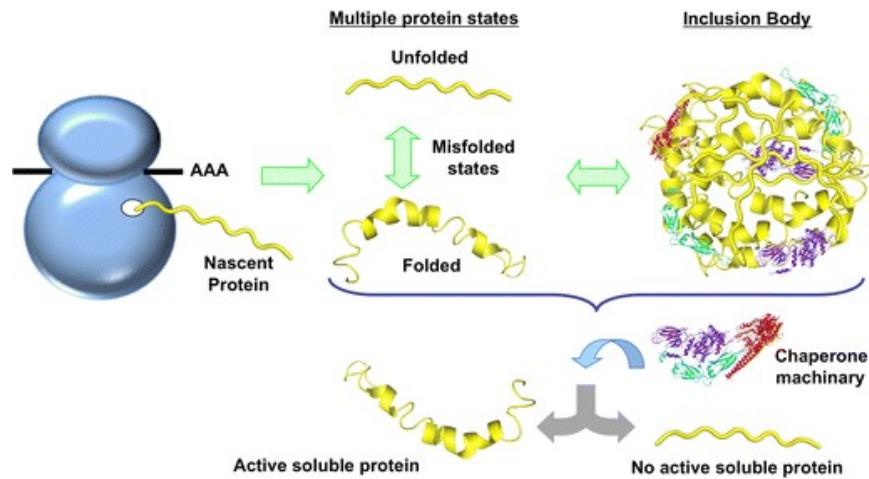


Fig. 1.9. Conformations adopted by a recombinant protein inside bacteria. The nascent polypeptide can fold or remain totally or partially unfolded. These structures can aggregate forming IBs, in which multiple conformations coexist. There is a dynamic equilibrium between soluble and insoluble fractions, promoted by the action of molecular chaperones. Thus, functional and inactive protein species are present both as soluble forms or embedded in the IBs [9].

This sequential specificity has been demonstrated by means of the simultaneous co-expression of two aggregation-prone proteins in the same cell. This experiment gave rise to the formation of two distinct aggregates in a single cell with different morphology and composition [81]. However, it can also result in the formation of one type of aggregate per cell, but with different spatial distribution of the two recombinant proteins. In this case, the most amyloidogenic protein is embedded in the inner core of the IB, while the less aggregation prone one remains attached to the aggregate surface (Fig. 1.10) [77].

1.3.3 Segregation of IBs during bacterial division

In a bacillus bacteria such as *E. coli*, cell poles are not equivalent entities, rather they can be differentiated: the old pole corresponds to the aging parental and the new pole to the progeny. One of the distinct traits is that old poles tend to accumulate IBs. This asymmetric deposition determines the asymmetric inheritance of protein deposits, which are more probable to be inherited by the old pole offspring cell. Hence, regardless of the initial position in the mother cell, through cycles of cell division, IBs eventually accumulate at the old pole (Fig. 1.11) [82].

Another differential trait between poles resides in the growth rate, as the old pole daughter cell divides slower than the new pole offspring, evidencing

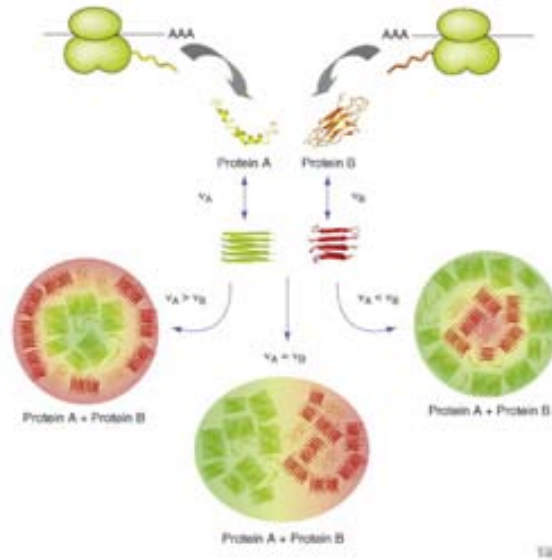


Fig. 1.10. The simultaneous coexpression of two recombinant proteins may result in the formation of heterogeneous IBs. The aggregation propensity of the proteins determine their spatial distribution. The more aggregation prone polypeptide and, thus, the one displaying higher self-assembly rate is embedded in the inner core of the IB [78].

a differential aging effect [80, 82]. The overlapping between IBs location and aging suggests that protein aggregation might represent a damaging burden associated to a loss in cell fitness. Thereby, the asymmetric distribution of protein deposits can be regarded as an evolutive mechanism to cope with the fitness lost, since it enables the appearance of a population, the new pole offspring, rejuvenated respectively to the mother cell.

1.3.4 Biotechnological exploitation of IBs

1.3.4.1 IBs contain active conformations

Beyond cell biology studies, biotechnology industry has also focused its attention on bacterial IBs. For a long time, they were considered an obstacle to overcome since aggregation is the main bottleneck in recombinant protein production. Hence, most of efforts were devoted to increase the solubility of the protein of interest. However, the demonstration that, in fact, IBs might contain native-like structure has represented a conceptual revolution in the field. A pioneering work reported that IBs formed by either enzymes or fluorescent proteins retained a certain degree of biological activity due to the dynamic equilibrium between soluble and insoluble forms (Fig. 1.9) [83]. Since then, an

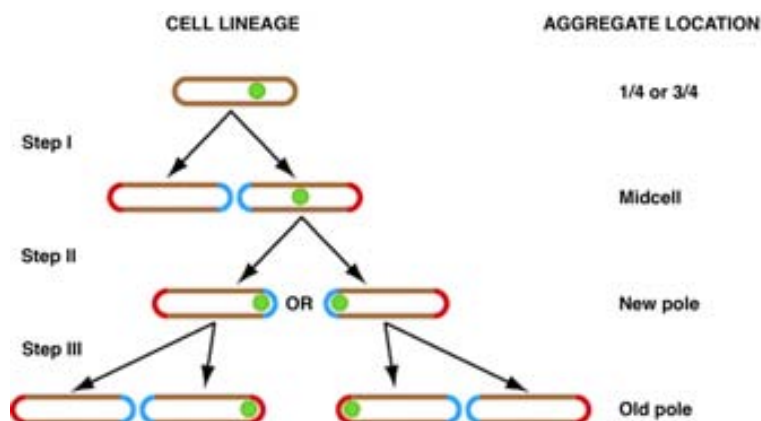


Fig. 1.11. Despite IBs can be initially formed at polar, mid-cell or quarter-cell positions, they are eventually located at one of the cell poles. Through cycles of cell divisions, they tend to accumulate at the parental pole thus being inherited by the old-pole offspring. Old and new poles are indicated in red and blue, respectively [82].

increasing number of independent groups have corroborated that part of the polypeptides embedded in these deposits is folded in native-like conformations using different and unrelated proteins [84–87].

These findings arise the possibility to employ IBs as functional nanomaterial, being especially suitable as biocatalysers and drug carriers. This last application intends to convert IBs into *nanopills*, becoming therapeutic agents capable of rescuing damaged mammalian cells [88, 89]. On this regard, IBs present some advantages since they are inert and stable but biologically tunable entities. Moreover, they are easily produced and purified from the cheap and fast-growing bacterial factories.

1.3.4.2 IBs in screening assays

Since IBs are amyloid-like and the protein embedded is partially active, they can be used as screening tools for amyloid aggregation inhibitors. Generally, these pharmaceutically important assays are cumbersome and performed with expensive synthetic peptides. Nevertheless, the broad knowledge achieved on bacterial aggregates has enabled their use in screening methods.

On one hand, fluorescent tag reporters of deposition have been employed to straightforward measure the assembly rate, as the final fluorescence of the aggregates is the result of a kinetic competition between folding and aggregation [90]. Thus, any compound that enhances or inhibits the latter reaction can be easily detected by spectrofluorometry. This new application has been exploited in a screening assay based on living bacterial cells expressing the

amyloid peptide fused to the GFP. The assay overcomes some of the problems traditionally associated to peptide-based assays, as bacterial cultures are cost-effective. Moreover, being performed *in vivo*, it has the advantage of working in biologically relevant conditions [91].

On the other hand, an interesting alternative assay has been recently described, which is independent of fluorescent tags. On the contrary, it makes use of the ThS dye, whose spectroscopic properties change upon binding to amyloid conformations. The method consists of the staining with this amyloid dye applied to intact bacterial cells containing cytoplasmic amyloid-like IBs. The coupling of this step to flow cytometry permits the rapid ThS staining analysis in large cell populations, since the spectroscopic change in ThS emission, indicative of the presence of amyloid-like structures, is detected cell-by-cell by flow cytometer. This methodology can be applied in the screening for amyloid aggregation inhibitors, representing an alternative *in vivo* assay suitable for high-throughput application, being non-invasive and chimeras-free (Fig. 1.12) [92].

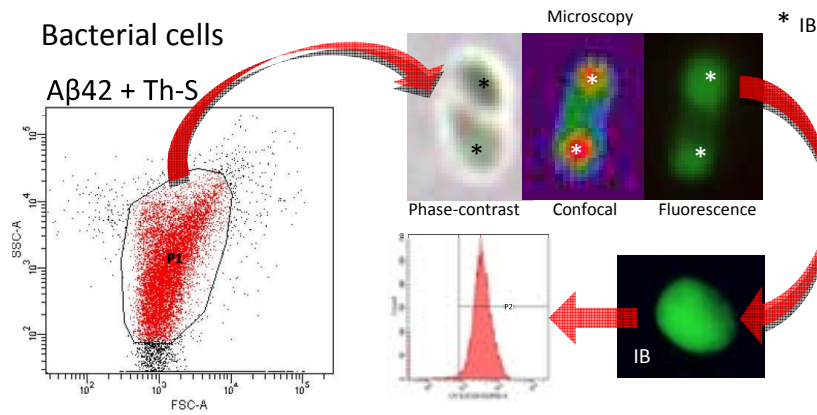


Fig. 1.12. ThS staining applied to cells containing amyloid aggregates is detected by means of flow cytometry analysis. This can be implemented as a high-throughput screening assay for amyloid modulator compounds [92].

1.4 Models used in the present thesis

1.4.1 Amyloid β -peptide

AD is the most common neurodegenerative disorder, affecting 50-70 % of the demented patients, most of them at the old age. It is a conformational disease or amyloidosis with two principal hallmarks in the patients brains: senile or amyloid plaques in the extracellular space, consisting of aggregated amyloid β -peptides, and intracellular neurofibrillar tangles formed by filaments of hyperphosphorylated tau protein (Fig. 1.13) [93].

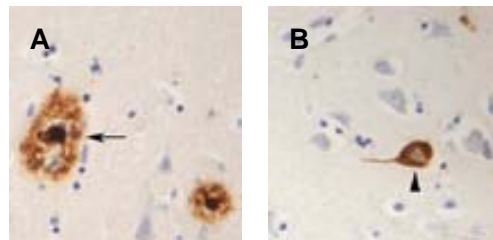


Fig. 1.13. Histological samples of neocortex from AD patients. A: $A\beta$ peptide immunohistochemistry demonstrating senile plaques (long arrow). B: hyperphosphorylated tau immunohistochemistry for neurofibrillar tangles (arrowhead) [94].

Amyloid β -peptides composing senile plaques are the result of the proteolytic processing of the amyloid precursor protein (APP), which is a cell membrane-anchored protein with unknown functionality. However, its conservation along species indicates that it may play an important biological role. Structural studies have revealed that APP is a growth factor resembling molecule and it has been demonstrated to be involved in the proliferation and differentiation of neural stem cells even at the early stages of embryogenesis. It has also been reported to play a role in cell adhesion, neuron migration and synaptogenesis [95]. The proteolytic cleavage of APP is enzymatically exerted by different secretases (Fig. 1.14). In physiological conditions, APP is mainly cleaved by α -secretase and γ - and ϵ -secretases releasing a non-pathological small peptide, P3. On the contrary, in the pathological pathway, β -secretase substitutes α -secretase in the proteolytic processing of APP, giving rise to a heterogeneous mixture of peptides (amyloid β or $A\beta$ peptides) variable in length, from 39 to 43 residues, but highly amyloidogenic [95–97]. The most abundant ones are $A\beta_{40}$ and $A\beta_{42}$ (40 and 42 residues length, respectively). Albeit the former is the most abundant in healthy brains, the latter is the

main component of senile plaques. Actually, A β 42 possesses two additional residues at the C-terminal end that strongly contribute to its higher aggregation propensity [98].

A β peptides, specially A β 42, have been widely employed in the field of protein aggregation, due to their high tendency to self-assembly and their biomedical relevance. Mutagenesis experiments have unraveled a central hydrophobic cluster composed by five residues (Leu-Val-Phe-Phe-Ala) that acts as a HS of aggregation [99]. Thereby, amino acid substitutions decreasing the hydrophobicity burden in this region slow down the peptide aggregation rate [100]. This type of experiments have been exploited to determine the intrinsic aggregation properties of the amino acids side-chains [101].

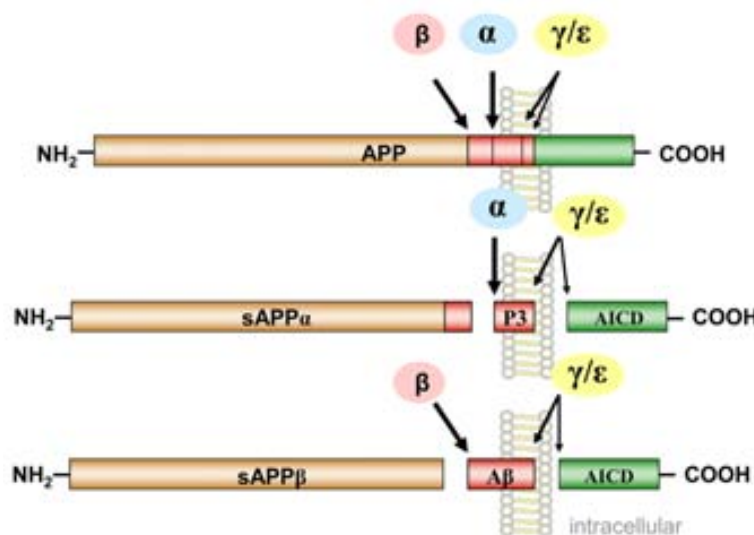


Fig. 1.14. Above: The entire APP anchored to the plasmatic membrane. Cleavage sites are indicated for the different secretases. Centre: APP is processed by α -secretases and γ - ϵ -secretases in the non-amyloidogenic pathway. Below: In the amyloidogenic pathway, β -secretase cleaves APP releasing the A β peptides [96].

1.4.2 Cellular models

1.4.2.1 *Escherichia coli*

E. coli constitutes the cellular model *par excellence* in the biochemistry field. This enterobacteria was discovered in 1885 as a diarrhea causative pathogen. Besides its medical interest, *E. coli* has been extensively exploited in molecular biology becoming the worldwide most studied microorganism. Cohen and

Boyer, using plasmids and restriction enzymes, implemented the recombinant DNA technology in this host, which resulted in the foundation of biotechnology and recombinant proteins production [102, 103].

E. coli is the most commonly used organism for heterologous protein expression. One of the main reasons resides in the profound genetic and physiological characterization achieved for a long a time, becoming a well established organism in laboratories. Principal advantages related to its culture are: short generation time, easy handling, availability of non-pathogenic strains and cheap culture media [104]. Expression in this organism allows recovering high yields of the recombinant protein, either in the cytoplasm or in the periplasmic space. However, too high expression levels usually saturate the cellular protein quality machinery leading to the aggregation of the protein of interest in IBs. This problem, representing the actual main bottleneck in the bacterial recombinant production, can be addressed by decreasing the vector copy number, changing the expression promoter, co-expressing molecular chaperones or varying the culture conditions. In contrast, sometimes the main problem may be a low production yield, which is usually caused by the difference in the codon usage of *E. coli* and the overexpressed protein. In this case, there exist strains, such as Rosetta, enriched in typical eukaryotic codons. Finally, changing the bacterial expression host might be a valuable alternative to take into consideration. On this regard, gram-negative Bacilli strains have emerged as a promising production host for therapeutic proteins, with some advantages over *E. coli*; for instance, the naturally high secretion ability and the lack of lipopolysaccharides in the outer membrane, typical of *E. coli*, that act as pyrogenic endotoxins in humans [105]. Despite their simplicity and cost-effective cultivation, prokaryotic cells have relevant limitations as production hosts. Perhaps, the most important ones are the reducing cytoplasm and the poor capacity for post-translational modifications [104].

Furthermore, the simple prokaryotic metabolism of *E. coli* may result too distinct to that of superior eukaryotes, thus it has been traditionally considered an irrelevant organism to model human disorders. However, the discovery of amyloid structure inside IBs and the similarities between these bacterial aggregates and the pathological human deposits, including nuclear inclusions and aggresomes, provide an interesting alternative to investigate the mechanism underlying protein aggregation in a biologically relevant context [106]. Among other examples, *E. coli* has been used to study the relationship between aging and protein aggregation [82, 107] or as a platform to screen for amyloid aggregation inhibitor compounds [91].

1.4.2.2 *Saccharomyces cerevisiae*

Like *E. coli*, yeasts are easy to genetically manipulate, grow cheaply and rapidly and allow to achieve high yields during heterologous proteins production. Although *Pichia Pastoris* is the most frequently used species for this purpose, the baker's yeast *Saccharomyces cerevisiae*, also known as budding

yeast, is the best genetically characterized eukaryotic organism and has also an essential role as production host as well as model for human diseases [104].

Regarding the production of recombinant proteins, yeast cells present some important advantages over bacteria. They possess complex post-translational modification pathways that are essential when producing high quality proteins for therapeutic purposes that require specific modification patterns to be fully active. Another advantage is their high secretory potential, which facilitates and reduces the cost of downstream processing. Also, yeast expression systems usually result in finely controlled levels of production which favour the solubility of the recombinant protein. *S. cerevisiae* is also the elected host in the production of metabolites like lipids, carbohydrates, amino acids, vitamins, antibiotics or organic acids. This preference over bacterial host is due to a better cellular robustness against adverse fermentation conditions and high concentration of the final compound [108].

The role of *S. cerevisiae* in the study of human disorders has been extensively addressed. Despite the simplicity of being a unicellular organism; due to its eukaryotic nature, it shares highly conserved molecular and cellular mechanisms of human cells, namely, mitochondria biogenesis, protein quality control, vesicular trafficking or autophagic pathways. Thus, budding yeast has become an invaluable model to study phenomena such as cell division, DNA replication, metabolism, protein folding and intracellular transport [109, 110].

Yeast models have been particularly implemented in the neurodegeneration field, with the objectives focused both on gaining insights in the pathological mechanisms and on screening approaches for the identification of novel therapeutic opportunities. In the case of some disorders, such as Friedreich's Ataxia, *S. cerevisiae* encodes homologues of disease-related proteins which permits the study of proteins functions and pathogenicity in their natural environment. However, in most of the cases, the human gene encoding the disease causing protein has no homologue in yeast and heterologous expression must be carried out resulting in the so-called *humanized yeast*. This is the case of common neurodegenerative disorders, such as AD, PD and HD, whose yeast models have been successfully developed (Fig. 1.15) [110, 111]. The yeast model of PD (expressing α -syn) recapitulates several pathogenic features including the formation of intracellular aggregates, trafficking defects and cell death in a concentration dependent manner. In the case of HD, the expression of mutant human htt exon 1 reproduces many of the pathological hallmarks, such as polyQ length dependent aggregation and toxicity [109]. Finally, the AD yeast model has enabled the study of APP and the role of post-translational processing secretases. Also, yeast has become a powerful cell system to study the cellular effects of the amyloidogenic peptides with highly relevant results indicating that oligomers are the most damaging species and that targeting these peptides to the secretory pathway, thus mimicking the trafficking in human cells, increases the toxicity [110, 112].

Since none of the currently available animal models for these neurodegenerative disorders fully recapitulates the spectrum of alterations observed in

humans, yeast models might provide important knowledge on the molecular mechanisms underlying pathogenesis and contribute to the development of better animal models [109].

1.4.2.3 Plants

Nowadays, practically all recombinant proteins on the market are produced in microorganisms (43 %) and animal cell cultures (57 %). However, both systems have important limitations such as incorrect post-translation modifications or expensive and cumbersome infrastructure. Within this scenario, during the last decade, plants have arisen as an interesting alternative for heterologous protein production [113].

Plants possess exceptional biosynthetic capacity and use cheap and simple media and infrastructure, which permits to achieve high amounts of biomass with low production costs. Moreover, being eukaryotic organisms, they overcome some of the problems derived from production in bacteria, such as the inability to perform post-translation modifications, critical when producing recombinant proteins with therapeutic interest. Another relevant advantage resides in their high flexibility, allowing adapting production scale, costs or safety depending on the destination of the recombinant production. Plants are versatile hosts and the production platform to employ is an important issue to be assessed. Among the most studied ones, there are the whole plants (ground or aquatic species), hairy roots or *in vitro* cultured plant cells or tissues [114].

The versatility of plants also includes the possibility to target the recombinant protein into chloroplasts, generating transplastomic plants. This novel but relevant strategy, first developed in tobacco, has been applied in other species, gaining relevance these last years and yielding commercialized products [113].

The molecular biology techniques to genetically modify plants have been deeply studied achieving great advances. However, the downstream processing of the production process is still an issue to be properly addressed, as purification methods developed for other production systems might be not totally suitable when applied to plants. Taking into consideration that downstream process contributes in the 65 - 95 % of the total cost, purification of recombinant proteins could become a bottleneck in the exploitation of plants as protein production systems [115].

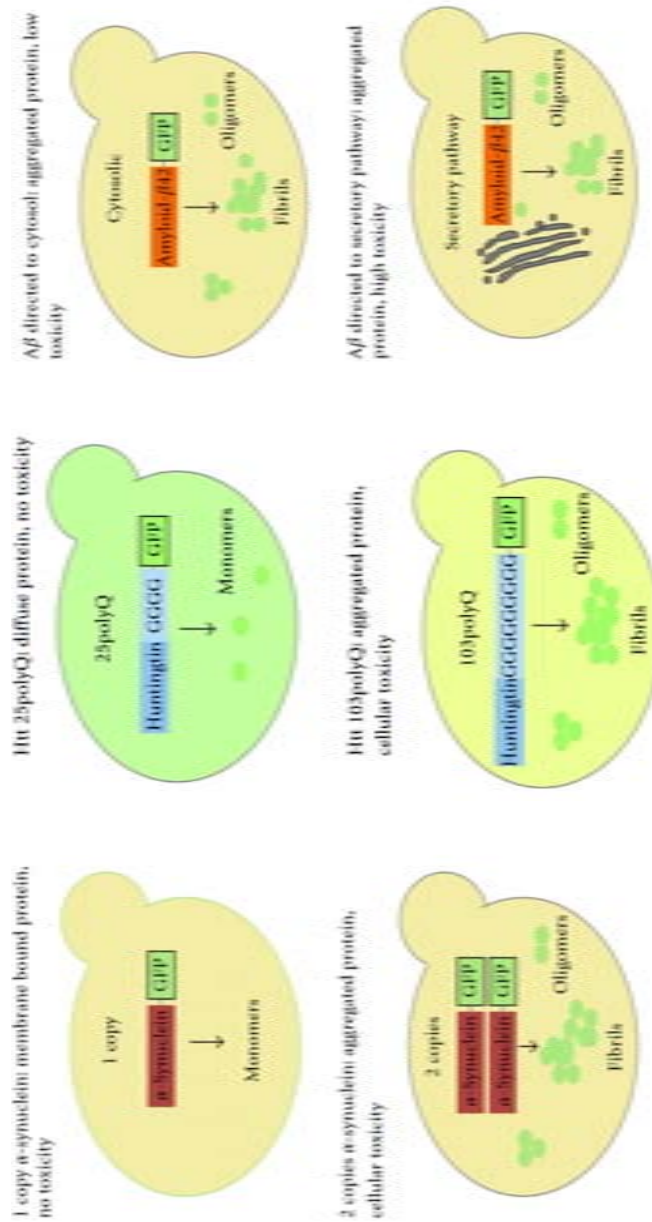


Fig. 1.15. Humanized yeast models of PD (left), HD (centre) and AD (right), based on the heterologous expression of α -syn, htt PolyQ tract and A β 42 peptide, respectively [110].

Aims

The general objective of this thesis is the study of intracellular protein deposition with amyloid features. We pretend to characterize the properties of the aggregates as well as their impact for the cellular machinery. Moreover, we aim to elucidate the role of the aggregation propensity intrinsic to the primary sequence, as one of the majors determinants of protein self-assembly. Since amyloid aggregation seems to be a generic characteristic of polypeptides, we pursue to use three distinct organisms in order to asses how this property is exerted in different cellular environments.

The principal aims of the thesis can summarized in the following items:

1. In bacteria:
 - To determine the influence of the aggregation propensity of the primary sequence in the assembly rate and in the properties of the formed protein deposits.
 - To analyze cell aging provoked by intracellular aggregates with amyloid features.
 - To study the alleviating effect promoted by molecular chaperones on the observed aging phenotype.
 - To develop a screening assay for amyloid aggregation inhibitors based on bacterial IBs.
2. In yeast:
 - To characterize protein deposition in yeast and the role of the intrinsic aggregation propensity in order to elucidate the differential influence exerted by eukaryotic and prokaryotic backgrounds.
 - To decipher the cellular mechanisms of protein deposits clearance and how they relate to the aggregation tendency of the assembled protein.
3. In plants:
 - To characterize the formation of intracellular protein deposits and their amyloid and cytotoxic features.

Publications

3.1 Publication I

The effect of amyloidogenic peptides on bacterial aging correlates with their intrinsic aggregation propensity.

Anna Villar-Piqué, Natalia S. de Groot, Raimon Sabate, Sergio P. Acebrón, Garbiñe Celaya, Xavier Fernández-Busquets, Aturo Muga and Salvador Ventura

J. Mol. Biol. 421(2-3): 270-81.
10 August 2012



The Effect of Amyloidogenic Peptides on Bacterial Aging Correlates with Their Intrinsic Aggregation Propensity

Anna Villar-Pique¹, Natalia S. de Groot¹, Raimon Sabaté¹, Sergio P. Acebrón², Garbiñe Celaya², Xavier Fernàndez-Busquets^{3,4}, Arturo Muga² and Salvador Ventura^{1*}

¹Institut de Biotecnologia i Biomedicina and Departament de Bioquímica i Biologia Molecular, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain

²Unidad de Biofísica (Consejo Superior de Investigaciones Científicas/Universidad del País Vasco-Euskal Herriko Unibertsitatea) and Departamento de Bioquímica y Biología Molecular, Universidad del País Vasco, Bilbao 48080, Spain

³Nanobioengineering Group, Institute for Bioengineering of Catalonia, Baldiri Reixac 10-12, Barcelona E08028, Spain

⁴Barcelona Center for International Health Research, CRESIB, Hospital Clínic, Universitat de Barcelona, Rosselló 132, Barcelona E08036, Spain

Received 23 September 2011;
received in revised form
6 December 2011;
accepted 7 December 2011
Available online
19 December 2011

Edited by S. Radford

Keywords:

protein aggregation;
amyloid fibrils;
inclusion bodies;
Escherichia coli;
chaperones

The formation of aggregates by misfolded proteins is thought to be inherently toxic, affecting cell fitness. This observation has led to the suggestion that selection against protein aggregation might be a major constraint on protein evolution. The precise fitness cost associated with protein aggregation has been traditionally difficult to evaluate. Moreover, it is not known if the detrimental effect of aggregates on cell physiology is generic or depends on the specific structural features of the protein deposit. In bacteria, the accumulation of intracellular protein aggregates reduces cell reproductive ability, promoting cellular aging. Here, we exploit the cell division defects promoted by the intracellular aggregation of Alzheimer's-disease-related amyloid β peptide in bacteria to demonstrate that the fitness cost associated with protein misfolding and aggregation is connected to the protein sequence, which controls both the *in vivo* aggregation rates and the conformational properties of the aggregates. We also show that the deleterious impact of protein aggregation on bacterial division can be buffered by molecular chaperones, likely broadening the sequential space on which natural selection can act. Overall, the results in the present work have potential implications for the evolution of proteins and provide a robust system to experimentally model and quantify the impact of protein aggregation on cell fitness.

© 2011 Elsevier Ltd. All rights reserved.

*Corresponding author. Institut de Biotecnologia i Biomedicina, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain. E-mail address: salvador.ventura@uab.es.

Present address: N. S. de Groot, Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, UK.

Abbreviations used: IB, inclusion body; A β , amyloid β ; GFP, green fluorescent protein; WT, wild type; Gdn-HCl, guanidinium hydrochloride; DAPI, 4',6-diamidino-2-phenylindole; PBS, phosphate-buffered saline.

Introduction

Protein misfolding and aggregation have a significant impact on cell fitness.¹ In humans, the formation of protein aggregates is associated with the onset of an increasing number of degenerative disorders, including Alzheimer's disease or diabetes.^{2,3} In bacteria, protein aggregation is linked to the loss of cell reproductive ability or aging.⁴ Importantly, the aggregates responsible for these detrimental phenotypes, namely amyloid fibrils in humans and inclusion bodies (IBs) in bacteria, share many conformational and functional properties, including neurotoxicity.^{5–7} The aggregated state represents, in fact, a free-energy ground state for proteins (alternative to that populated by the native state) and, accordingly, protein aggregation reactions are now recognized as major contributors to the shaping of protein free-energy landscapes. *In vitro* and *in silico* studies have shown that the aggregation propensities of polypeptides, as well as the conformational and toxic properties of the resulting aggregates, are strongly dependent on their sequences.^{8–13} Therefore, not only the native conformation but also the ability to form aggregated assemblies are somehow encoded in the sequence. Actually, an inverse correlation between protein abundance and sequential aggregation propensity has been observed to occur in both human and bacteria proteomes,^{14–17} suggesting that mutations that increase polypeptide aggregation propensities are more stringently purged by natural selection in highly expressed genes, whose products are at higher risk for aggregation, than in lesser expressed genes. Similarly, proteins with high turnover rate and thus short lifetime have, on average, a higher aggregation propensity than long-living proteins.¹⁸ Overall, it appears that protein sequences might be shaped by natural selection in order to provide the encoded polypeptides with an intrinsic solubility that allows these molecules to perform their function at the cellular concentrations, times, and locations required to maintain protein homeostasis and cellular fitness.¹⁹

Despite evidence indicating that intracellular protein aggregation is inherently deleterious, how the sequential and associated structural features of aggregates impact cell fitness is hardly quantifiable and still poorly understood. Deciphering this issue constitutes a challenging task because, in contrast to *in vitro* conditions where sequential contributions to aggregation can be easily dissected, in the cell, the intrinsic aggregation properties of the proteins are modulated by the action of the highly conserved protein quality machinery.²⁰ In this context, bacterial systems might provide simplified models to address this question under near-physiological conditions.²¹ Here, using a battery of amyloid β (A β) peptide mutants displaying a broad range of

intrinsic aggregation propensities as a model polypeptide and using *Escherichia coli* as a host, we demonstrate experimentally that protein aggregation arrests cell division in this organism, that this effect correlates linearly with the conformational stability of accumulated intracellular aggregates, and that the molecular folding machinery acts by alleviating the detrimental impact of protein aggregation. Overall, we provide a privileged platform to study and model how protein aggregation affects cell fitness and shapes protein sequences.

Results and Discussion

The stability of A β -green fluorescent protein bacterial intracellular aggregates correlates with their aggregation propensity

We have shown previously that the bacterial IBs formed by the Alzheimer's-disease-related A β peptide display an amyloid-like structure when this peptide is expressed alone⁷ or as a fusion to different fluorescent proteins.²² We have constructed a library of 20 different A β -green fluorescent protein (GFP) mutants differing only in the residue in position 19 of the A β peptide.²³ This residue is located in the central hydrophobic cluster of the peptide and has been shown to affect the folding, self-assembly, and fibril structure of A β .^{24,25} All these fusions are expressed at comparable levels in *E. coli*²³ and accumulate in the insoluble cellular fraction as IBs.²⁶ However, the presence of active GFP in such aggregates differs significantly (Fig. 1a and b) and displays a high correlation with the aggregation propensity of each A β mutant.²⁶ This observation suggested the existence of an *in vivo* kinetic competition between the folding of the GFP domain and the aggregation of the fusion protein, directed by the A β moiety. The faster the fusion protein aggregates, the lower is the IB fluorescence emission and vice versa, in such a way that the fluorescence of IBs would report on the *in vivo* protein aggregation kinetics.^{26,27} Here, we used time-lapse confocal microscopy to confirm this hypothesis. We observed that after the induction of protein expression at time points where the formation of IBs was already detectable at the poles of cells expressing the wild-type (WT) A β -GFP fusion, the mutant F19D, displaying the lowest predicted aggregation propensity of the mutant set, was still completely soluble (Fig. 1c).

The different *in vivo* aggregation kinetics of WT and F19D A β variants impact not only the functionality of the resulting IBs but also the structure of these aggregates.²⁸ The competition between folding and aggregation likely reflects a balance between the formation of native intramolecular

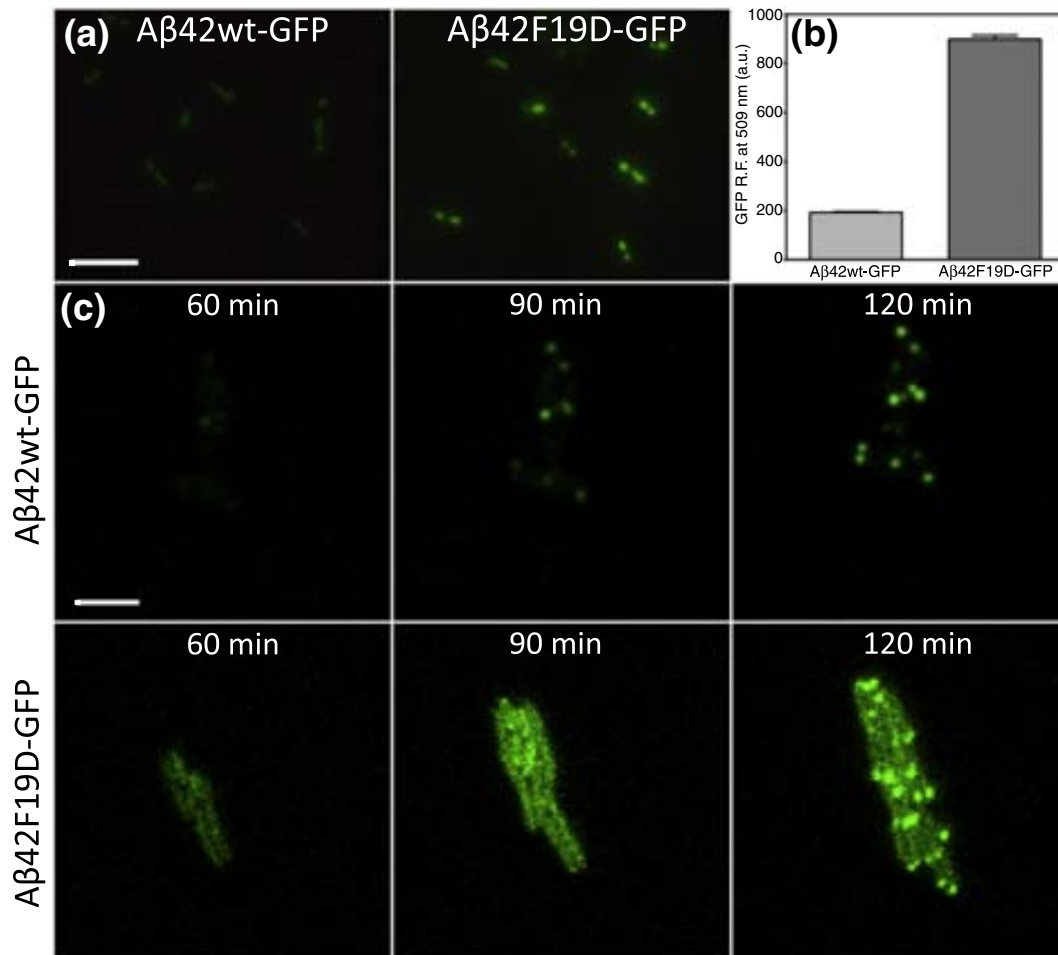


Fig. 1. Fluorescence emission of IBs formed by A β 42-GFP variants. (a) Visualization of GFP fluorescence in the cytoplasmic IBs of cells expressing WT (left) and F19D (right) A β 42-GFP by optical fluorescence microscopy 240 min after induction of protein expression. Scale bar represents 10 μ m. (b) GFP relative fluorescence of WT and F19D mutant by fluorescence spectroscopy. (c) Fluorescence images of WT (upper series) and F19D (lower series) A β 42-GFP taken during a time-lapse confocal microscopy experiment 60 min, 90 min, and 120 min after induction of protein expression. Scale bar represents 5 μ m.

interactions, which lead to functional GFP fusions, and anomalous intermolecular contacts, which lead to the formation of the amyloid-like β -sheet structure—a conformation recurrently observed in the bacterial IBs formed by sequentially and structurally unrelated polypeptides.^{5,6} Like in amyloid fibrils, the network of intermolecular contacts between different polypeptide chains glues the structure of IBs.²² Here, we analyzed if the predicted aggregation propensity for the complete mutant set correlates with the conformational stability of their intracellular aggregates. To this aim, the different IBs were dissolved in 2 M guanidinium hydrochloride (Gdn·HCl), and the amount of remaining aggregated material was measured by monitoring the decrease in IB solution turbidity at 360 nm after equilibrium had been reached. A striking correlation ($r=0.94$, $p<10^{-5}$) between the predicted aggrega-

tion propensity—and, by inference, the *in vivo* aggregation rate—of the mutant and the stability of the resulting A β -GFP intracellular aggregates was observed (Fig. 2a). Atomic force microscopy can be used to measure single-molecule or single-particle rupture forces. We have shown previously that the fibrillar material inside the aggregates is mainly responsible for the mechanical resistance of WT A β -GFP IBs to atomic force microscopy tip indentation.²² We selected four IBs of mutants displaying different resistances against chemical disintegration and measured their relative stability in mechanical disruption (Fig. 2b). Again, a clear direct correlation between the predicted aggregation propensity of the A β peptide and the resistance of the corresponding IBs to mechanical disruption was observed ($r=0.99$, $p<0.0008$). As expected, the mechanical and chemical stabilities of the IBs are

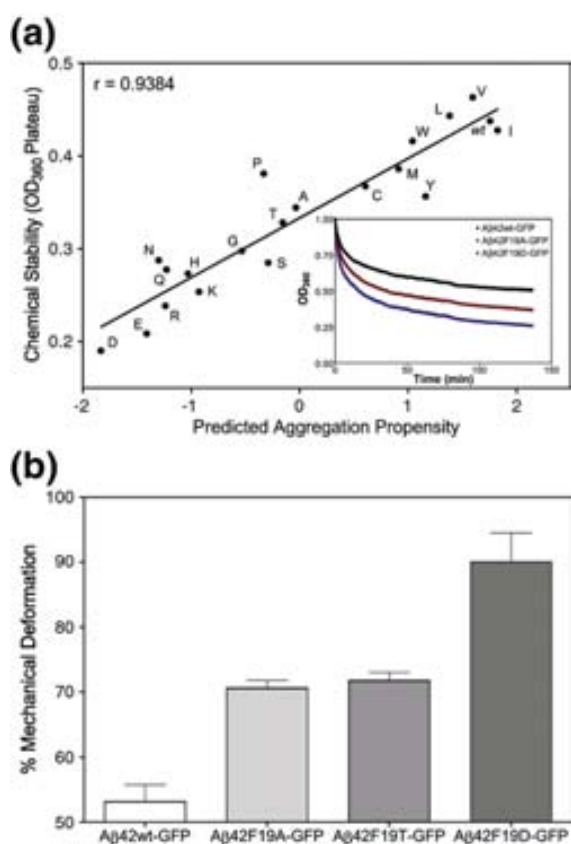


Fig. 2. Chemical and mechanical stabilities of IBs formed by A β 42-GFP variants. (a) Correlation between the aggregation propensity (calculated using AGGRESCAN⁸) and the chemical denaturation resistance of purified IBs. The remaining turbidity at the plateau phase after the incubation of IBs in the presence of 2 M Gdn-HCl has been employed as an indicator of the stability of the aggregates during chemical denaturation. Inset: The resultant denaturation curves fitted to a double-exponential decay curve. The legend indicates the residue at the 19th position of each peptide variant. (b) Mechanical deformation of several A β 42-GFP variants by force spectroscopy.

also correlated ($r=0.99$, $p<0.0006$). Our results indicate that the intracellular aggregates formed by different polypeptides should not be treated as identical entities, since the conformational properties of these ensembles depend on the intrinsic properties of the embedded proteins, and even point mutations can alter their physical properties significantly.

The disaggregating capability of molecular chaperones depends on the conformational stability of the aggregate

Efficient reactivation of protein aggregates requires the collaboration of members of the chaperone families Hsp40, Hsp70, and Hsp100.^{29–31} In

prokaryotes, these are chaperone DnaK (Hsp70; DnaK system), cochaperones DnaJ (Hsp40) and GrpE (KJE system), and the Hsp100 chaperone ClpB. Using thermal protein aggregates formed *in vitro*, we have shown previously that chaperone association with protein aggregates proceeds sequentially.^{32,33} In the first step, aggregate-bound DnaJ drives DnaK to the aggregate surface. Subsequently, DnaK mediates ClpB binding to the aggregate, where the combined effort of the two chaperone (KJE and ClpB) systems extracts unfolded substrate molecules for subsequent folding. However, the effect of this bichaperone network (KJE and ClpB) on the reactivation of physiologically relevant aggregates, such as IBs, has not been assayed yet. We tested the reactivation activity on WT A β -GFP IBs by monitoring the recovery of functional GFP, resulting in a fluorescence increase in the sample. As shown in Fig. 3a, in the absence of ClpB, the DnaK system displays poor reactivating activity, and the refolding yield was not significantly enhanced with a 2-fold increase in the concentration of the DnaK system components (DnaK, DnaJ, and GrpE) (data not shown). When the same experiments were carried out in the presence of ClpB, the reactivation efficiency of the bichaperone system was significantly enhanced. Like amyloid fibrils, IBs are highly enriched in intermolecular β -sheet structure,^{6,22,34} extracting unfolded monomers from these stable aggregates would be energetically more demanding than extracting unfolded monomers from unstructured aggregates, since more and stronger nonnative interactions must be disrupted. As a consequence, the DnaK system is not competent, by itself, to reactivate the IBs and requires ClpB to do it efficiently. This observation is in line with the ClpB requirement to reactivate *in-vitro*-formed aggregates of G6PDH³³ or luciferase³⁵ enriched in β -sheet secondary structure. Accordingly, we could demonstrate that the chaperones DnaK, DnaJ, and ClpB all bind to the surface of the IBs during the refolding reaction (Fig. S1).

If the limiting step for IB reactivation by molecular chaperones is the extraction of individual polypeptide chains from the aggregates, as our results suggest, it is expected that IBs differing in conformational stability would impose dissimilar challenges to the cellular quality-control machinery. To test this possibility, we compared the bichaperone network activity on WT A β -GFP IBs to that on the F19D variant (Fig. 3b). As for the WT IBs, ClpB is required for F19D IB reactivation. However, kinetic analysis reveals that, for any particular ClpB concentration and at any time point, the reactivation yield is around 2-fold higher for the mutant IBs, indicating that the chaperone systems work more efficiently on the less stable F19D aggregates. To confirm that the observed differences in the refolding efficiencies of the IBs are due to the different

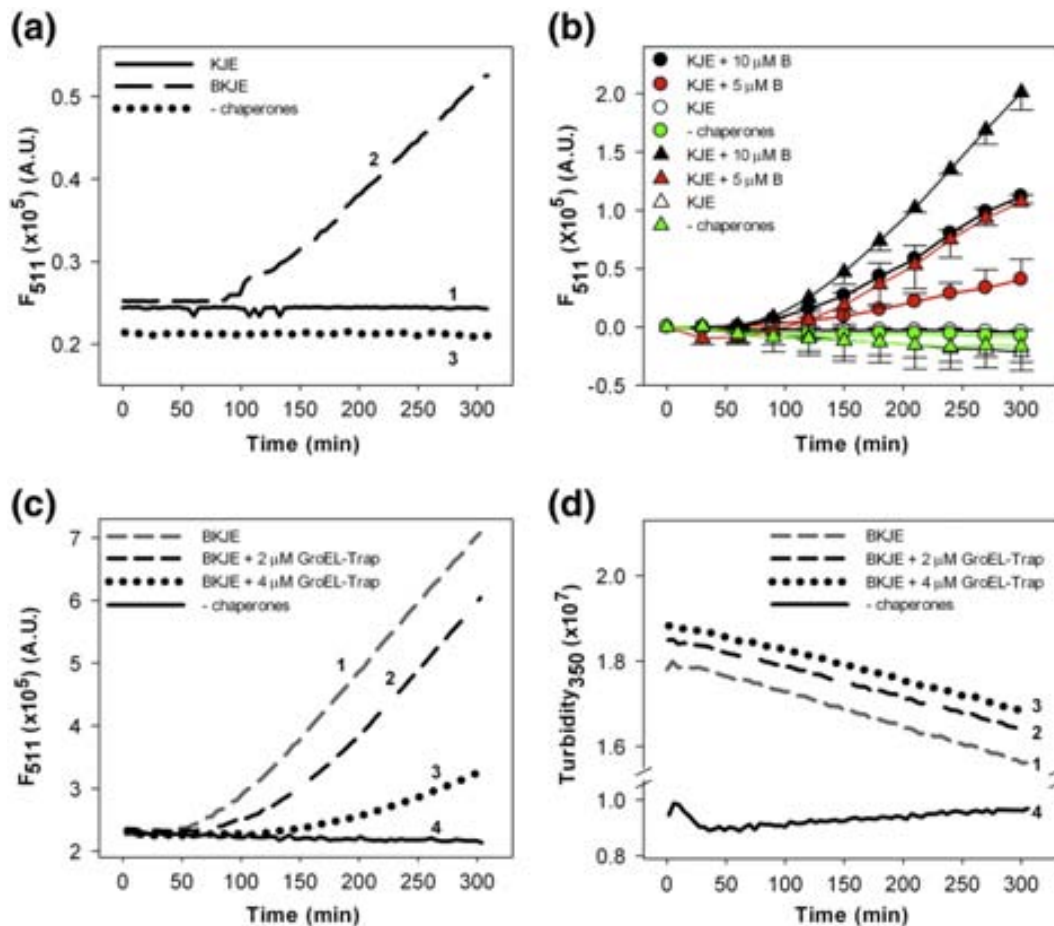


Fig. 3. Chaperone-mediated reactivation of IBs. (a) Refolding of A β 42-GFP aggregates by the DnaK system alone (DnaK–DnaJ–GrpE (KJE); trace 1) and by the DnaK/ClpB bichaperone system (BKJE; trace 2). The fluorescence of the aggregates in the absence of chaperones was also run as a control (trace 3). (b) Time dependence of the refolding of A β 42-GFP (circles) and A β 42-F19D-GFP (triangles) aggregates by the DnaK system alone (white) and in the presence of 5 μ M (red) or 10 μ M (black) ClpB monomer. As reference, the fluorescence of the aggregates in the absence of chaperones is shown (green). (c) Effect of GroEL-trap on the chaperone-assisted refolding of A β 42-GFP aggregates. The refolding mixtures contained the DnaK/ClpB bichaperone system alone (trace 1) and in the presence of 2 μ M (trace 2) or 4 μ M (trace 3) GroEL-trap monomer. Trace 4 corresponds to the fluorescence monitored in the absence of chaperones. (d) Turbidity of the same samples shown in (c) along the refolding reaction.

work the chaperones have to perform to detach individual polypeptidic chains from the aggregate, we used a trap mutant of the GroEL chaperone (GroEL-trap). This mutant binds soluble unfolded polypeptides without releasing them.³⁶ When GroEL-trap was present in the reaction mixture, a significant reduction in fluorescence recovery was detected (Fig. 3c), whereas the reduction in turbidity along the reaction was independent of GroEL-trap (Fig. 3d). This indicates that, *in vitro*, the refolding of the GFP moiety occurs spontaneously in solution, after the extraction of the protein fusion from the aggregate by the DnaK/ClpB bichaperone system. GroEL-trap prevents its refolding by trapping newly released unfolded molecules. This stable interaction might help to keep the substrate in solution, since the change in

turbidity is similar regardless of the presence of GroEL-trap. Data also indicate that, in the absence of GroEL-trap, disaggregation and refolding act as coupled reactions. Overall, the results indicate that the ability of molecular chaperones to undo the accumulation of misfolded polypeptides in intracellular aggregates is strongly dependent on their structural properties.

Impairment of cell division by A β -GFP bacterial intracellular aggregates correlates with their aggregation propensity

The intracellular aggregation of proteins seems to exert a generic fitness cost,¹ which, in the case of prokaryotic cells, has been shown to be associated with cellular aging, since the presence of protein

aggregates appears to be significantly correlated with the loss of reproductive ability.^{4,21} However, it is not known whether this deleterious effect on cellular fitness is of the same magnitude for any protein aggregate or depends on the specific nature of the deposit. When we observed cells expressing the highly aggregation-prone WT A β -GFP fusion by phase-contrast and fluorescence microscopies 4 h after the induction of protein expression, we often detected the presence of cells containing three or more IBs, with many of them displaying a filamentous morphology (Fig. 4a). The formation of cytoplasmic IBs usually begins specifically at one

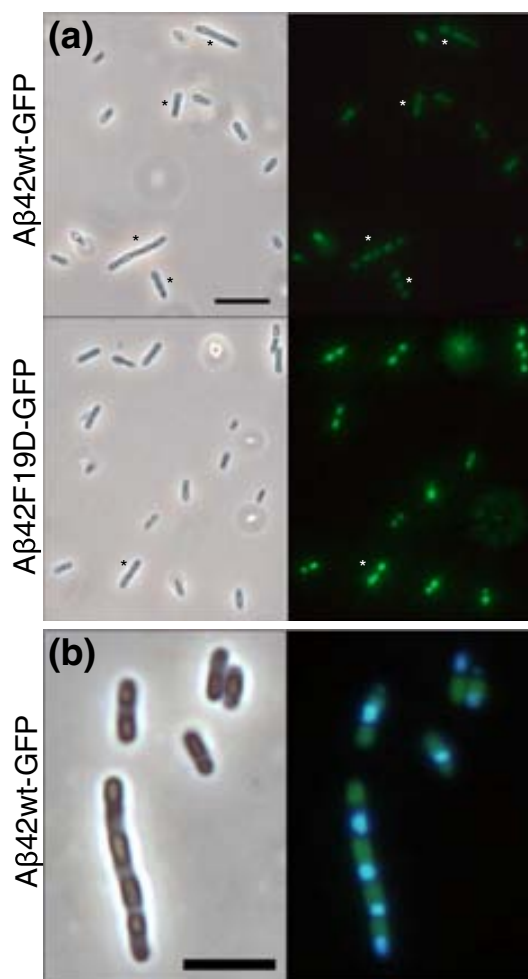


Fig. 4. (a) Visualization of GFP fluorescence in the cytoplasmic IBs of cells expressing WT (upper) and F19D (lower) A β 42-GFP. Left images correspond to phase-contrast microscopy, and right images correspond to fluorescence microscopy under UV light. Scale bar represents 10 μ m. (b) Localization of cytoplasmic IBs and nucleoids in cells expressing A β 42-GFP. Images were taken using phase-contrast microscopy (left) and fluorescence microscopy under UV light (right; merged image of DAPI and GFP fluorescence). Scale bar represents 5 μ m.

of the bacterial poles, but IBs can appear at both poles during cell development.^{37,38} Therefore, after division, each of the daughter cells can inherit, at maximum, one of the parental IBs. Accordingly, the presence of three or more IBs in a single cell indicates a division defect and the inability to reproduce by binary fission (aging).⁴ 4',6-Diamidino-2-phenylindole (DAPI) staining indicates that these cells duplicate their genetic material but are unable to segregate it into new cells (Fig. 4b), in line with evidence indicating that chromosome replication does not trigger cell division in *E. coli* but that the chromosome replication and cell division cycles of *E. coli* run in parallel, independently of each other.³⁹ The mechanism for aggregate localization in *E. coli* remains controversial, with some data indicating that an active energy-driven process is necessary for polar localization,³⁷ while others support nucleoid occlusion as the primary parameter influencing both the number and the position of heat-induced aggregates in *E. coli*⁴⁰ and, thus, a passive mechanism for aggregate localization. In our system, IBs localize specifically in the nucleoid-free space (Fig. 4b), in agreement with the nucleoid segregation-driven model. Interestingly, qualitatively, this phenotype was more rare in cells expressing the less amyloidogenic F19D variant (Fig. 4a), suggesting that the effect on cell division can be related to the conformational properties of the aggregates. To confirm this hypothesis, we monitored the morphology of 1000 IB-containing cells for each of the 20 A β -GFP variants. As shown in Fig. 5, we found a significant correlation between the proportion of cells displaying abnormalities in division and the predicted aggregation propensity of the peptide ($r=0.87$, $p<10^{-5}$) (Fig. 5a) or the stability of the aggregate ($r=0.78$, $p<10^{-5}$) (Fig. 5b), thus indicating that the fitness cost of protein aggregation is strongly dependent on the nature of the aggregating polypeptide.

Chaperones alleviate cell division defects

ClpB and DnaK have been shown to be specifically associated with IBs inside bacterial cells,³⁸ suggesting that they may play a detoxifying role by disaggregating or remodeling intracellular protein deposits, thus reducing their impact on cell physiology. We coexpressed the complete chaperone network, including DnaK, DnaJ, GrpE, ClpB, GroEL, and GroES, and monitored its influence on IB conformation and cell division arrest. A strong increase in the fluorescence of IBs and a drastic reduction in the population of cells displaying division defects were observed (Fig. 6). Still, the F19D IBs formed in this background were more active and less toxic than the WT under the same conditions, indicating that the detected beneficial effect of chaperones on cell division was sequence

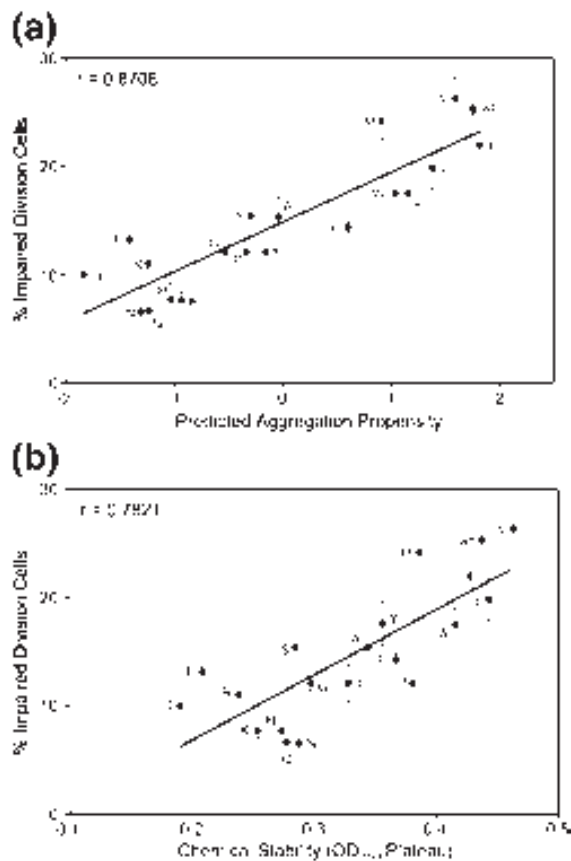


Fig. 5. Correlation between the percentage of *E. coli* cells containing more than two IBs and (a) the aggregation propensity of A β 42-GFP variants (calculated using AGGRESCAN⁸) or (b) their chemical stability. The legend indicates the residue at the 19th position of each peptide variant.

dependent. Again, IBs containing more active GFP were less toxic, suggesting that, even for the same polypeptide sequence, the toxic effect depends on the specific conformational properties of its intracellular aggregates.

Although the overexpression of the complete chaperone set has a clear impact on cell division, a surprising observation was that most of the active GFP moieties still remained embedded in the IBs (Fig. 6) without significant mobilization of fluorescent A β -GFP fusions to the cytoplasm. This observation suggests several possibilities. (i) Reactivation of misfolded molecules would occur at the aggregate surface. However, our *in vitro* data seem to contradict this view, since extraction of individual polypeptides appears to be a requirement for refolding, as demonstrated in the GroEL-trap experiments. (ii) Refolding of the globular GFP domain effectively occurs at the cytosol after extraction, but the high aggregation propensity of the A β domain would promote reincorporation of

the protein fusion (now displaying a functional GFP domain) into the IBs. As a consequence, these intracellular aggregates would become enriched in functional protein, thus increasing the proportion of intramolecular interactions in the IBs at the expense of intermolecular associations and decreasing their conformational stability. (iii) Because DnaK and GroELs assist folding and try to prevent the aggregation of newly synthesized polypeptides,⁴¹ the observed effect on IB activity and cell division would not correspond to the processing of already formed aggregates but to a chaperone-mediated modulation of the inherent polypeptide aggregation propensity. To decipher between these two last possibilities, we coexpressed the WT and F19D A β -GFP variants together with both the DnaK system and the GroELs system, but in the absence of ClpB overexpression (Fig. 6). The levels of IB fluorescence and cell division recovery were identical with those obtained in the presence of the complete chaperone complement, suggesting that the disaggregating activity of the Hsp100 machinery does not play a crucial role in alleviating the cell fitness cost associated with protein aggregation in our model system. Accordingly, no significant changes in IB fluorescence were observed for any of the variants when ClpB was overexpressed alone. Also, the proportion of cells suffering from division problems was similar or higher in ClpB-overexpressing cells than in controls (Fig. 6). According to our data, the most likely explanation for the observed adaptive effect of molecular chaperones is that they interact with the polypeptide fusion, rendering its aggregation slower and/or assisting the folding of the GFP moiety during the aggregation process. Any of these effects would result in the formation of more active IBs (glued by a reduced number of intermolecular contacts) and therefore less stable and—according to our results—less toxic aggregates. Our data contribute additional evidence to the hypothesis stating that chaperones are able to buffer the destabilizing effect of genetic mutations in proteins, therefore playing a critical role in the maintenance of genetic diversity and accelerating the rate of adaptation.^{42,43} Mutations that provide proteins with new or improved reaction properties are generally destabilizing and tend to occur in more buried residues than nonadaptive mutations.⁴⁴ Destabilization of the buried core has been shown to be the triggering factor of amyloid formation for a large number of proteins in *in vitro* studies.^{45–47} Therefore, it is tempting to propose that the effect of chaperones on genetic diversity is exerted not only by increasing the soluble fraction of stability-impaired mutant proteins but also by reducing the fitness cost associated with their aggregation, favoring the maintenance of these variants in the cell population until additional adaptive mutations could allow their structural stabilization and genetic fixation.

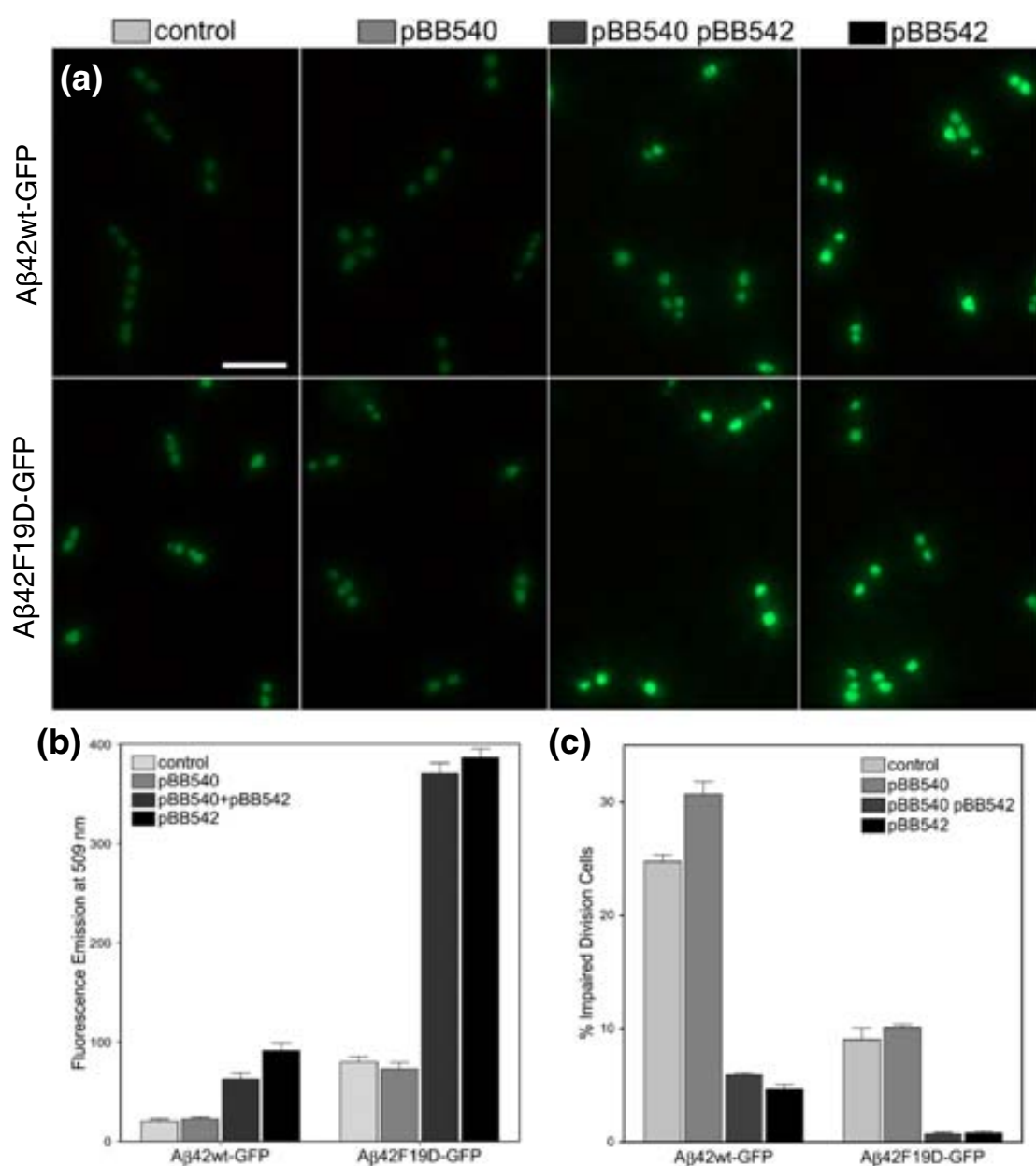


Fig. 6. Chaperones' effect on the formation of Aβ42-GFP IBs. (a) Visualization of GFP fluorescence in the cytoplasmic IBs of cells expressing WT (upper series) and F19D (lower series) Aβ42-GFP under UV light. From left to right, images correspond to control cells expressing Aβ42-GFP, cells coexpressing plasmids pBB540 (containing *grpE* and *clpB* genes) and pBB540, and cells coexpressing pBB542 (containing *dnaK*, *dnaJ*, and *groESL* genes) and pBB542, respectively. Scale bar represents 5 μm. Exposure time: 222 ms for WT; 69 ms for F19D. (b) Relative fluorescence determination. (c) Percentage of cells with impeded cell division in the presence or in the absence of chaperones.

Conclusions

Although it is widely accepted that protein misfolding and aggregation reduce cell fitness and that increasing evidences converge to suggest that selection against protein aggregation might act as an important constraint in the evolution of protein sequences, the precise fitness cost associated with

protein aggregation has been traditionally difficult to quantify experimentally. In a recent study, Geiler-Samerotte *et al.* measured the growth defects resulting from the expression of four different misfolded YFP mutants in yeast and demonstrated that cells respond to steady-state protein misfolding and aggregation by increasing the levels of cytosolic chaperones.⁴⁸ Despite the relevance of this work, the

relationship between the aggregation propensity of the different variants, the conformational properties of the resulting aggregates, and the final fitness cost were not addressed. Here, we exploit the cell division defects exerted by the intracellular aggregation of the A β peptide and its fusion to GFP in bacteria to demonstrate that the impact of protein aggregates on cell physiology is tightly regulated by the protein sequence, which encodes *in vivo* aggregation rates and, consequently, the structural and functional properties of protein deposits. The data confirm that chaperones can modulate intrinsic sequential traits, reducing the cellular impact of protein aggregation even in cases where they cannot avoid it and likely allowing the evolution of protein sequences that otherwise would be purged from the genome due to their associated toxicity, thus relaxing the global constraint imposed by protein aggregation.

A question that arises recurrently when addressing the impact of protein misfolding and aggregation on cell fitness, as formulated by Geiler-Samerotte *et al.*,⁴⁸ is whether it is caused by the stimulus or the response. In other words, do the deleterious effects result from the intrinsic toxicity of the aggregates or from the cost of the unfolded protein response they induce? Our data suggest that, at least in our model system, the particular properties of the protein aggregates are mainly responsible for the observed phenotypes because even when we saturate the cells with molecular chaperones, the impact on cell division correlates with the structural properties of the aggregate. Overall, we think that our findings have potential implications on the natural evolution of proteins, providing an experimental system to model the impact of protein aggregation on fitness costs and protein sequences in different genetic backgrounds or under diverse physiological conditions by means of large-scale competition experiments (work in progress). Moreover, rationalizing the effect of mutations on the *in vivo* conformational properties of aggregates and their relationship with physiological defects might provide a means to develop computational strategies for the prediction of protein aggregation effects on living organisms.

Materials and Methods

Cloning and protein expression

E. coli BL21(DE3) competent cells were transformed with pET-28a vectors (Novagen, Inc., Madison, WI) encoding for 20 different variants of A β 42–GFP fusions, differing only in residue 19 of the A β 42 domain, as previously described.²³ pBB540 (*clpB* and *grpE*) and pBB542 (*dnaK*, *dnaJ*), and *groESL* plasmids were used for

the expression of molecular chaperones.⁴⁹ Cultures were grown at 37 °C and 250 rpm in LB medium containing appropriate antibiotics. At an OD₆₀₀ of 0.5, protein expression was induced with 1 mM isopropyl β -D-1-thiogalactopyranoside (IPTG) for 4 h.

Purification of IBs

Cells were harvested by centrifugation, and pellets were resuspended in native buffer [150 mM NaCl and 10 mM Tris (pH 7.5)]. IBs were purified from cell extracts using a detergent-based procedure, as described previously.⁵⁰ Briefly, cell pellets from 10 mL of culture were resuspended in lysis buffer [100 mM NaCl, 50 mM Tris (pH 8), and 1 mM ethylenediaminetetraacetic acid]. Cells were incubated at 37 °C for 30 min after the addition of the protease inhibitor PMSF and lysozyme at the final concentrations of 15 mM and 300 μ g/mL, respectively. NP-40 was added at 1%. After incubation at 4 °C for 30 min, the cells were treated with DNase and RNase, both at 15 μ g/mL. Protein aggregates were harvested by centrifugation at 12,000g and washed with lysis buffer containing 0.5% Triton X-100. Finally, IBs were washed with phosphate-buffered saline (PBS; pH 7) to remove the remaining detergent.

Fluorescence measurements

The emission spectra of GFP in cultures expressing A β 42–GFP fusion variants were measured on a Perkin-Elmer 650-40 spectrofluorimeter (Perkin-Elmer, Wellesley, MA). The GFP fluorescence of purified IBs at an OD₆₀₀ of 0.1 was recorded in the range 500–600 nm, using an excitation wavelength of 470 nm. For microscopy analysis, 10 μ L of culture was collected after 4 h of induction, washed in native buffer, and deposited on top of glass slides. Images were obtained under UV light or using phase-contrast microscopy in a Leica fluorescence DMBR microscope (Leica Microsystems, Mannheim, Germany). Bacterial nucleoids were stained by incubation at 37 °C for 1 h with DAPI at a final concentration of 2 μ g/mL. The number of IBs per cell was counted in 20 different fields for each culture, and the data collected represent the average of about 1000 cell counts for each sample.

Time-lapse confocal microscopy of living bacteria

Cultures were grown at 37 °C and 250 rpm until an OD₆₀₀ of 0.5 had been reached. Immediately after addition of IPTG, 10 μ L of each culture was placed on a sterilized glass slide with a layer of solidified medium (LB, 2% agarose, 50 μ g/mL kanamycin, and 1 mM IPTG) and covered with a sterilized coverslip. The glass slides with cells expressing A β 42–GFP fusions were placed at 37 °C in the incubation chamber of a TCS-SP5 AOBS confocal laser scanning microscope (Leica Microsystems Heidelberg GmbH, Mannheim, Germany). GFP fluorescence was excited using a 458-nm argon laser (497–555 nm emission bandwidth collected). Images were digitally captured at specified time intervals for 4 h and analyzed with LAS AF Lite Software (Leica Microsystems CMS GmbH, Mannheim, Germany).

Stability of IBs during chemical solubilization

One way to determine the stability of an IB is by performing an equilibrium study of its solubilization by a chaotropic agent, assuming a two-state mechanism in which the protein is either in an aggregated state that contributes to turbidity or in a soluble state with no contribution to the signal.^{7,27,28,51} Here, 750 μL of a 1.35 OD_{360} solution of purified and homogenized IBs in PBS (pH 7) was added to 250 μL of 8 M Gdn-HCl solution ($\text{OD}_{360}=1.0$ and $[\text{Gdn-HCl}]=2$ M at the beginning of the reaction) for kinetic denaturation experiments. The reaction was monitored for 180 min by measuring the changes in OD_{360} in a Cary-100 Varian spectrophotometer (Varian, Inc., Oxford, UK), with stirring by microstir bars at ~ 300 rpm. The remaining OD_{360} at the end of the reaction was taken as a value that reflects the differential stability of the assayed IBs against chemical denaturation.²⁸ No significant changes in OD_{360} were observed in the absence of denaturant. The reactions could be fitted to double-exponential decay curves using GraphPad Prism 3.02 (GraphPad Software, La Jolla, CA).

Mechanical stability by force spectroscopy

Force spectroscopy analysis of IBs was performed with a multimode atomic force microscope (Veeco Instruments, Inc., Santa Barbara, CA) equipped with a 12- μm scanner (E-scanner) using gold-coated V-shaped cantilevers with unsharpened Si_3N_4 tips. The images were taken in liquid media using a liquid cell without the O-ring seal. Veeco NP-S probes were used to scan the samples in tapping mode at a scan rate of 0.5 Hz or 1 Hz. For sample preparation, about 50 μL of a 1 OD_{350} solution of purified and homogenized WT, F19A, F19T, and F19D A β 42-GFP IBs in PBS (pH 7) was deposited on cleaved, highly oriented pyrolytic graphite (NT-MDT Co., Moscow, Russia) and allowed to adsorb for about 20 min before the measurements were started. As standard IB indentation procedure, a cantilever was used to apply increased force until total deformation of each IB had been achieved, thus obtaining approach and retraction curves showing tip force (nN) as a function of sample deformation (nm). The sample deformation at a fixed applied tip force of 2.5 nN (force at which all IBs show some degree of deformation) was recorded, and the percentage of deformation of each IB was calculated in order to compare various IBs. Each data set was collected from 30 IBs within the size range 15–25 nm for each sample.

In vitro assays with chaperones

DnaK, DnaJ, GrpE, ClpB, and the GroEL Δ 87K variant were obtained as previously reported.^{52–54} Purified IBs (0.2 μM) were resuspended in buffer [50 mM Tris-HCl (pH 7.5), 20 mM MgCl_2 , 150 mM KCl, and 10 mM DTT] containing DnaK (3.5 μM), DnaJ (0.7 μM), and GrpE (0.35 μM) in the absence or in the presence of ClpB (5 μM or 10 μM monomer). Reactivation was started by adding 4 mM ATP to the reaction mixture that also contained an ATP regenerating system (15 mM phosphoenolpyruvate and 20 ng/mL pyruvate kinase). GFP refolding was measured at 30 $^\circ\text{C}$ following the fluorescence intensity at

511 nm, using an excitation wavelength of 470 nm, in a Fluorolog-3 spectrofluorimeter (HORIBA; Jobin Yvon). Excitation and emission slits were set to 3 nm. Sample turbidity was measured under the same experimental conditions used in the refolding assays, with both excitation and emission wavelengths set to 500 nm.

Acknowledgements

We thank Ario de Marco for kindly providing us the chaperone-encoding plasmids, and Arthur L. Horwich for the GroEL-trap plasmid. This work was supported by grants BFU2010-14901 (to S.V.) and BFU2010-15443 (to A.M.) from Ministerio de Ciencia e Innovación (Spain) and by grant 2009-SGR 760 from AGAUR (Generalitat de Catalunya). S.V. was granted an ICREA ACADEMIA award (ICREA).

Supplementary Data

Supplementary data to this article can be found online at [doi:10.1016/j.jmb.2011.12.014](https://doi.org/10.1016/j.jmb.2011.12.014)

References

1. Olzscha, H., Schermann, S. M., Woerner, A. C., Pinkert, S., Hecht, M. H., Tartaglia, G. G. *et al.* (2011). Amyloid-like aggregates sequester numerous metastable proteins with essential cellular functions. *Cell*, **144**, 67–78.
2. Chiti, F. & Dobson, C. M. (2006). Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.* **75**, 333–366.
3. Fernandez-Busquets, X., de Groot, N. S., Fernandez, D. & Ventura, S. (2008). Recent structural and computational insights into conformational diseases. *Curr. Med. Chem.* **15**, 1336–1349.
4. Lindner, A. B., Madden, R., Demarez, A., Stewart, E. J. & Taddei, F. (2008). Asymmetric segregation of protein aggregates is associated with cellular aging and rejuvenation. *Proc. Natl Acad. Sci. USA*, **105**, 3076–3081.
5. de Groot, N. S., Sabate, R. & Ventura, S. (2009). Amyloids in bacterial inclusion bodies. *Trends Biochem. Sci.* **34**, 408–416.
6. Wang, L., Maji, S. K., Sawaya, M. R., Eisenberg, D. & Riek, R. (2008). Bacterial inclusion bodies contain amyloid-like structure. *PLoS Biol.* **6**, e195.
7. Dasari, M., Espargaro, A., Sabate, R., Lopez Del Amo, J. M., Fink, U., Grelle, G. *et al.* (2011). Bacterial inclusion bodies of Alzheimer's disease beta-amyloid peptides can be employed to study native-like aggregation intermediate states. *ChemBioChem*, **12**, 407–423.
8. Conchillo-Sole, O., de Groot, N. S., Aviles, F. X., Vendrell, J., Daura, X. & Ventura, S. (2007). AGGRES-CAN: a server for the prediction and evaluation of "hot spots" of aggregation in polypeptides. *BMC Bioinf.* **8**, 65.

9. Sabate, R., Espargaro, A., de Groot, N. S., Valle-Delgado, J. J., Fernandez-Busquets, X. & Ventura, S. (2010). The role of protein sequence and amino acid composition in amyloid formation: scrambling and backward reading of IAPP amyloid fibrils. *J. Mol. Biol.* **404**, 337–352.
10. Trovato, A., Seno, F. & Tosatto, S. C. (2007). The PASTA server for protein aggregation prediction. *Protein Eng. Des. Sel.* **20**, 521–523.
11. Maurer-Stroh, S., Debulpaep, M., Kuemmerer, N., Lopez de la Paz, M., Martins, I. C., Reumers, J. *et al.* (2010). Exploring the sequence determinants of amyloid structure using position-specific scoring matrices. *Nat. Methods*, **7**, 237–242.
12. Tartaglia, G. G. & Vendruscolo, M. (2008). The Zyggregator method for predicting protein aggregation propensities. *Chem. Soc. Rev.* **37**, 1395–1401.
13. Monsellier, E., Ramazzotti, M., de Laureto, P. P., Tartaglia, G. G., Taddei, N., Fontana, A. *et al.* (2007). The distribution of residues in a polypeptide sequence is a determinant of aggregation optimized by evolution. *Biophys. J.* **93**, 4382–4391.
14. Tartaglia, G. G., Pechmann, S., Dobson, C. M. & Vendruscolo, M. (2007). Life on the edge: a link between gene expression levels and aggregation rates of human proteins. *Trends Biochem. Sci.* **32**, 204–206.
15. Tartaglia, G. G., Pechmann, S., Dobson, C. M. & Vendruscolo, M. (2009). A relationship between mRNA expression levels and protein solubility in *E. coli*. *J. Mol. Biol.* **388**, 381–389.
16. Castillo, V., Grana-Montes, R. & Ventura, S. (2011). The aggregation properties of *Escherichia coli* proteins associated with their cellular abundance. *Biotechnol. J.* **6**, 752–760.
17. de Groot, N. S. & Ventura, S. (2010). Protein aggregation profile of the bacterial cytosol. *PLoS One*, **e9383**, 5.
18. De Baets, G., Reumers, J., Delgado Blanco, J., Dopazo, J., Schymkowitz, J. & Rousseau, F. (2011). An evolutionary trade-off between protein turnover rate and protein aggregation favors a higher aggregation propensity in fast degrading proteins. *PLoS Comput. Biol.* **7**, e1002090.
19. Tartaglia, G. G. & Vendruscolo, M. (2009). Correlation between mRNA expression levels and protein aggregation propensities in subcellular localisations. *Mol. Biosyst.* **5**, 1873–1876.
20. Buchberger, A., Bukau, B. & Sommer, T. (2010). Protein quality control in the cytosol and the endoplasmic reticulum: brothers in arms. *Mol. Cell*, **40**, 238–252.
21. Mogk, A., Huber, D. & Bukau, B. (2011). Integrating protein homeostasis strategies in prokaryotes. *Cold Spring Harbor Perspect. Biol.* **3**.
22. Morell, M., Bravo, R., Espargaro, A., Sisquella, X., Aviles, F. X., Fernandez-Busquets, X. & Ventura, S. (2008). Inclusion bodies: specificity in their aggregation process and amyloid-like structure. *Biochim. Biophys. Acta*, **1783**, 1815–1825.
23. de Groot, N. S., Aviles, F. X., Vendrell, J. & Ventura, S. (2006). Mutagenesis of the central hydrophobic cluster in Abeta42 Alzheimer's peptide. Side-chain properties correlate with aggregation propensities. *FEBS J.* **273**, 658–668.
24. Morimoto, A., Irie, K., Murakami, K., Masuda, Y., Ohigashi, H., Nagao, M. *et al.* (2004). Analysis of the secondary structure of beta-amyloid (Abeta42) fibrils by systematic proline replacement. *J. Biol. Chem.* **279**, 52781–52788.
25. Bitan, G., Vollers, S. S. & Teplow, D. B. (2003). Elucidation of primary structure elements controlling early amyloid beta-protein oligomerization. *J. Biol. Chem.* **278**, 34882–34889.
26. de Groot, N. S. & Ventura, S. (2006). Protein activity in bacterial inclusion bodies correlates with predicted aggregation rates. *J. Biotechnol.* **125**, 110–113.
27. de Groot, N. S. & Ventura, S. (2006). Effect of temperature on protein quality in bacterial inclusion bodies. *FEBS Lett.* **580**, 6471–6476.
28. Espargaro, A., Sabate, R. & Ventura, S. (2008). Kinetic and thermodynamic stability of bacterial intracellular aggregates. *FEBS Lett.* **582**, 3669–3673.
29. Glover, J. R. & Lindquist, S. (1998). Hsp104, Hsp70, and Hsp40: a novel chaperone system that rescues previously aggregated proteins. *Cell*, **94**, 73–82.
30. Mogk, A., Tomoyasu, T., Goloubinoff, P., Rudiger, S., Roder, D., Langen, H. & Bukau, B. (1999). Identification of thermolabile *Escherichia coli* proteins: prevention and reversion of aggregation by DnaK and ClpB. *EMBO J.* **18**, 6934–6949.
31. Sabate, R., de Groot, N. S. & Ventura, S. (2010). Protein folding and aggregation in bacteria. *Cell. Mol. Life Sci.* **67**, 2695–2715.
32. Acebron, S. P., Fernandez-Saiz, V., Taneva, S. G., Moro, F. & Muga, A. (2008). DnaJ recruits DnaK to protein aggregates. *J. Biol. Chem.* **283**, 1381–1390.
33. Acebron, S. P., Martin, I., del Castillo, U., Moro, F. & Muga, A. (2009). DnaK-mediated association of ClpB to protein aggregates. A bichaperone network at the aggregate surface. *FEBS Lett.* **583**, 2991–2996.
34. Carrio, M., Gonzalez-Montalban, N., Vera, A., Villaverde, A. & Ventura, S. (2005). Amyloid-like properties of bacterial inclusion bodies. *J. Mol. Biol.* **347**, 1025–1037.
35. Lewandowska, A., Matuszewska, M. & Liberek, K. (2007). Conformational properties of aggregated polypeptides determine ClpB-dependence in the disaggregation process. *J. Mol. Biol.* **371**, 800–811.
36. Fenton, W. A., Kashi, Y., Furtak, K. & Horwich, A. L. (1994). Residues in chaperonin GroEL required for polypeptide binding and release. *Nature*, **371**, 614–619.
37. Rokney, A., Shagan, M., Kessel, M., Smith, Y., Rosenshine, I. & Oppenheim, A. B. (2009). *E. coli* transports aggregated proteins to the poles by a specific and energy-dependent process. *J. Mol. Biol.* **392**, 589–601.
38. Ventura, S. & Villaverde, A. (2006). Protein quality in bacterial inclusion bodies. *Trends Biotechnol.* **24**, 179–185.
39. Bernander, R. & Nordstrom, K. (1990). Chromosome replication does not trigger cell division in *E. coli*. *Cell*, **60**, 365–374.
40. Winkler, J., Seybert, A., Konig, L., Pruggnaller, S., Haselmann, U., Sourjik, V. *et al.* (2010). Quantitative and spatio-temporal features of protein aggregation in *Escherichia coli* and consequences on protein quality control and cellular ageing. *EMBO J.* **29**, 910–923.
41. Kerner, M. J., Naylor, D. J., Ishihama, Y., Maier, T., Chang, H. C., Stines, A. P. *et al.* (2005). Proteome-wide

- analysis of chaperonin-dependent protein folding in *Escherichia coli*. *Cell*, **122**, 209–220.
42. DePristo, M. A., Weinreich, D. M. & Hartl, D. L. (2005). Missense meanderings in sequence space: a biophysical view of protein evolution. *Nat. Rev. Genet.* **6**, 678–687.
 43. Tokuriki, N. & Tawfik, D. S. (2009). Chaperonin overexpression promotes genetic variation and enzyme evolution. *Nature*, **459**, 668–673.
 44. Tokuriki, N., Stricher, F., Serrano, L. & Tawfik, D. S. (2008). How protein stability and new functions trade off. *PLoS Comput. Biol.* **4**, e1000002.
 45. Pallares, I., Vendrell, J., Aviles, F. X. & Ventura, S. (2004). Amyloid fibril formation by a partially structured intermediate state of alpha-chymotrypsin. *J. Mol. Biol.* **342**, 321–331.
 46. Eichner, T., Kalverda, A. P., Thompson, G. S., Homans, S. W. & Radford, S. E. (2011). Conformational conversion during amyloid formation at atomic resolution. *Mol. Cell*, **41**, 161–172.
 47. Dhulesia, A., Cremades, N., Kumita, J. R., Hsu, S. T., Mossuto, M. F., Dumoulin, M. *et al.* (2010). Local cooperativity in an amyloidogenic state of human lysozyme observed at atomic resolution. *J. Am. Chem. Soc.* **132**, 15580–15588.
 48. Geiler-Samerotte, K. A., Dion, M. F., Budnik, B. A., Wang, S. M., Hartl, D. L. & Drummond, D. A. (2011). Misfolded proteins impose a dosage-dependent fitness cost and trigger a cytosolic unfolded protein response in yeast. *Proc. Natl Acad. Sci. USA*, **108**, 680–685.
 49. de Marco, A., Deuerling, E., Mogk, A., Tomoyasu, T. & Bukau, B. (2007). Chaperone-based procedure to increase yields of soluble recombinant proteins produced in *E. coli*. *BMC Biotechnol.* **7**, 32.
 50. Carrio, M. M., Cubarsi, R. & Villaverde, A. (2000). Fine architecture of bacterial inclusion bodies. *FEBS Lett.* **471**, 7–11.
 51. Sabate, R., Espargaro, A., Saupe, S. J. & Ventura, S. (2009). Characterization of the amyloid bacterial inclusion bodies of the HET-s fungal prion. *Microbial Cell Fact.* **8**, 56.
 52. Mehl, A. F., Heskett, L. D. & Neal, K. M. (2001). A GrpE mutant containing the NH(2)-terminal “tail” region is able to displace bound polypeptide substrate from DnaK. *Biochem. Biophys. Res. Commun.* **282**, 562–569.
 53. Zylicz, M., Yamamoto, T., McKittrick, N., Sell, S. & Georgopoulos, C. (1985). Purification and properties of the dnaJ replication protein of *Escherichia coli*. *J. Biol. Chem.* **260**, 7591–7598.
 54. Woo, K. M., Kim, K. I., Goldberg, A. L., Ha, D. B. & Chung, C. H. (1992). The heat-shock protein ClpB in *Escherichia coli* is a protein-activated ATPase. *J. Biol. Chem.* **267**, 20429–20434.

3.2 Publication II

Using bacterial inclusion bodies to screen for amyloid aggregation inhibitors.

Anna Villar-Piqué*, Alba Esparagaró*, Raimon Sabaté, Natalia S. de Groot and Salvador Ventura

Microb. Cell Fact. 11(1): 55-65.
3 May 2012

* Equal contribution

TECHNICAL NOTES

Open Access

Using bacterial inclusion bodies to screen for amyloid aggregation inhibitors

Anna Villar-Piqué^{1,2†}, Alba Espargaró^{2†}, Raimon Sabaté^{2,3}, Natalia SdeGroot^{2,4} and Salvador Ventura^{1,2*}

Abstract

Background: The amyloid- β peptide (A β 42) is the main component of the inter-neuronal amyloid plaques characteristic of Alzheimer's disease (AD). The mechanism by which A β 42 and other amyloid peptides assemble into insoluble neurotoxic deposits is still not completely understood and multiple factors have been reported to trigger their formation. In particular, the presence of endogenous metal ions has been linked to the pathogenesis of AD and other neurodegenerative disorders.

Results: Here we describe a rapid and high-throughput screening method to identify molecules able to modulate amyloid aggregation. The approach exploits the inclusion bodies (IBs) formed by A β 42 when expressed in bacteria. We have shown previously that these aggregates retain amyloid structural and functional properties. In the present work, we demonstrate that their *in vitro* refolding is selectively sensitive to the presence of aggregation-promoting metal ions, allowing the detection of inhibitors of metal-promoted amyloid aggregation with potential therapeutic interest.

Conclusions: Because IBs can be produced at high levels and easily purified, the method overcomes one of the main limitations in screens to detect amyloid modulators: the use of expensive and usually highly insoluble synthetic peptides.

Keywords: Amyloids, Inclusion bodies, Protein folding, Protein aggregation, Metals, Alzheimer

Background

In the last few years, protein aggregation has emerged from a neglected area of protein chemistry as a transcendental issue in biological and medical sciences, mainly because the deposition of proteins into insoluble amyloid fibrils is being found behind an increasing number of human diseases such as Alzheimer's disease (AD) or type II diabetes [1–4]. Therefore, there is an increasing interest in developing methods to identify molecules that trigger the aggregation of these proteins inside the organism as well as to discover chemical compounds that can interfere with these pathways, having thus therapeutic potential [5–7].

The pathological hallmark of AD is brain deposition of amyloid plaques composed predominantly by the A β 42 peptide isoform [8–10]. The origin of these insoluble

extracellular neurotoxic deposits is still not completely clear, and multiple factors such as pH, peptide concentration, oxidative stress and metal ions have been reported to trigger their formation [11,12]. Here we present a fast, cost-effective high-throughput approach to study conditions and molecules that affect A β 42 aggregation. The assay is based on the use of the inclusion bodies (IBs) formed by an A β 42-GFP fusion protein in bacteria. IBs formation has long been regarded as an unspecific process relaying on the establishment of hydrophobic contacts [13,14]. However, there are now strong evidences demonstrating that bacterial IBs formation shares a number of common features with the formation of the highly ordered and pathogenic amyloid fibrils linked to human diseases [15–18]. Therefore, IBs have become an attractive model to study protein aggregation and their consequences in simple but biologically relevant environments [19–21]. IBs are formed inside the cell when the folding of proteins into native conformations is competed by a faster establishment of anomalous intermolecular interactions that leads to the formation of insoluble aggregates [22]. In the present work,

* Correspondence: salvador.ventura@uab.es

[†]Equal contributors

¹Departament de Bioquímica i Biologia Molecular, Facultat de Ciències, Universitat Autònoma de Barcelona, E-08193, Bellaterra, Spain

²Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, E-08193, Bellaterra, Spain

Full list of author information is available at the end of the article

we exploit this kinetic competition *in vitro* to screen for compounds that can modulate protein aggregation. As a proof of principle, we demonstrate the ability of the approach to detect the effect of metal ions on A β 42 aggregation as well as to identify compounds that block this metal-induced reaction.

Results and Discussion

Refolding A β 42-GFP IBs is sequence specific

We have previously shown that the IBs formed by A β 42 display amyloid-like properties whether the peptide is expressed alone [23] or fused to fluorescent proteins

[16,24]. We have constructed a set of 20 different A β 42-GFP variants, which differ only in a single residue in the peptide's central hydrophobic region [25]. All these proteins are expressed at similar levels in *E. coli* and form insoluble IBs [25]. Nevertheless, the fraction of active GFP in those aggregates is significantly different (Figure 1). The IBs fluorescence correlates with the aggregation propensity of the specific A β 42 mutant [26]. This correlation is the result of a kinetic competition between the folding of the GFP domain and the aggregation of the fusion protein, which is driven by the A β 42 moiety. Therefore, the slower the fusion protein aggregates, the higher the IB fluorescence

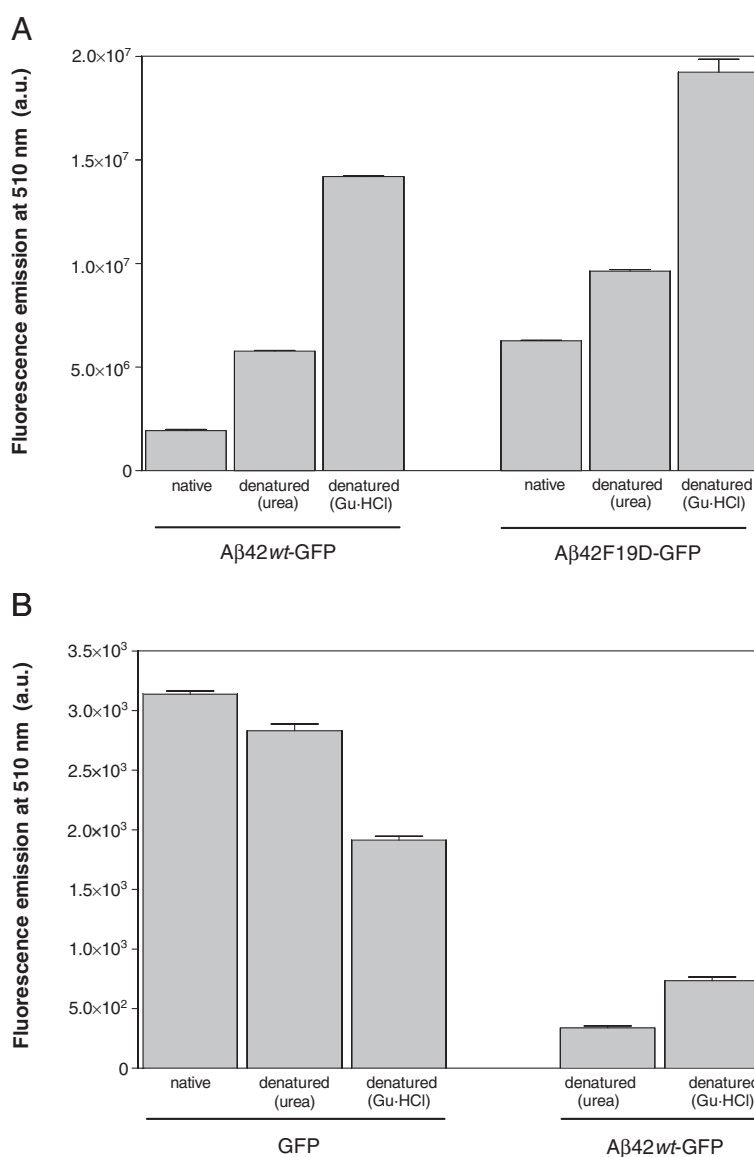


Figure 1 Fluorescence recovery after denaturation. (A) Purified IBs were incubated in PBS in the absence (native) and presence of 8 M Gu-HCl or 10 M urea for 4 h and diluted 100-fold in PBS. (B) Purified untagged GFP (left) and IBs (right) were incubated in PBS in the absence (native) and presence of 8 M Gu-HCl or 10 M urea for 4 h and diluted 100-fold in PBS. In all cases, after incubation for 16 h fluorescence was recorded at 510 nm.

emission is and *vice versa*. In this way, IBs fluorescence reports on intracellular aggregation kinetics [22,26,27].

We wondered if the kinetic competition between GFP folding and A β 42 aggregation can be reproduced *in vitro*. To this aim we used the IBs formed by the *wild-type* peptide fusion (A β 42*wt*-GFP) and the F19D mutant (A β 42F19D-GFP), which display the highest and lowest aggregation propensities in our library, respectively [22]. Purified IBs were denatured to remove the polypeptide contacts supporting the aggregate structure. This provides unfolded and isolated species for the subsequent *in vitro* refolding step and guarantees that all

inter- or intra-molecular contacts are established *de novo* as it happens after protein synthesis in the cell. IBs were chemically denatured using two chaotropic agents, 10 M urea and 8 M Gu-HCl. Each unfolded A β 42-GFP fusion was diluted in refolding buffer and the amount of recovered active GFP monitored using fluorescence spectroscopy (see Methods). The same conditions were used to unfold and refold equimolar concentrations of native untagged GFP. As it can be seen in Figure 1A, independently of the IBs peptide variant, the level of recovered GFP activity was higher when Gu-HCl was used as denaturant. This is in contrast with the results

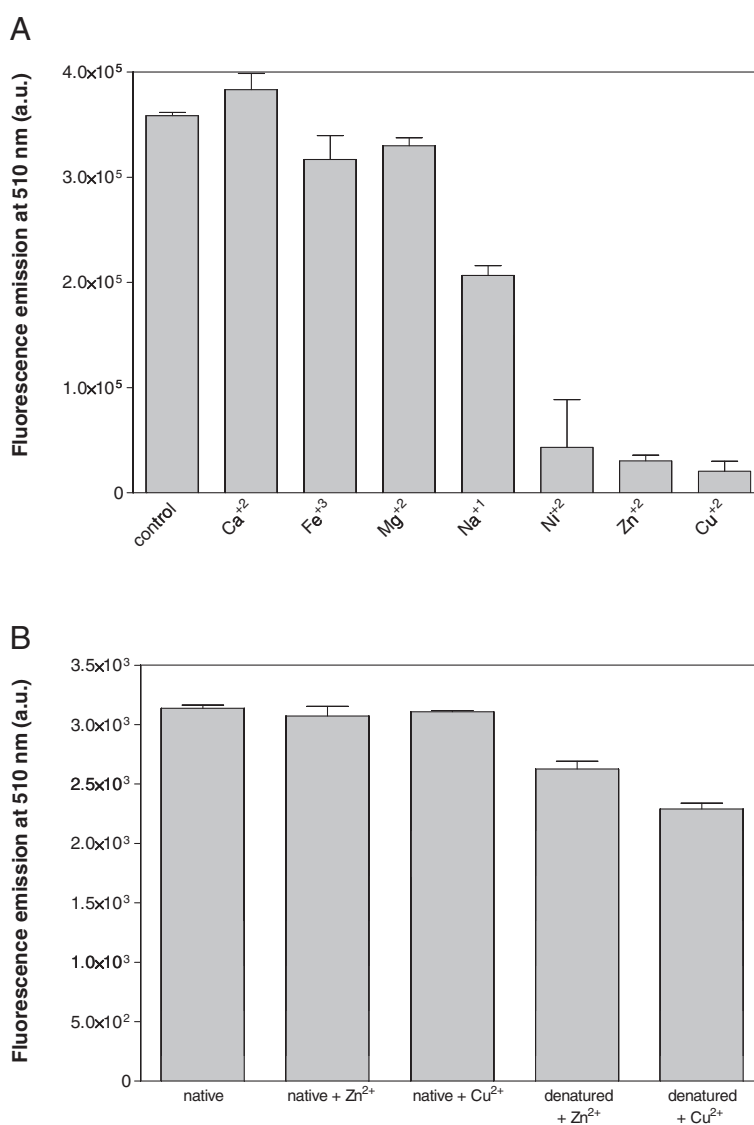


Figure 2 Fluorescence recovery in the presence of metallic ions. (A) A β 42*wt*-GFP IBs were denatured in 8 M Gu-HCl for 4 h and diluted 100-fold in PBS (control) or in PBS containing different metallic ions at 25 μ M final concentration. **(B)** Purified untagged GFP and IBs were incubated in PBS in the absence (native) and presence of 8 M Gu-HCl for 4 h and diluted 100-fold in PBS containing Cu²⁺ and Zn²⁺ at 25 μ M final concentration. In all cases, after incubation for 16 h fluorescence was recorded at 510 nm.

obtained with untagged GFP, for which denaturation with urea resulted in higher fluorescence recovery (Figure 1B), suggesting that the used denaturant might affect the aggregation/refolding pathway. The proportion of fluorescent

GFP recovered after refolding was always higher than that in the original IB (Figure 1A). Aggregation usually corresponds to a second or higher order reaction and therefore, aggregation rates are extremely dependent

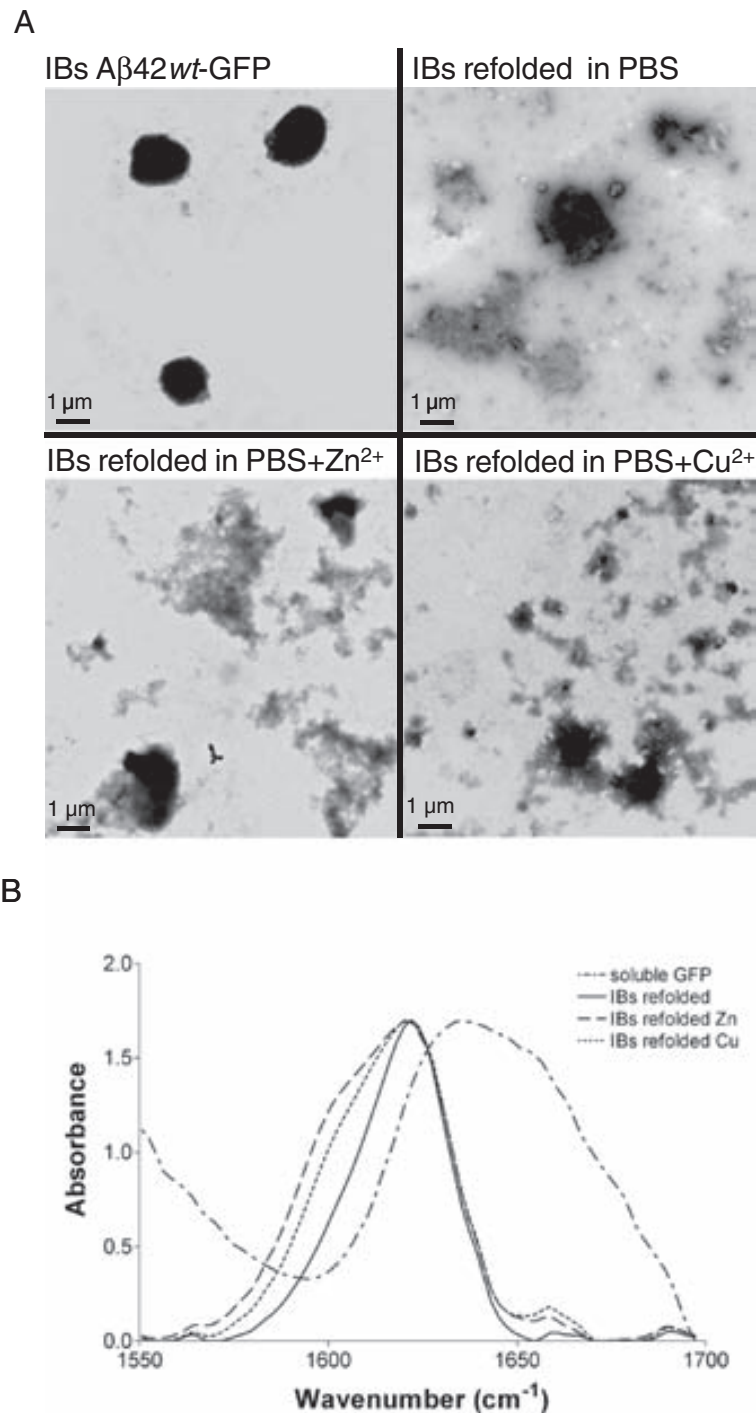


Figure 3 Morphology and secondary structure of aggregates. (A) Morphology of purified IBs and aggregates in refolding solutions in the absence and presence Cu $^{+2}$ and Zn $^{+2}$. (B) Analysis of the secondary structure of aggregates in refolding solutions in the absence and presence Cu $^{+2}$ and Zn $^{+2}$ by FT-IR spectroscopy in the amide I region of the spectra. The spectrum of native GFP is shown as a control.

on protein concentrations [28]. Since the protein concentrations used during *in vitro* refolding are much lower than those existent *in vivo*, the folding of the GFP domain can compete more efficiently with the aggregation process, providing a larger dynamic response than in bacteria. However, the refolding efficiency of A β 42-GFP IBs is about ~10-fold and ~4-fold lower than this of untagged GFP after denaturation in urea and Gu-HCl, respectively, suggesting that, as it happens *in vivo*, the aggregation of the A β 42 moiety competes the folding of GFP. Importantly, the activity recovery from the mutant IBs is higher than that from IBs formed by the *wild-type* sequence, supporting a kinetic competition between GFP folding and A β 42 aggregation *in vitro*. The predicted lower aggregation rate of the mutant would account for the higher fluorescence recovery. By analogy, any agent that would increase the intrinsic aggregation rate of A β 42 will decrease the final amount of functional GFP and *vice versa*, allowing to screen for promoters or inhibitors of the protein aggregation process.

Detection of the A β 42 aggregation-promoting effect of ionic metals

Endogenous transition metals can bind amyloid peptides, like A β 42, promoting their aggregation and the formation of amyloid fibers [29]. We analyzed if this pro-aggregating effect can be monitored using the above-described approach. Purified and Gu-HCl denatured A β 42 $_{wt}$ -GFP IBs were allowed to refold in PBS in the absence and in the presence of Ca $^{2+}$, Cu $^{2+}$, Fe $^{3+}$, Mg $^{2+}$, Na $^{+}$, Ni $^{2+}$ and Zn $^{2+}$. A highly significant decrease of GFP activity was observed in the presence of the divalent cations Cu $^{2+}$, Ni $^{2+}$ and Zn $^{2+}$ (Figure 2A). This result validates the method since there are strong evidences that zinc and copper enhance amyloid aggregation of A β 42 and are a component of the senile plaques of Alzheimer's disease patients [30]. In the case of nickel, despite being a metal that lacks physiological relevance, it has also been described to bind A β 42 and enhance the peptide cytotoxicity, via nanoscale oligomer formation, with the same potency than Cu $^{2+}$ [29]. Neither Zn $^{2+}$ nor Cu $^{2+}$ quenched the fluorescence of native untagged GFP (Figure 2B). Moreover, although the presence of Zn $^{2+}$ and Cu $^{2+}$ reduced untagged GFP fluorescence recovery, its effect was clearly lower than the one exerted on the refolding of A β 42 $_{wt}$ -GFP IBs (Figure 2B), indicating that in both cases the A β 42 peptide is a main player in the observed metal promoted aggregation. We analyzed the presence and morphology of aggregates in refolding solutions in the presence and absence of Zn $^{2+}$ and Cu $^{2+}$ by Transmission Electron Microscopy (Figure 3A). In contrast to intact IBs, which appear as electrodense spherical individual entities, all the aggregates in refolding solutions had an amorphous morphology. Nevertheless, Fourier Transformed Infrared Spectroscopy (FT-IR) analysis of the

secondary structural features of the aggregated material shows that, in all the cases, the spectra in the amide I region is dominated by a band at 1620–1625 cm $^{-1}$, typically attributed to the presence of intermolecular β -sheet, which is accompanied by a minor band at 1690 cm $^{-1}$ corresponding to the splitting of the main β -sheet signal (Figure 3B). These two bands are considered a hallmark of the presence of amyloid-like contacts. The spectra of these aggregates are significantly different from that of native GFP, in which these signatures are absent (Figure 3B).

We explored if the approach allows visualizing a concentration dependent effect of Zn $^{2+}$ and Cu $^{2+}$ on the aggregation of the target at cation concentrations in the range of the physiological levels in human brain [31]. As shown in Figure 4, the approach is highly sensitive to metal concentrations. The titration curves indicate that the impact of Cu $^{2+}$ on aggregation is somehow higher than that of Zn $^{2+}$. Curve fitting to one site binding equation renders apparent dissociation constants of 0.6 and 1.9 μ M for copper and zinc, respectively. These data are in good agreement with early reports stating that, despite the two cations bind to equivalent sites in the A β peptide, the dissociation constant for copper (0.4 μ M) is lower than that of zinc (1.2 μ M), as measured by fluorescence and H-NMR at pH 7.2 [32]. Interestingly, despite our assay is not intended for calculating dissociation constants, the ratio between the copper and zinc binding values is also ~3. Overall, the approach provides a fast qualitative assessment of metals effect on protein aggregation.

Identification of inhibitors of metal-triggered A β 42 aggregation

The identification of small molecules able to interfere protein aggregation is one of the approaches towards

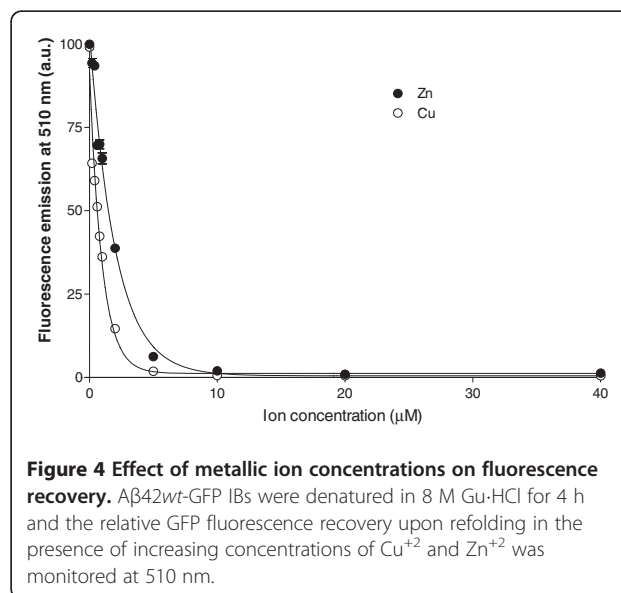


Figure 4 Effect of metallic ion concentrations on fluorescence recovery. A β 42 $_{wt}$ -GFP IBs were denatured in 8 M Gu-HCl for 4 h and the relative GFP fluorescence recovery upon refolding in the presence of increasing concentrations of Cu $^{2+}$ and Zn $^{2+}$ was monitored at 510 nm.

Table 1 Chemical structure of the small chemical compounds used in the present study

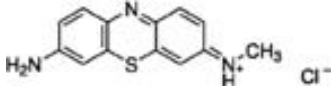
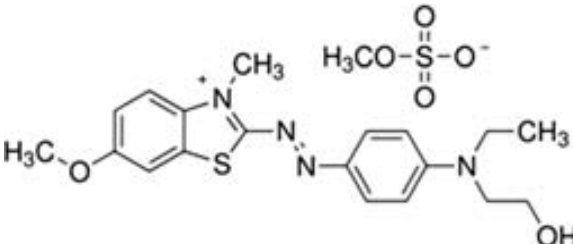
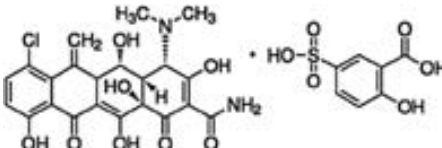
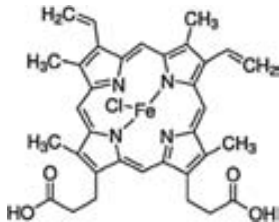
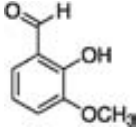
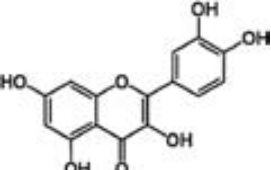
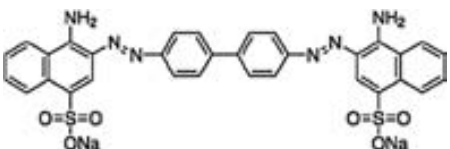
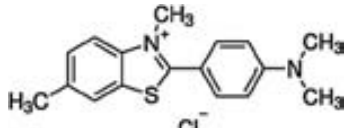
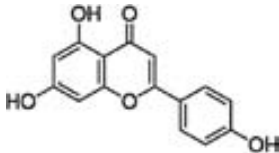
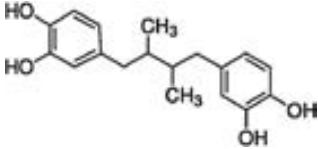
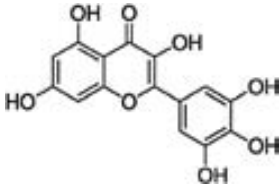
	Compound	Formula
C1	Azure C	
C2	Basic blue 41	
C3	Meclocline sulfosalicylate	
C4	Hemin	
C5	o-Vanillin	
C6	Quercetin	
C7	Congo Red	
C8	Thioflavin T	

Table 1 Chemical structure of the small chemical compounds used in the present study (Continued)

C9	Apigenin	
C10	Nordihydroguaiaretic acid	
C11	Myricetin	

therapeutic treatment in amyloid disorders [30,31]. In principle, the outlined assay could be used to screen for such compounds. In particular, in the present work we focused in validating the approach for the identification of inhibitors of copper and zinc promoted aggregation. Despite divalent chelating molecules would work *in vitro*, we discarded the study of this type of molecules since *in vivo* they have shown to sequester cofactors that are essential for the cell physiology [32]. Instead, as a test case, the IBs refolding assay was performed in the presence of selected concentrations of a collection of small compounds that have been reported previously to bind synthetic amyloid A β peptides or to modulate their aggregation and/or toxicity [33–37] but have never been assayed before in the presence of metals. The chemical formulae of the different compounds are shown in Table 1. Among the twelve tested compounds only meclocycline sulfosalicylate promoted a significant change in the final levels of GFP fluorescence in the presence of copper (Figure 5A). This compound was also active in the presence of zinc but the strongest effect in the presence of this cation was observed for *o*-Vanillin (2-Hydroxy-3-methoxybenzaldehyde) (Figure 5B). *o*-Vanillin has a cyclic structure that might quench GFP fluorescence. Effectively, the presence of 25 μ M concentration of the compound quenched 15 % of the native untagged GFP fluorescence (Figure 5C). We monitored the effect of *o*-Vanillin on the refolding of A β 42*wt*-GFP IBs in the absence of metals. The compound did not exhibit any positive effect on GFP recovery by itself and again a

17 % decrease in final fluorescence, mostly attributable to quenching, was observed. Overall, these data indicate that the presence of the compound reduces the metal-promoted aggregation effect by more than 15 fold, allowing to recover about 95% of the GFP-fluorescence observed in the absence of zinc and presence of *o*-Vanillin (Figure 5D). Interestingly, the *o*-Vanillin effect seems to be specific for zinc, with a negligibly effect for copper. This result is in agreement with previous data indicating that zinc and copper A β 42 induced aggregation pathways differ in the nature of their intermediate species and suggest that the natural product *o*-Vanillin targets specifically zinc promoted misfolding intermediates, which are characterized by a larger exposition of hydrophobic residues relative to those promoted by copper [38].

Although, to our knowledge, no *in vivo* effects of *o*-Vanillin on A β 42 promoted neuronal toxicity have been reported so far (work in progress). A closely related compound differing only in a CH₂ group, 2-Hydroxy-3-ethoxybenzaldehyde, completely blocked the neurotoxicity of the peptide to rat hippocampal neurons in culture [39], indicating that despite the simplicity of our assay, it may identify physiologically relevant hit compounds.

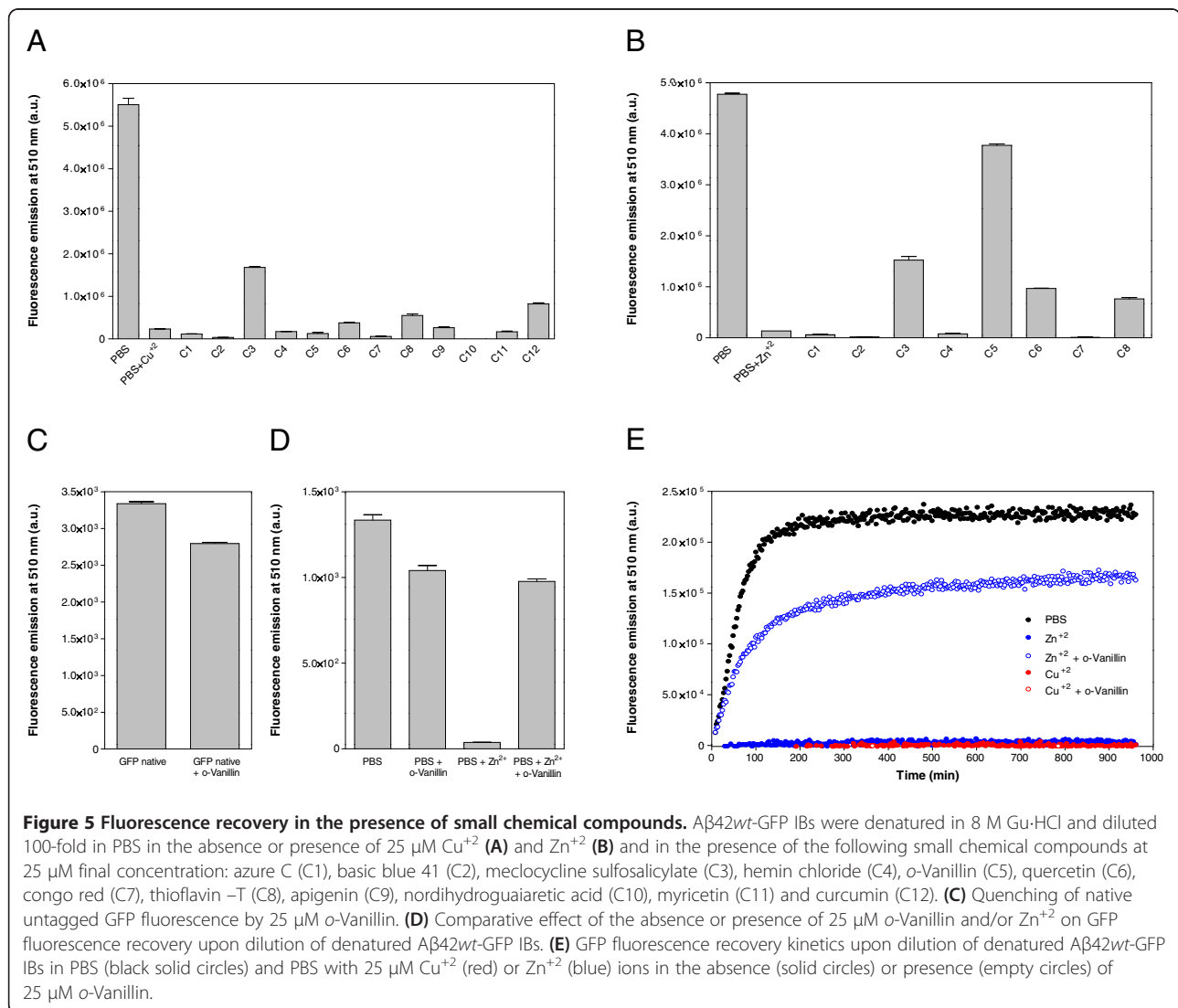
To obtain further insights on the effects of copper, zinc and *o*-Vanillin on A β 42 aggregation, we monitored the kinetics of GFP refolding after IBs denaturation in the presence and absence of these molecules by following the changes in fluorescence emission (Figure 5E). In PBS, GFP fluorescence was recovered following a

double exponential curve with a rate constant of $0.90 \pm 0.02 \text{ s}^{-1}$ and a half-life of 46.21 min for the fast reaction phase. The presence of both copper and zinc abrogated completely the fluorescence recovery already at the beginning of the refolding reaction, likely indicating that they promote a very fast aggregation of the fusion protein that totally competes the GFP domain folding reaction. The presence of *o*-Vanillin has a negligible effect on copper containing solutions. In contrast, this molecule allows recovery of 70 % of the fluorescence at the end of the reaction in the presence of zinc. The rate constants and half-life for the fast phase were very close to those exhibited in the absence of metals, with values of $0.87 \pm 0.03 \text{ s}^{-1}$ and 48.29 min, respectively. This indicates that this compound acts interfering with zinc promoted A β 42 aggregation without affecting GFP folding. Interestingly, the GFP fluorescence recovery reaction is completed after 3.5 h, being thus a faster assay than those

relying on the aggregation of synthetic peptides, which usually require at least overnight incubation [40]. We used the metallochromic Zincon reagent [41] to quantify the free levels of Zn^{2+} and Cu^{2+} in the absence and presence of *o*-Vanillin using spectrophotometry. No differences in free ion metal levels were observed (data not shown) suggesting that the compound does not act as a chelator but rather affects the refolding/aggregation kinetics of misfolded GFP fusions.

Conclusions

Based in our previous knowledge on the amyloid-like nature of the IBs formed by A β peptides [16,23] and the *in vivo* correlation between the aggregation rates and the total IBs activity [22,26], we describe here a straightforward approach to identify compounds that modulate A β aggregation using bacterial IBs. The method is implemented using 96 well plates and the



reaction takes less than four hours, making it suitable for high-throughput screening (Figure 6). Because most amyloid proteins and peptides form IBs when expressed in bacteria [17], the approach may have, in principle, a broad applicability in the search for aggregation modulators in conformational disorders. The assay does not require a detailed understanding of the structure of the

aggregating species, and can provide an unbiased method for the discovery of hit compounds. IBs can be produced and purified in large amounts, making the method cost-effective, especially when compared with the use of synthetic peptides. Despite its simplicity, the approach allows to distinguish between aggregation pathways and to identify inhibitors with therapeutic potential.

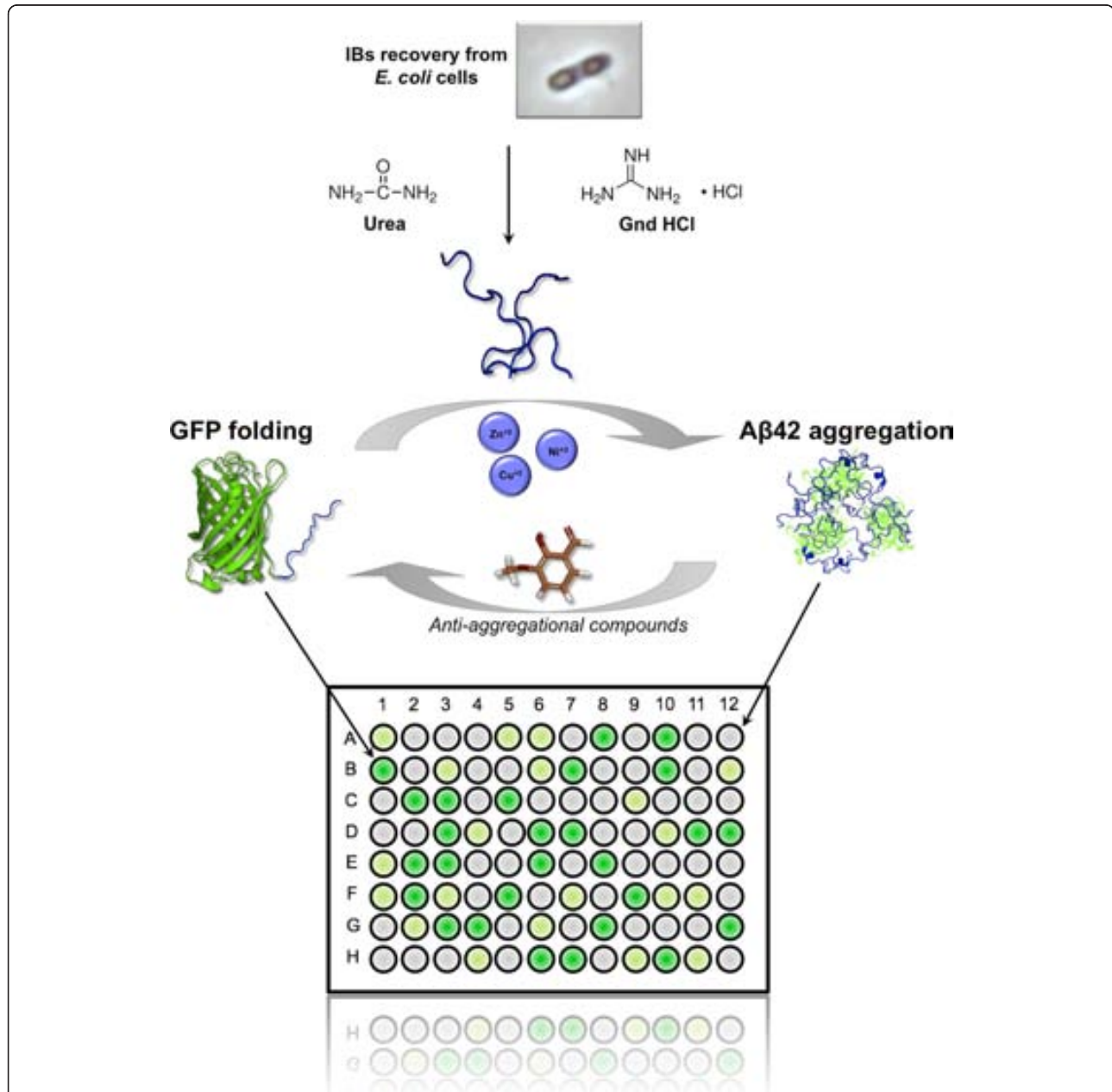


Figure 6 Outline of the screening method to identify molecules able to modulate amyloid aggregation. The method is based on the kinetic competition between the aggregation promoted by the Aβ42 moiety and the folding of the GFP domain after denaturation and refolding of Aβ42wt-GFP IBs. Molecules that accelerate the aggregation reaction result in low fluorescence recovery and *vice versa*.

Methods

Production and purification of inclusion bodies

Escherichia coli BL21DE3 competent cells were transformed with pET28 vectors (Novagen, Inc., Madison, WI, USA) encoding the sequences for A β 42 wt -GFP fusion and the mutant A β 42F19D-GFP, as previously described [25].

10 mL of bacterial cultures were grown at 37°C and 250 rpm in LB medium containing 50 μ g/mL of kanamycin. At an OD₆₀₀ of 0.5, 1 mM of isopropyl- β -D-1-thiogalactopyranoside (IPTG) was added to induce recombinant protein expression.

After 4 hours, cells were harvested by centrifugation and pellets were re-suspended in lysis buffer (100 mM NaCl, 1 mM EDTA and 50 mM Tris pH 8) to purify intracellular inclusion bodies (IBs), as previously described [41]. Briefly, protease inhibitor PMSF and lysozyme were added at the final concentrations of 15 mM and 300 μ g/mL, respectively. After incubating at 37°C for 30 min, detergent NP-40 was added at 1 % and cells were incubated at 4°C for 50 min under mild agitation. To remove nucleic acids, cells were treated with DNase and RNase at 15 μ g/mL at 37°C for 30 min. IBs were collected by centrifugation at 12,000xg for 10 min and washed with lysis buffer containing 0.5 % Triton X-100. Finally, they were washed three times with PBS to remove remaining detergent.

In vitro refolding assay

15 μ L of purified IBs at OD₃₆₀ = 10 were centrifuged for 10 min at 12000xg. To denature the aggregates, the pellets were re-suspended in 10 μ L of 8 M Gu-HCl or 10 M urea and incubated at room temperature for 4 h. For the refolding process, denatured aggregates were dissolved in 990 μ L of refolding buffer. These buffers were based on PBS, previously treated with Chelex 100 chelating resin from Sigma-Aldrich (St. Louis, MO, USA), and the following salts and compounds according to the different refolding assays: CaCl₂, FeCl₃, MgCl₂, NaCl, NiCl₂, ZnCl₂, CuCl₂, apigenin, azure C, basic blue 41, congo red, curcumin, hemin chloride, meclocycline sulfosalicylate, myricetin, nordihydroguaiaretic acid, *o*-Vanillin (2-hydroxy-3-methoxybenzaldehyde), thioflavin -T and quercetin, all obtained from Sigma-Aldrich (St. Louis, MO, USA). Equimolar concentrations of purified untagged GFP were used in control experiments. GFP fluorescence of the solutions containing refolded IBs or untagged GFP were measured in a 96 well plate in a Victor 3 Plate Reader (Perkin-Elmer, Inc., Waltham, MA, USA) using excitation and emission wavelength filters of 405 nm and 510 nm, respectively or in a Jasco FP-8200 spectrofluorometer using excitation and emission wavelengths of 480 nm and 510 nm, respectively. Measurements were performed in triplicate. For kinetic experiments, the refolding step was

followed using the same parameters and reading the fluorescence emission every 2 min for 16 h. In order to homogenize the samples, these were briefly shaken (for 5 s) before each determination.

Transmission electronic microscopy

IBs or aggregates containing solutions were placed on carbon-coated copper grids, and left to stand for five minutes. The grids were washed with distilled water and stained with 2 % (w/v) uranyl acetate for another two minutes before analysis using a HitachiH-7000 transmission electron microscope (Hitachi, Tokyo, Japan) operating at accelerating voltage of 75 kV.

Secondary structure determination

Aggregates present in refolding solutions were precipitated by centrifugation at 12.000 xg (g en cursiva i sense espais) for 30 min, resuspended in Milli-Q water and analyzed, together with purified untagged GFP, by FT-IR spectroscopy using a Bruker Tensor 27 FT-IR Spectrometer (Bruker Optics Inc) with a Golden Gate MKII ATR accessory. Each spectrum consists of 16 independent scans, measured at a spectral resolution of 2 cm⁻¹ within the 1700–1500 cm⁻¹ range. All spectral data were acquired and normalized using the OPUS MIR Tensor 27 software.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

This work was supported by grants BFU2010-14901 from Ministerio de Ciencia e Innovación (Spain) and 2009-SGR 760 from AGAUR (Generalitat de Catalunya). RS is recipient of a contract from the Ramón y Cajal Programme from Ministerio de Ciencia e Innovación. NSG is recipient of a FEBS long-term fellowship. SV has been granted an ICREA ACADEMIA award (ICREA).

Author details

¹Departament de Bioquímica i Biologia Molecular, Facultat de Biociències, Universitat Autònoma de Barcelona, E-08193, Bellaterra, Spain. ²Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, E-08193, Bellaterra, Spain. ³Present address: Departament de Físicoquímica, Facultat de Farmàcia, Universitat de Barcelona, Avda. Joan XXIII, 08028, Barcelona, Spain. ⁴Present address: Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge, CB2 0QH, United Kingdom.

Authors' contributions

SV supervised the project, designed the study and drafted the manuscript. AVP and AE carried out all experiments and drafted the manuscript. RS and NSG critically revised and corrected the manuscript. All authors read and approved the final manuscript.

Received: 21 January 2012 Accepted: 3 May 2012

Published: 3 May 2012

References

1. Chiti F, Dobson CM: Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem* 2006, **75**:333–366.
2. Fernandez-Busquets X, de Groot NS, Fernandez D, Ventura S: Recent structural and computational insights into conformational diseases. *Curr Med Chem* 2008, **15**(13):1336–1349.
3. Friedman R: Aggregation of amyloids in a cellular context: modelling and experiment. *Biochem J* 2011, **438**(3):415–426.

4. Ross CA, Poirier MA: **Protein aggregation and neurodegenerative disease.** *Nat Med* 2004, **10**(Suppl):S10–S17.
5. Lee LL, Ha H, Chang YT, DeLisa MP: **Discovery of amyloid-beta aggregation inhibitors using an engineered assay for intracellular protein folding and solubility.** *Protein Sci* 2009, **18**(2):277–286.
6. Morell M, de Groot NS, Vendrell J, Aviles FX, Ventura S: **Linking amyloid protein aggregation and yeast survival.** *Mol Biosyst* 2011, **7**(4):1121–1128.
7. Amijee H, Madine J, Middleton DA, Doig AJ: **Inhibitors of protein aggregation and toxicity.** *Biochem Soc Trans* 2009, **37**(Pt 4):692–696.
8. Hardy J, Selkoe DJ: **The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics.** *Science* 2002, **297**(5580):353–356.
9. Kuperstein I, Broersen K, Benilova I, Rozenski J, Jonckheere W, Debulpaep M, Vandersteen A, Segers-Nolten I, Van Der Werf K, Subramaniam V, et al: **Neurotoxicity of Alzheimer's disease Abeta peptides is induced by small changes in the Abeta42 to Abeta40 ratio.** *EMBO J* 2010, **29**(19):3408–3420.
10. Karran E, Mercken M, De Strooper B: **The amyloid cascade hypothesis for Alzheimer's disease: an appraisal for the development of therapeutics.** *Nat Rev Drug Discov* 2011, **10**(9):698–712.
11. Bonda DJ, Lee HG, Blair JA, Zhu X, Perry G, Smith MA: **Role of metal dyshomeostasis in Alzheimer's disease.** *Metallomics* 2011, **3**(3):267–270.
12. Jomova K, Vondrakova D, Lawson M, Valko M: **Metals, oxidative stress and neurodegenerative disorders.** *Mol Cell Biochem* 2010, **345**(1–2):91–104.
13. de Groot NS, Espargaro A, Morell M, Ventura S: **Studies on bacterial inclusion bodies.** *Future Microbiol* 2008, **3**:423–435.
14. Ventura S, Villaverde A: **Protein quality in bacterial inclusion bodies.** *Trends Biotechnol* 2006, **24**(4):179–185.
15. Carrio M, Gonzalez-Montalban N, Vera A, Villaverde A, Ventura S: **Amyloid-like properties of bacterial inclusion bodies.** *J Mol Biol* 2005, **347**(5):1025–1037.
16. Morell M, Bravo R, Espargaro A, Sisquella X, Aviles FX, Fernandez-Busquets X, Ventura S: **Inclusion bodies: specificity in their aggregation process and amyloid-like structure.** *Biochim Biophys Acta* 2008, **1783**(10):1815–1825.
17. de Groot NS, Sabate R, Ventura S: **Amyloids in bacterial inclusion bodies.** *Trends Biochem Sci* 2009, **34**(8):408–416.
18. Wang L, Maji SK, Sawaya MR, Eisenberg D, Riek R: **Bacterial inclusion bodies contain amyloid-like structure.** *PLoS Biol* 2008, **6**(8):e195.
19. Sabate R, de Groot NS, Ventura S: **Protein folding and aggregation in bacteria.** *Cell Mol Life Sci* 2010, **67**(16):2695–2715.
20. Garcia-Fruitos E, Sabate R, de Groot NS, Villaverde A, Ventura S: **Biological role of bacterial inclusion bodies: a model for amyloid aggregation.** *FEBS J* 2011, **278**(14):2419–2427.
21. Lotti M: **Bacterial inclusion bodies as active and dynamic protein ensembles.** *FEBS J* 2011, **278**(14):2407.
22. Villar-Pique A, de Groot NS, Sabate R, Acebron SP, Celaya G, Fernandez-Busquets X, Muga A, Ventura S: **The Effect of Amyloidogenic Peptides on Bacterial Aging Correlates with Their Intrinsic Aggregation Propensity.** *J Mol Biol* 2011. doi:http://dx.doi.org/10.1016/j.jmb.2011.12.014.
23. Dasari M, Espargaro A, Sabate R, Lopez Del Amo JM, Fink U, Grelle G, Bieschke J, Ventura S, Reif B: **Bacterial Inclusion Bodies of Alzheimer's Disease beta-Amyloid Peptides Can Be Employed To Study Native-Like Aggregation Intermediate States.** *Chem Bio Chem* 2011, **12**(3):407–423.
24. Garcia-Fruitos E, Gonzalez-Montalban N, Morell M, Vera A, Ferraz RM, Aris A, Ventura S, Villaverde A: **Aggregation as bacterial inclusion bodies does not imply inactivation of enzymes and fluorescent proteins.** *Microb Cell Fact* 2005, **4**:27.
25. de Groot NS, Aviles FX, Vendrell J, Ventura S: **Mutagenesis of the central hydrophobic cluster in Abeta42 Alzheimer's peptide, Side-chain properties correlate with aggregation propensities.** *FEBS J* 2006, **273**(3):658–668.
26. de Groot NS, Ventura S: **Protein activity in bacterial inclusion bodies correlates with predicted aggregation rates.** *J Biotechnol* 2006, **125**(1):110–113.
27. de Groot NS, Ventura S: **Effect of temperature on protein quality in bacterial inclusion bodies.** *FEBS Lett* 2006, **580**:6471–6476.
28. Kieflhaber T, Rudolph R, Kohler HH, Buchner J: **Protein aggregation in vitro and in vivo: a quantitative model of the kinetic competition between folding and aggregation.** *BioTechnology* 1991, **9**(9):825–829.
29. Jin L, Wu WH, Li QY, Zhao YF, Li YM: **Copper inducing Abeta42 rather than Abeta40 nanoscale oligomer formation is the key process for Abeta neurotoxicity.** *Nanoscale* 2011, **3**(11):4746–4751.
30. Zhang X, Smith DL, Meriin AB, Engemann S, Russel DE, Roark M, Washington SL, Maxwell MM, Marsh JL, Thompson LM, et al: **A potent small molecule inhibits polyglutamine aggregation in Huntington's disease neurons and suppresses neurodegeneration in vivo.** *Proc Natl Acad Sci U S A* 2005, **102**(3):892–897.
31. Chen J, Armstrong AH, Koehler AN, Hecht MH: **Small molecule microarrays enable the discovery of compounds that bind the Alzheimer's Abeta peptide and reduce its cytotoxicity.** *J Am Chem Soc* 2010, **132**(47):17015–17022.
32. Hegde ML, Bharathi P, Suram A, Venugopal C, Jagannathan R, Poddar P, Srinivas P, Sambamurti K, Rao KJ, Scancar J, et al: **Challenges associated with metal chelation therapy in Alzheimer's disease.** *J Alzheimer's Dis* 2009, **17**(3):457–468.
33. Necula M, Kaye R, Milton S, Glabe CG: **Small molecule inhibitors of aggregation indicate that amyloid beta oligomerization and fibrillation pathways are independent and distinct.** *J Biol Chem* 2007, **282**(14):10311–10324.
34. Necula M, Breydo L, Milton S, Kaye R, van der Veer WE, Tone P, Glabe CG: **Methylene blue inhibits amyloid Abeta oligomerization by promoting fibrillation.** *Biochemistry* 2007, **46**(30):8850–8860.
35. Kim H, Park BS, Lee KG, Choi CY, Jang SS, Kim YH, Lee SE: **Effects of naturally occurring compounds on fibril formation and oxidative stress of beta-amyloid.** *J Agric Food Chem* 2005, **53**(22):8537–8541.
36. Jagota S, Rajadas J: **Effect of Phenolic Compounds Against Abeta Aggregation and Abeta-Induced Toxicity in Transgenic C. elegans.** *Neurochem Res* 2012, **37**(1):40–48.
37. Matsuzaki K, Noguch T, Wakabayashi M, Ikeda K, Okada T, Ohashi Y, Hoshino M, Naiki H: **Inhibitors of amyloid beta-protein aggregation mediated by GM1-containing raft-like membranes.** *Biochim Biophys Acta* 2007, **1768**(1):122–130.
38. Chen WT, Liao YH, Yu HM, Cheng IH, Chen YR: **Distinct effects of Zn²⁺, Cu²⁺, Fe³⁺, and Al³⁺ on amyloid-beta stability, oligomerization, and aggregation: amyloid-beta destabilization promotes annular protofibril formation.** *J Biol Chem* 2011, **286**(11):9646–9656.
39. De Felice FG, Vieira MN, Saraiva LM, Figueroa-Villar JD, Garcia-Abreu J, Liu R, Chang L, Klein WL, Ferreira ST: **Targeting the neurotoxic species in Alzheimer's disease: inhibitors of Abeta oligomerization.** *FASEB J* 2004, **18**(12):1366–1372.
40. Rodriguez-Rodriguez C, Sanchez De Groot N, Rimola A, Alvarez-Larena A, Lloervera V, Vidal-Gancedo J, Ventura S, Vendrell J, Sodupe M, Gonzalez-Duarte P: **Design, selection, and characterization of thioflavin-based intercalation compounds with metal chelating properties for application in Alzheimer's disease.** *J Am Chem Soc* 2009, **131**(4):1436–1451.
41. Hilario E, Romero I, Celis H: **Determination of the physicochemical constants and spectrophotometric characteristics of the metallochromic Zincon and its potential use in biological systems.** *J Biochem Biophys Methods* 1990, **21**(3):197–207.

doi:10.1186/1475-2859-11-55

Cite this article as: Villar-Piqué et al.: Using bacterial inclusion bodies to screen for amyloid aggregation inhibitors. *Microbial Cell Factories* 2012 **11**:55.

Submit your next manuscript to BioMed Central and take full advantage of:

- **Convenient online submission**
- **Thorough peer review**
- **No space constraints or color figure charges**
- **Immediate publication on acceptance**
- **Inclusion in PubMed, CAS, Scopus and Google Scholar**
- **Research which is freely available for redistribution**

Submit your manuscript at
www.biomedcentral.com/submit



3.3 Publication III

Linking intracellular protein aggregation propensity and protein quality control degradation.

Anna Villar-Piqué and Salvador Ventura

Manuscript for incoming submission.

This publication is included in this as annex and, thus, it is located in appendices.

3.4 Publication IV

Amyloid-like protein inclusions in tobacco transgenic plants.

Anna Villar-Piqué, Raimon Sabaté, Oriol Lopera, Jordi Gibert, Josep Maria Torné,
Mireya Santos and Salvador Ventura

PLoS One 5(10): e13625.
26 October 2010

Amyloid-Like Protein Inclusions in Tobacco Transgenic Plants

Anna Villar-Piqué¹, Raimon Sabaté¹, Oriol Lopera², Jordi Gibert², Josep Maria Torne², Mireya Santos², Salvador Ventura^{1*}

1 Institut de Biotecnologia i de Biomedicina and Departament de Bioquímica i Biologia Molecular, Universitat Autònoma de Barcelona, Bellaterra, Spain, **2** Centre for Research in Agricultural Genomics (CRAG) CSIC-IRTA-UAB, Molecular Genetics Laboratory, Barcelona, Spain

Abstract

The formation of insoluble protein deposits in human tissues is linked to the onset of more than 40 different disorders, ranging from dementia to diabetes. In these diseases, the proteins usually self-assemble into ordered β -sheet enriched aggregates known as amyloid fibrils. Here we study the structure of the inclusions formed by maize transglutaminase (TGZ) in the chloroplasts of tobacco transplastomic plants and demonstrate that they have an amyloid-like nature. Together with the evidence of amyloid structures in bacteria and fungi our data argue that amyloid formation is likely a ubiquitous process occurring across the different kingdoms of life. The discovery of amyloid conformations inside inclusions of genetically modified plants might have implications regarding their use for human applications.

Citation: Villar-Piqué A, Sabaté R, Lopera O, Gibert J, Torne JM, et al. (2010) Amyloid-Like Protein Inclusions in Tobacco Transgenic Plants. PLoS ONE 5(10): e13625. doi:10.1371/journal.pone.0013625

Editor: Anna Mitragi, University of Crete, Greece

Received: May 28, 2010; **Accepted:** October 1, 2010; **Published:** October 26, 2010

Copyright: © 2010 Villar-Piqué et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by grants BIO2007-68046, BFU2006-15115-CO2-01/BMC and BFU2009-08575 from Ministerio de Ciencia e Innovación (Spain), by grant 2009-SGR 760 from AGAUR(Agencia de Gestió d'Ajuts Universitaris i de Recerca-Generalitat de Catalunya). AV was beneficiary of a PIF fellowship from Universidad Autónoma de Barcelona (UAB). RS was beneficiary of an I3 contract (UAB-Generalitat de Catalunya). JG was beneficiary of a FPI fellowship awarded by Generalitat de Catalunya. SV has been granted with an ICREA ACADEMIA award (ICREA-Institució Catalana de Recerca i Estudis Avancats). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: salvador.ventura@uab.es

Introduction

The intracellular aggregation of polypeptides is a pathogenic feature of cellular degeneration in many human degenerative disorders [1,2,3]. Intracellular protein aggregates are formed when misfolded polypeptides accumulate in the cells due to malfunctioning or overloading of the protein quality control machinery or of the components of the degradative pathway [4]. Many disease-associated protein aggregates are composed of filaments known as amyloid fibrils. Amyloid fibrils bind to Thioflavin T (Th-T) and Congo red (CR) due to their repetitive intermolecular β -sheet architecture [5,6]. It has been shown that the ability to self-assemble into amyloid-like structures is not an unusual feature exhibited by a reduced set of disease-associated molecules with special sequence or structural properties, but rather a property shared by many polypeptides [7]. In addition, the formation of amyloid-like aggregates in living cells is not restricted to animals but has also been observed in fungi [8,9] and bacteria [10,11,12]. Although the formation of amyloids by plant pathogenic bacteria in infected leaves has been recently reported [13], to the best of our knowledge, the formation of amyloid-like deposits in plants by plant-encoded proteins has not been described yet.

The ability to genetically modify plants has allowed the bioproduction of heterologous proteins. In the last decade, plants have become an alternative source for the cost effective production of recombinant polypeptides for therapeutics in animal and human health and diagnostics [14]. The chloroplasts of higher plants are bounded by two envelope membranes that surround an aqueous matrix, the stroma, and the internal photosynthetic membranes, the

thylakoids. In chloroplast transformation, and differing from nuclear transformation, the transgene is integrated in the plastid genome via homologous recombination. The flanking sequences of the transformation vector, homologous to the plastid genome, direct the transgene to a specific and unique location without gene silencing, permitting the expression of the desired protein into the chloroplast without needing many generations of gene selection [15]. Transglutaminases (TGases) catalyse post-translational modification of structural proteins by establishing ϵ -(γ -glutamyl) links and covalent conjugation of polyamines. These proteins are widely distributed in bacteria, animals and plants. Human TGase has been associated to the progression of several neurodegenerative diseases [16]. In plants, this enzyme is poorly characterized and only the maize plastidial TGase gene (*tgz*) has been cloned to date (Patent number WO03102128) [17]. Variants of this TGase have been expressed recombinantly in *Escherichia coli* [18] and *tgz*-transplastomic tobacco plants engineered [19]. Here we use Th-T and CR binding, Fourier Transformed Infrared Spectroscopy (FT-IR) and Transmission Electronic Microscopy (TEM) to study the conformational properties of the protein deposits formed by maize transglutaminase (TGZ) *in vitro* and in the chloroplasts of transplastomic plants, demonstrating that in both cases they exhibit characteristic amyloid features.

Results

Maize transglutaminase forms amyloid-like aggregates *in vitro*

We have used two different bioinformatic approaches to detect the presence amyloidogenic regions in TGZ, namely the

AGGRESCAN [20] and TANGO [21] algorithms. Both programs coincide to indicate the concentration of aggregation promoting sequences at the C-terminus of the protein. The sequence stretch comprising residues 466–477 is consistently predicted to be a region with high amyloidogenic propensity (Figure S1) suggesting that TGZ might have the capability to aggregate into structures displaying amyloid features.

TGZ was recombinantly produced in *E. coli*, purified from the insoluble fraction and unfolded in 6 M guanidine hydrochloride. After refolding at 4°C, the protein self-assembles into observable aggregates when incubated at 25°C for one week. The secondary structure content of these aggregates was evaluated by ATR FT-IR in the amide I region of the spectrum (Figure 1A). The second derivative of the absorbance spectrum in this region is dominated

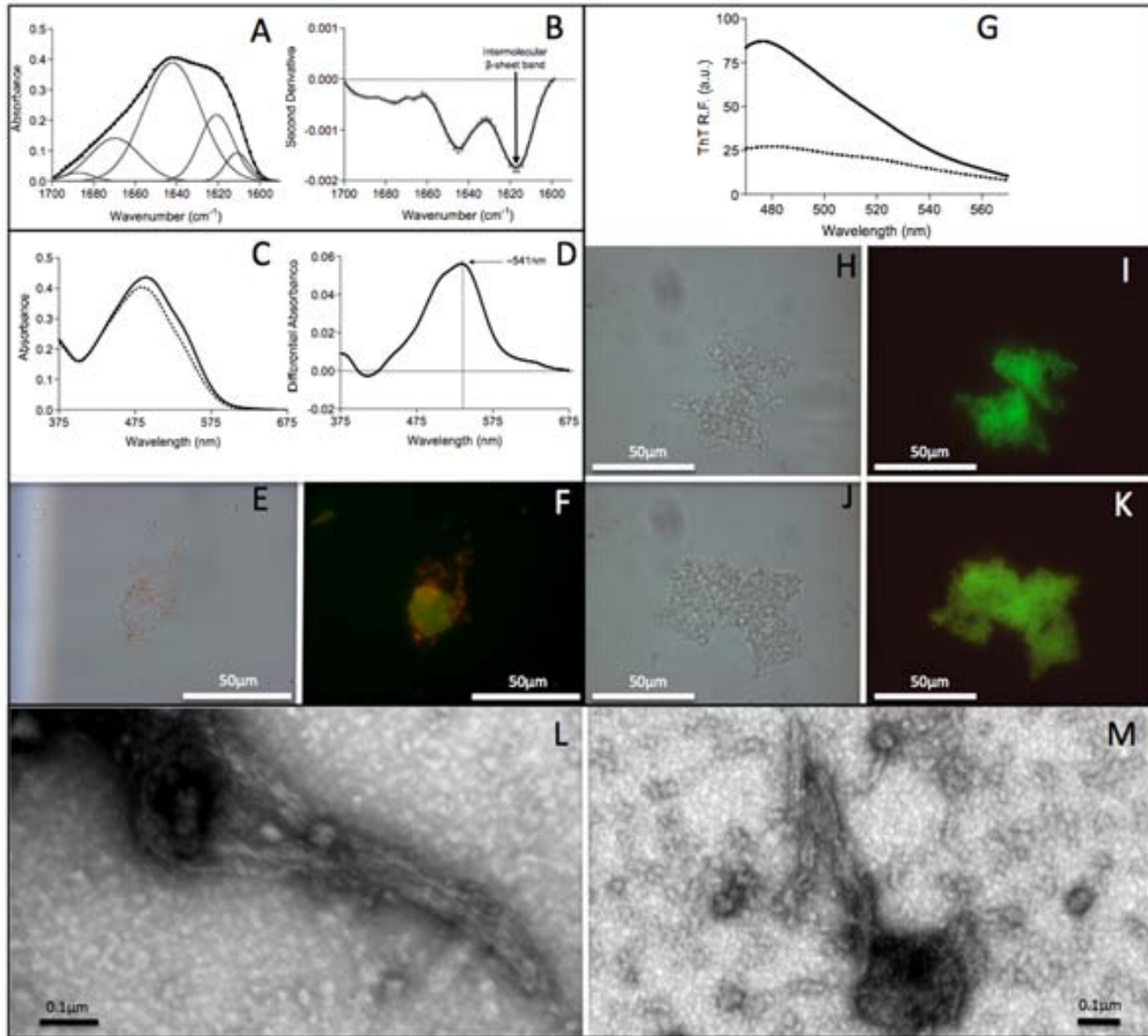


Figure 1. Amyloid-like properties of *in vitro* maize TGZ aggregates. A) Secondary structure of *in vitro* TGZ aggregates as measured by FTIR absorbance in the amide I region (solid thick lines) showing the spectral component bands (solid thin lines) and the characteristic bands corresponding to intermolecular β -sheet conformations. B) Second derivative of the FTIR absorbance spectra shown in panel A. C) Absorption spectra of Congo red (CR) in the presence (solid line) and absence (dashed line) of TGZ aggregates. Changes in λ_{\max} and intensity in CR spectra are observed in the presence of *in vitro* formed aggregates. D) Difference absorbance spectra of Congo red in the presence of TGZ aggregates showing the characteristic amyloid band at ~ 541 nm. E) Bright field image of TGZ aggregates stained with Congo red. F) The same image under cross-polarized light showing the characteristic amyloid birefringence. Both images at 40-fold magnification. G) Fluorescence emission spectra of Th-T in the presence (solid line) and absence (dashed line) of TGZ aggregates. The fluorescence intensity increases by three fold in the presence of *in vitro* formed aggregates. H) Bright field image of TGZ aggregates stained with Th-T. I) The same aggregates viewed by fluorescence microscopy under UV light. Both images at 100-fold magnification. J) Bright field image of TGZ aggregates stained with Th-S. K) The same aggregates viewed by fluorescence microscopy under UV light. Both images at 100-fold magnification. L and M) Fibrillar morphology of *in vitro* TGZ aggregates as monitored by transmission electronic microscopy. doi:10.1371/journal.pone.0013625.g001

by a peak at $\sim 1618\text{ cm}^{-1}$ (Figure 1B). This signal is typically associated to the presence of intermolecular β -sheet structure. Deconvolution of the absorbance spectrum into its main components (Figure 1A) suggests that this peak arises from the combination of two signals at ~ 1610 and ~ 1620 , both indicative of the existence of short hydrogen bonds between β -strands and compatible with an amyloid like-conformation in the aggregates.

The *in vitro* formed aggregates of TGZ bind to the amyloid diagnostic dye Congo Red (CR) as evidenced by the increase in the absorbance signal and shift of the spectrum towards higher wavelengths (Figure 1C). The different spectrum between the dye in the presence and absence of aggregates allows detecting the characteristic amyloid-like band at $\sim 541\text{ nm}$ (Figure 1D). In addition, TGZ aggregates incubated with CR display a characteristic amyloid-like green-yellow birefringence when illuminated under cross-polarized light (Figure 1E and 1F).

We further explored the properties of *in vitro* TGZ aggregates by measuring their binding to Th-T. A threefold increase in the maximum emission at 482 was observed (Figure 1G). This change in fluorescence is consistent with TGZ being in an amyloid conformation. The binding of Th-T and the related amyloid dye Thioflavin-S (Th-S) to aggregates was also visualized by using fluorescence microscopy (Figure 1H to 1K). In both cases areas rich in aggregated material were stained with Th-T giving a bright green or green-yellow fluorescence against a dark background.

We monitored the morphology of *in vitro* formed TGZ aggregates by TEM. The presence of abundant bundles of fibrillar structures with dimensions compatible with an amyloid nature could be observed (Figure 1L and 1M).

A short C-terminal peptide of maize transglutaminase forms amyloid fibrils

Polypeptide sequences might contain local regions with high aggregation propensity that can nucleate the early steps of aggregation [22,23]. As described above, the region comprising residues 466–477 at the C-terminus has the highest predicted aggregation propensity in the TGZ sequence. To assay if this region has the ability to self-assemble and act as possible nucleation element in the aggregation process of TGZ we synthesized and characterized the amyloidogenic properties of the peptide QLVVLDILLGKFS corresponding to TGZ residues 465–477 (Glu-465 was included to provide solubility to the peptide during its synthesis). The peptide was incubated in 50 mM TRIS at pH 7.5, 150 mM NaCl at 25°C for 48 h at 100 μM in quiescent or agitated conditions. In both cases the formation of fibrillar structures with size and morphology compatible with an amyloid nature could be observed by TEM. Fibrils formed under quiescent conditions were longer (Figure 2A and 2B) than those in agitated samples, which tend to cluster together (Figure 2C and 2D). We analyzed the secondary structure content of quiescent fibrils by ATR FT-IR in the amide I region of the spectrum (Figure 2E and 2F). The second derivative of the absorbance spectrum in this region is dominated by a peak at $\sim 1626\text{ cm}^{-1}$ (Figure 2F) confirming the presence of intermolecular β -sheet structure in the fibrils. Deconvolution of the absorbance spectrum into its main components (Figure 2E) results in five main signals corresponding to the presence of extended β -sheets (1604 , 1628 and 1692 cm^{-1}) and turns (1660 and 1680 cm^{-1}) compatible with an amyloid conformation of the peptide inside the fibrils. The strong changes promoted by the fibrils in the absorbance spectrum of CR (Figure 2G and 2H) and the fluorescence spectrum of Th-T (Figure 2I) confirm the amyloidogenic properties of the fibrillar structures formed by the most aggregation-prone sequence of TGZ.

Maize transglutaminase forms amyloid-like aggregates in the leaves of transplastomic tobacco plants

Homoplasmic tobacco *tgz*-transgenic plants presented abnormal phenotype with respect to the leaf colour, having pigment deficiencies (Fig. 3A) and thylakoid appression abnormalities (Figure 3B). The TGZ protein was immunolocalized into chloroplast inclusion bodies (Figure 3C insert), suggesting that in the plant the protein is present in an at least partially aggregated state, which might coexist with functional conformations as shown for bacterial inclusion bodies [24]. The thylakoids in the chloroplasts of non-transgenic plants displayed a normal arrangement with grana stacks consisting of 15–20 tightly appressed thylakoid membranes interconnected by stroma thylakoids (Figure 3D).

We analyzed the protein content of the soluble and insoluble fractions of transgenic plants by SDS-PAGE and Western Blot using an anti-TGZ antibody to determine if TGZ is effectively found in an aggregated state *in vivo* (Figure 4). In spite of the much higher protein content of the soluble fraction, TGZ is absolutely absent in this fraction and localizes exclusively into the insoluble fraction, in which constitutes a major protein component. Three different types of TGZ bands are detected by Western Blot in the insoluble fraction upon SDS-denaturation: a first band corresponding to a truncated species, according to its smaller size when compared with purified TGZ, a second band corresponding to the full length monomeric protein and several intense high molecular bands corresponding to SDS-resistant aggregated species. This SDS-resistant species resemble the oligomeric species found in aggregated solutions of amyloid proteins like A β -peptide [25]. Like in the case of amyloid assemblies, in addition to SDS, high chaotropic reagent concentrations are required to disrupt these aggregated species, indicating that they are stabilized by strong intermolecular interactions.

To analyze if the aggregates formed by TGZ in transplastomic tobacco plants display amyloid features similar to those observed *in vitro* we isolated the protein insoluble fraction. The same amount of WT tobacco plant leaves were fractionated and analyzed simultaneously as a negative control. The ATR FT-IR spectrum in the amide I region of transplastomic aggregates is significantly different from that of WT aggregates (Figure 5A). The spectrum of transplastomic aggregates is dominated by an intermolecular β -sheet band at $\sim 1620\text{ cm}^{-1}$ whereas that of WT plants presents a major band at $\sim 1656\text{ cm}^{-1}$ associated to unstructured and/or α -helical conformation. The second derivative of the spectrum confirms that the intermolecular β -sheet is the main secondary component of transplastomic protein aggregates (Figure 5B).

Consistently with a β -sheet enriched architecture, transplastomic protein aggregates bind to Th-T (Figure 5C) and CR (Figure 5D) promoting the expected spectral changes for amyloid-like structures. Although WT aggregates also exhibit some binding to these dyes the spectral changes they promote are much weaker than in the case of transgenic plants (Figure 5C and 5D).

We further analyzed the morphology of both types of protein deposits by TEM. In both cases we could detect the presence of aggregates. However, they display different structural properties. Whereas, in most of the transplastomic samples the presence of abundant fibrillar material could be observed (Figure 5F), these types of structures are absent in WT samples and their aggregates appear as amorphous material (Figure 5E). In addition, the presence of isolated bundles of fibrillar structures with dimensions and morphology compatible with amyloids could be observed in transplastomic aggregates (Figure 5F) while despite the analysis of large number of fields we could not observe such fibrils in WT samples and its aggregates appear to lack any regular structure (Figure 5E).

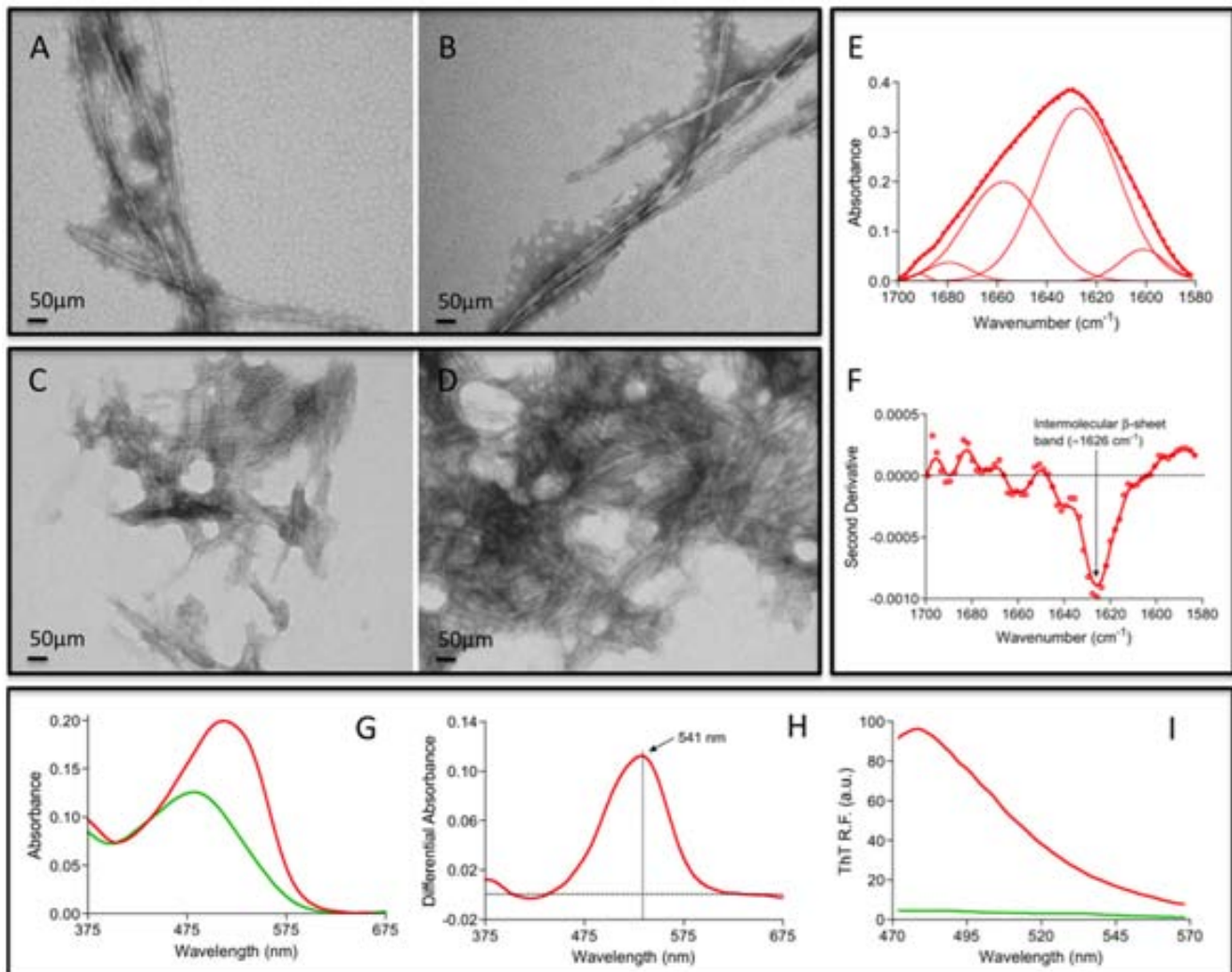


Figure 2. Amyloid-like properties of a C-terminal peptide of TGZ. The peptide QLVLVDILLGKFS corresponds to residues 465–477 of TGZ. A–D) Fibrillar morphology of peptide aggregates formed under quiescence (A and B) and agitation at 500 rpm (C and D) as monitored by transmission electronic microscopy. E–F) Secondary structure of peptide fibrils measured by FTIR absorbance (E) in the amide I region (solid thick lines) showing the spectral component bands (solid thin lines) and second derivative of the FTIR absorbance spectra (F) showing the characteristic bands corresponding to intermolecular β -sheet conformations. G–I) Amyloid specific dyes staining of peptide fibrils. G) Absorption spectra of Congo Red (CR) in presence (in red) and absence (in green) of peptide fibrils. Changes in λ_{\max} and intensity in CR spectra are observed in presence of fibrils. H) Difference absorbance spectra of Congo red in the presence of peptide fibrils showing the characteristic amyloid band at ~ 541 nm. I) Thioflavin-T (Th-T) fluorescence emission spectra in the presence (in green) and absence (in red) of peptide fibrils. The fluorescence intensity increases by 20-fold in the presence of these aggregates.

doi:10.1371/journal.pone.0013625.g002

Discussion

In the present work we show that maize TGase (TGZ) displays an intrinsic propensity to form aggregates displaying amyloid-like features when it is refolded *in vitro* in the absence of the components of the protein quality control machinery. The C-terminal TGZ sequence stretch comprising residues 465–477 is able to form highly ordered amyloid fibrils and may act as trigger of the aggregation process. *In vivo*, at any given time, the concentration of a protein in its native state results from a balance between the rate of protein synthesis, the rate of de novo folding, the stability of the protein conformation and the degradation rate [26]. The continuous and high translation rates occurring during the expression of proteins in transplastomic plants tends to unbalance this equilibrium, saturating and/or de-coordinating the mechanisms to assist the process of protein folding or the

degradation of misfolded species, which ultimately would result in the accumulation of a significant population of non-native conformers, that in the case of TGZ, due to its intrinsic tendency to aggregate, accumulate as insoluble inclusion bodies into the chloroplast. These aggregates consist, at least in part, of fibrillar material displaying amyloid-like features. Although TGZ species appears as major component of the insoluble fraction in the leaves of transplastomic plants, the possibility that the observed amyloid structures would contain other endogenous plant proteins that co-aggregate with TGZ as observed for protein aggregates formed in bacteria [27] or in mammalian cells [28] cannot be discarded. The saturation of the folding machinery and degradative pathways by misfolded TGZ species as well as the refractivity of TGZ aggregates to proteolysis might be all factors contributing to the accumulation of the observed insoluble material. The presence of truncated forms of TGZ in insoluble fraction of transgenic plant

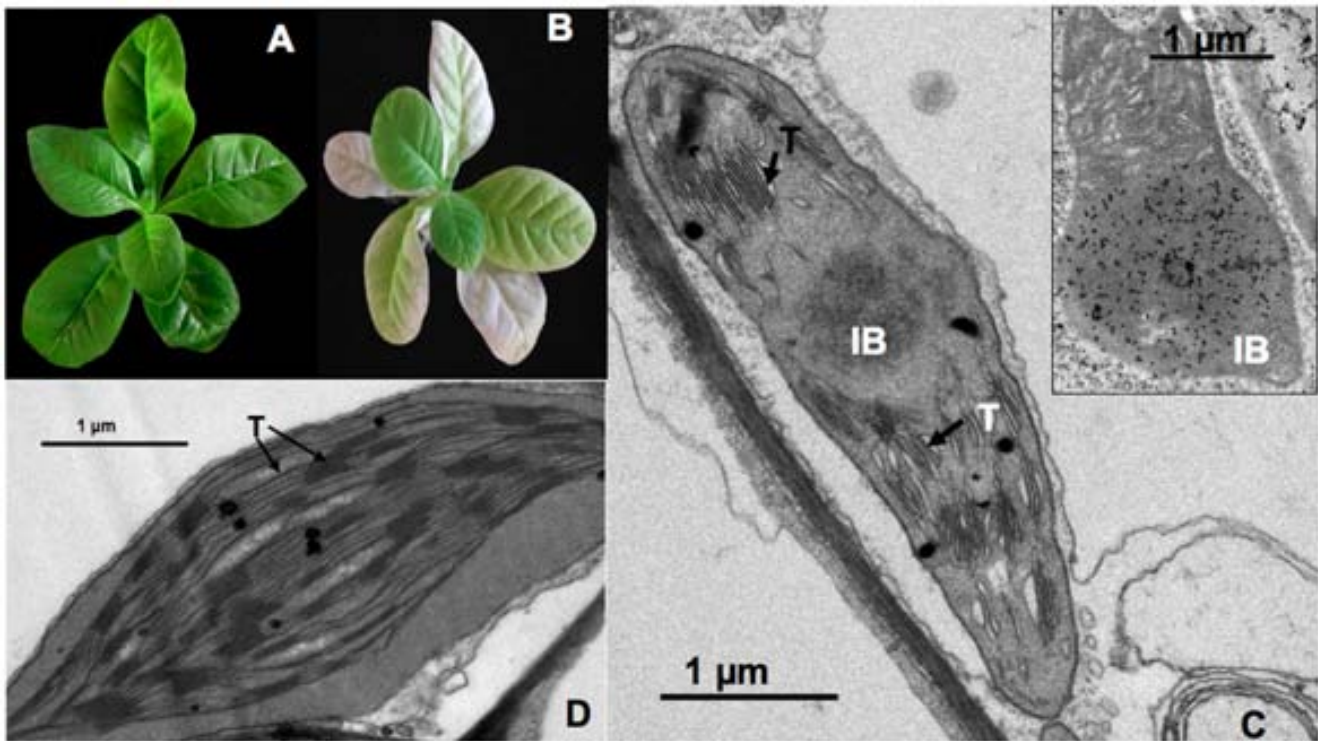


Figure 3. Maize TGZ forms protein inclusions in transplastomic tobacco chloroplasts. A) Aspect of wild type tobacco plant. B) Aspect of *tgz*-transplastomic tobacco plant. C) TEM image of a *tgz*-transplastomic tobacco chloroplast. A high number of appressed thylakoid membranes (arrows), membrane interruptions and IB presence are shown. Inside: subcellular immunolocalization of TGZ protein in the IB of a tobacco *tgz*-transformed chloroplast using an anti-TGZ antibody (1:3000) (see M & M). D) TEM image of a WT tobacco chloroplast showing normal thylakoid membranes and interconnexions. IB, inclusion body; T, thylakoids.
doi:10.1371/journal.pone.0013625.g003

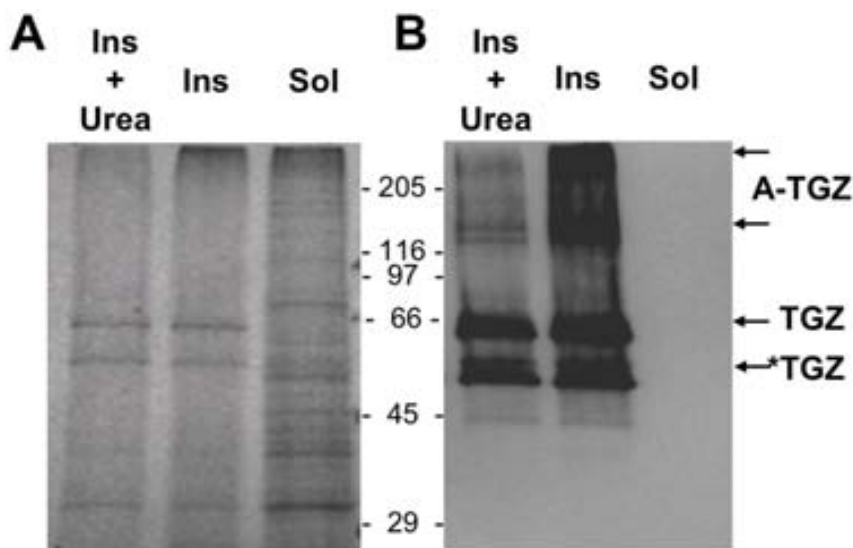


Figure 4. Localization of maize TGZ in the insoluble fraction of leaves in transplastomic tobacco plants. A) SDS-PAGE analysis of the soluble (Sol) and insoluble fraction of transplastomic leaves before (Ins) and after 8 M Urea incubation (Ins + Urea). B) Western Blot of the soluble (Sol) and insoluble fraction of transplastomic leaves before (Ins) and after 8 M Urea incubation (Ins + Urea) using a polyclonal antibody raised against maize TGZ protein (see M & M). Three main types of TGZ species are detected: *TGZ corresponds to a truncated species, TGZ to the full-length protein and A-TGZ to SDS-resistant aggregated forms. The amount of A-TGZ decreases significantly in the presence of 8 M Urea. The positions of molecular weight markers are indicated (in kDa).
doi:10.1371/journal.pone.0013625.g004

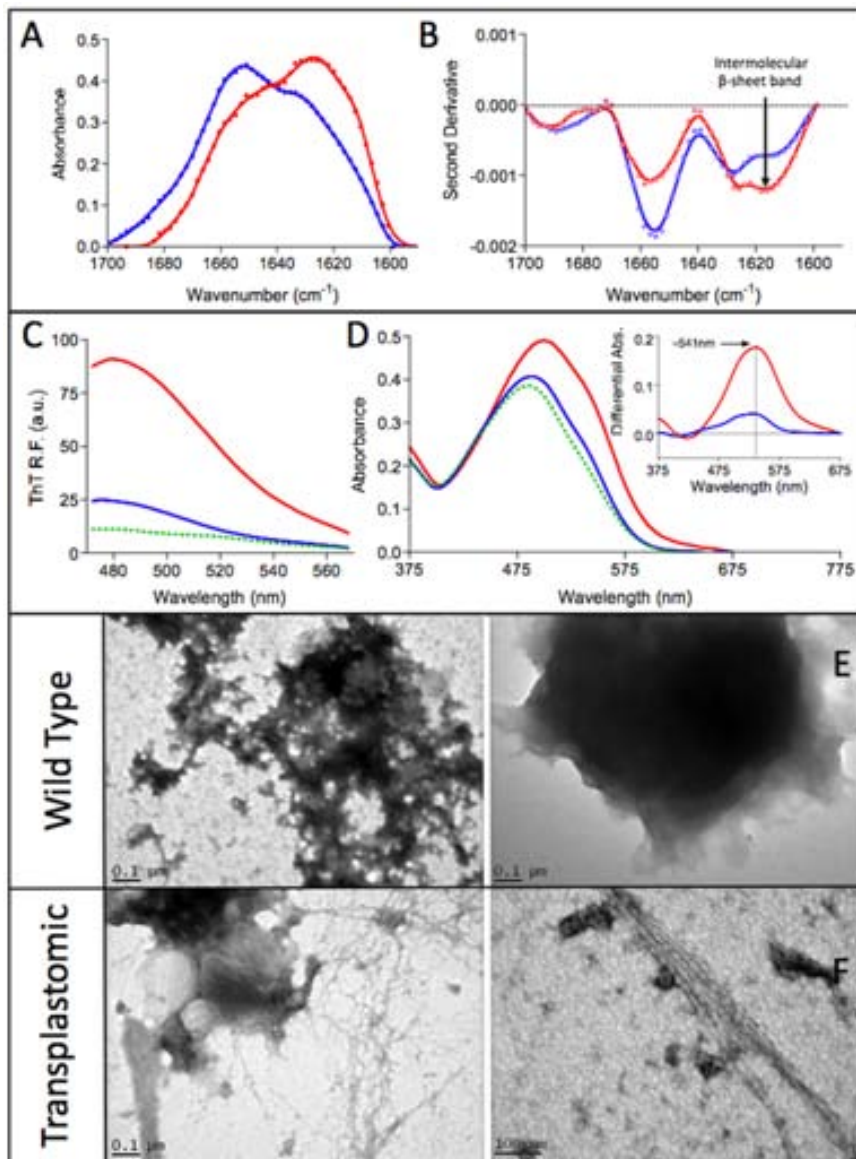


Figure 5. Amyloid-like properties of transplastomic tobacco protein aggregates. A) Secondary structure of transplastomic (red) and WT (blue) plants protein aggregates as measured by FTIR absorbance in the amide I region. B) Second derivative of the FTIR absorbance spectra shown in panel A. C) Fluorescence emission spectra of Th-T in the absence (green) and presence of transplastomic (red) and WT (blue) plants protein aggregates. The fluorescence intensity increases in the presence of protein aggregates. D) Absorption spectra of Congo red (CR) in the absence (green) and presence of transplastomic (red) and WT (blue) plants protein aggregates. Changes in λ_{max} and intensity in CR spectra are observed in the presence of protein aggregates, in the inset it is shown the difference absorbance spectra of Congo red in the presence of transplastomic (red) and WT (blue) plants protein aggregates, showing the characteristic amyloid band at ~ 541 nm. E and F) Morphology of transplastomic and WT plants protein aggregates as monitored by transmission electronic microscopy. doi:10.1371/journal.pone.0013625.g005

leaves suggests that the proteolytic cellular response targets misfolded TGZ species. However, the role of the different components of the protein quality control machinery in the formation of amyloid-like conformations should be addressed in depth before we can assemble a more precise picture of the process of protein aggregation in plant cells.

To the best of our knowledge, the present report describes for the first time the formation in plants of protein aggregates sharing structural features with the protein deposits associated to human disorders like Alzheimer's or Parkinson's diseases. The formation of amyloid-like aggregates has been recurrently observed in animal cells [1,29], described in fungi [9] and more recently in bacteria

[24,30]. Our data indicate that also in plants the accumulation of misfolded proteins after their synthesis at the ribosome might trigger their self-assembly into ordered β -sheet enriched macromolecular structures and therefore suggests that amyloid formation is an omnipresent process occurring across all the kingdoms of life.

The accumulation of TGZ in the chloroplasts has dramatic effects for the plant phenotype. In a first developmental phase, TGZ over-expression promotes an increase in the granum size (i.e. increase of the number of stacked thylakoids) with a concomitant decrease of stroma thylakoids and some impairment in the photochemistry of photosynthesis [19], finally the extended effect of this over-expression result in increased oxidative stress

symptoms and progressive cell degeneration [31]. Although initially this phenotype was attributed to an increase in the TGZ activity, a more generic toxic effect of the amyloid containing aggregates formed by TGZ on chloroplast membranes cannot be discarded. Accordingly, A β -peptide, α -synuclein, and prion proteins are known to promote the formation of distinct amyloid structures that destabilize and disrupt membrane structures [32] and cellular oxidative stress constitutes a common characteristic of amyloidogenic disorders including Alzheimer's disease [33]. Interestingly enough, the expression of A β -peptide in transgenic rice plants has been shown to promote significant endoplasmic reticulum (ER) stress suppressing the synthesis and targeting of secretory proteins including storage proteins during seed development, resulting in alternation of grain phenotype and changing the expression of genes and proteins [34]. Although the aggregated state of A β -peptide was not evaluated, the data in this study together with these in the present work suggest that the expression of aggregation-prone proteins in transgenic plants may have a generic detrimental effect in cell homeostasis. Despite no reports exists to date, in light of these data the existence of plant disorders associated to the aggregation of natural endogenous proteins cannot be completely discarded.

As in the case of bacteria, the formation of heterologous protein aggregates is a frequent observation during bioproduction in transgenic plants [35,36,37]. Importantly, the inclusion bodies formed in bacteria by proteins unrelated to any human disease have been shown to be toxic for mammalian cells [38], supporting the view that general mechanisms appear to underlie the cytotoxicity of many amyloid-like protein aggregates for mammalian cells independently of their source or sequence [39]. In addition, the uptake of heterologous amyloid-like material has been recently suggested to be a factor of risk for the initiation of human amyloid diseases [40]. The extent to which the presence of amyloid-like conformations is a common or anecdotic characteristic of protein deposits in genetically modified plants should be further studied for other polypeptides, tissues and species. In any case, the present data argue that the conformational properties of individual recombinant proteins produced in transgenic plant systems should be investigated in depth before its use for animal or human applications.

Materials and Methods

Protein expression in *Escherichia coli* and purification

E. coli strains DH5 α and BL21(DE3) were used as cloning hosts for construction of expression plasmids and for protein expression, respectively. For protein expression, transformed (pET28-TGZ) *E. coli* BL21 colonies were grown in LB medium containing 30 μ g kanamycin/ml to an OD₆₀₀ of 0.4 induced for 3 h with 0.4 mM IPTG, and finally harvested by centrifugation [18]. Intracellular recombinant proteins were released with CelLytic BII reagent (Sigma). TGZ was expressed as a histidine-tagged fusion and purified from the insoluble fraction under denaturing conditions (20 mM TRIS at pH 8, 0.5 M NaCl, 20 mM imidazole and 6 M GndHCl) by affinity chromatography on FF-Histrap histidine-tag resin (Amersham). Buffer was exchanged by gel filtration on Sephadex G-25 column (Amersham) with 50 mM TRIS at pH 7.5, 150 mM NaCl at 4°C.

In vitro protein aggregation

TGZ fibrils were assembled at 25°C at a final protein concentration of 20 μ M. TGZ aggregation was followed by measuring the transition from the soluble to the aggregated state by UV/Vis spectrophotometry at 350 nm. After a week, the

fibrillation reaction was considered to be finished since no further increase in the scattering signal was observed.

Plant transformation

To transfer the maize (*Zea mays*) *tgz* gene [17] to tobacco chloroplasts, the gene was PCR amplified, fused to the promoter and 5' untranslated region of the *psbA* gene and then introduced into the multiple cloning site of pAF [15], to give the final vector, pAF-*tgz13* [19]. Plant regenerants analysis was performed as previously described [19]. After transplanting homoplasmic *tgz*-transgenic plants were not able to set seed, and died, so were maintained *in vitro* [19].

Extraction of insoluble protein fraction from tobacco leaves

To extract the insoluble protein fraction from tobacco leaves, transformed and untransformed leaves from *in vitro*-grown plants were ground in liquid nitrogen and 100 mg resuspended in 7 vols. of protein extraction buffer consisting on 20 mM Tris-HCl, 150 mM NaCl, pH 7.5, plus protease inhibitors. The homogenate was filtered through two layers of cheesecloth and 3 μ L of DNase I and RNase from 1 mg/mL stock (25 μ g/mL final concentration) and 3 μ L of 1 M MgSO₄ were added and the resulting mixture was further incubated at 37°C for 30 min. Protein aggregates were separated by centrifugation at 10 000 \times g for 15 min at 4°C. Finally, the aggregates were washed with the same buffer containing 0.5% Triton X-100 and twice with sterile PBS. After a final centrifugation at 10 000 \times g for 15 min, pellets were stored at -20°C until analysis. The frozen pellets were reconstituted in 50 mM TRIS at pH 7.5, 150 mM NaCl.

TEM observations

Tobacco leaf thin sections (less than 0.5 mm) were fixed by vacuum infiltration with 2% paraformaldehyde and 2.5% glutaraldehyde in 0.1 M phosphate buffer pH 7.4. After washing, they were fixed in osmium tetroxide for 2 h in the same buffer, dehydrated through an acetone series and embedded in Spurr resin by infiltration. The blocks were polymerized for 48 h at 60°C. Ultrathin sections were obtained with an Ultracut UCT ultramicrotome (Leica) using a diamond knife, and mounted on gold grids (200 mesh). To immunolocalize subcellularly TGZ into tobacco chloroplasts, leaf slices were fixed with 4% paraformaldehyde and 0.5% glutaraldehyde in 0.1 M phosphate buffer (pH 7.4) for 2 h at 4°C. After washing, samples were dehydrated through an ethanol series and embedded in Lowicryl K4M resin (Pelco International, Redding, Calif., U.S.A.) at -35°C. Blocks were polymerized under a UV lamp at -20°C for 24 h and ultrathin sections were mounted on gold grids. The primary antibody AbTGZ4 [18] was used at 1:5.000 dilution and a solution of 12 nm diameter colloidal gold-affinipure anti-mouse IgG (Jackson ImmunoResearch) diluted at 1:30 in the blocking solution was used as the secondary antibody. Control samples were treated only with blocking solution or pre-immune serum following the same protocol. The sections were examined under a Jeol-JEM-1010 transmission electron microscope at 80 kV.

FT-IR spectroscopy

Attenuated total reflectance (ATR) FT-IR spectroscopy analyses of purified insoluble protein fractions of bacteria and plant leaves were performed using a Bruker Tensor 27 FT-IR Spectrometer (Bruker Optics Inc) with a Golden Gate MKII ATR accessory. Each spectrum consists of 20 independent scans, measured at a spectral resolution of 1 cm⁻¹ within the 1800–1500 cm⁻¹ range. All spectral data were acquired and normalized using the OPUS

MIR Tensor 27 software. Second derivatives of the spectra were used to determine the frequencies at which the different spectral components were located. Infrared spectra were fitted overlapping Gaussian curves and the amplitude, centre and bandwidth at half of the maximum amplitude and area of each Gaussian function were calculated using a non-linear peak fitting program (PeakFit package, Systat Software, San Jose, CA, USA).

Congo Red Assay

CR interaction with the purified protein fractions of bacteria and plant leaves was tested using a Cary100 (Varian Inc., Palo Alto, CA, USA) UV/Vis spectrophotometer by recording the absorbance spectra from 375 nm to 675 nm using a matched pair of quartz cuvettes of 1 cm optical length placed in a thermostated cell holder at 25°C. Final CR and protein concentrations were 10 µM and 0,04 mg/ml in BUFFER. Spectra were recorded after 2 min equilibration and solutions without protein and solutions without Congo red were used as negative controls. Binding of CR to a 10 µM amylin amyloid fibril solution was used as positive control. For optical microscopy analysis, proteins were incubated for 1 h in the presence of 50 µM CR. After centrifugation (14 000xg for 5 min), the precipitated fraction was placed on a microscope slide and sealed. The CR birefringence was detected under cross-polarized light using an optic microscope (Leica DMRB, Heidelberg, Germany).

Thioflavin-T Assay

Th-T binding to the purified protein fractions of bacteria and plant leaves was recorded using a Varian spectrofluorometer (Cary Eclipse) (Varian Inc., Palo Alto, CA, USA) with a excitation wavelength of 445 nm and emission range from 470 nm to 570 nm and the emission at 480 nm was recorded. Final Th-T and protein concentrations were 25 µM and 0,04 mg/ml in BUFFER, respectively. Spectra were recorded after 2 min equilibration and solutions without protein were used as negative controls. Binding of Th-T to a 10 µM amylin amyloid fibril solution was used as positive control. For microscopy analysis, proteins were incubated for 1 h in presence of 125 µM of Th-T. After centrifugation (14 000xg for 5 min), the precipitated fraction was placed on a microscope slide and sealed. Th-T relative fluorescence images of purified IBs were obtained at 40-fold magnification under UV light in a Leica fluorescence DMBR microscope (Leica Microsystems, Mannheim, Germany).

Electronic transmission microscopy

The purified protein fractions of bacteria and plant leaves were diluted ten fold in water, placed on carbon-coated copper grids and left to stand for five minutes; then, the grids were washed with distilled water and stained with 2% (w/v) uranyl acetate for another two minutes before analysis using a HitachiH-7000 transmission electron microscope operating at an accelerating voltage of 75 kV.

Prediction of aggregation-prone regions

Two different algorithms were used to predict the potential presence of aggregation-prone regions in the sequence of transglutaminase: TANGO (<http://tango.crg.es>), which is based on the physicochemical principles underlying β -sheet formation, extended by the assumption that the core regions of an aggregate are fully buried and AGGRESCAN (<http://bioinf.uab.es/aggrescan/>) which uses an aggregation-propensity scale for natural amino acids derived from in vivo experiments. The default prediction parameters were used for both programs.

Preparation and aggregation of a C-terminal peptide of TGZ

The peptide QLVVLDILLGKFS corresponding to residues 465–477 of maize TGZ was obtained from EZBiolab Inc. (Carmel, IN, USA) with a purity of 92.62%. Stock solutions were prepared at 5 mM in 1,1,1,3,3,3-hexafluoro-2-propanol (HFIP), then centrifuged at 15 000 g at 4°C for 15 min and finally were filtrated through millex-GV 0.22 µm filters in order to avoid the presence of pre-aggregated species in the assays. HFIP was removed by evaporation under a gentle stream of nitrogen and the samples were stored at –80°C until analysis. The peptide was resuspended in 1 mL of 50 mM TRIS at pH 7.5, 150 mM NaCl at 4°C at 100 µM and the samples bathsonicated for 5 min previous to aggregation assays. The peptide aggregation was carried out at 25°C without and with agitation (500 rpm) for 48 h and amyloid properties evaluated as described above.

Western blot analysis

Proteins of the soluble and insoluble fraction were quantified by the Bradford method using the Bio-Rad® reactive. Proteins were separated by SDS-PAGE in a Mini-Protean III system (Bio-Rad, Hercules, CA, USA) adding 3 mg protein per ml of buffer in the presence or absence of 8 M Urea. About 100 µg total protein was added per well. Separated proteins were further transferred to nitrocellulose membrane (GE Healthcare, Little Chalfont, UK) on wet system (Bio-Rad, Hercules, CA, USA) according to manufacturer's instructions. Membrane blocking was performed with non-fat dry milk (5%, w/v) in PBS 1× and washed in PBS with Tween 20 (0.1 to 0.3%, v/v). Immunodetection was carried out by using a polyclonal antibody raised against maize TGZ protein expressed and purified from *E. coli* [18] as primary antibody diluted 1/5000, and a peroxidase-conjugated goat anti-rabbit IgG (AO545, Sigma-Aldrich, Spain) at 1/15000 dilution as the secondary antibody. Immunodetection was obtained by chemiluminescence (ECL, Amersham Pharmacia Biotech®) following the manufacturer's instructions. Western-blot images were acquired in a LAS-3000 Fuji (Japan) Imaging System.

Supporting Information

Figure S1 Prediction of aggregation prone regions in maize TGZ sequence. A) Amino acid sequence of TGZ. The regions with the highest predicted aggregation propensities are shown in red. B) AGGRESCAN aggregation profile of the 150 C-terminal residues of TGZ (residues 1 and 151 in the profile correspond to residues 384 and 534 in TGZ, respectively). C) TANGO aggregation profile of the 150 C-terminal residues of TGZ (residues 1 and 151 in the profile correspond to residues 384 and 534 in TGZ, respectively).

Found at: doi:10.1371/journal.pone.0013625.s001 (0.20 MB PDF)

Acknowledgments

We thank N. Cortadellas, E. Fernandez and A. García (Serveis Científico-Tècnics, UB) for their technical assistance in the TEM images. Authors specially thank Drs. Veramendi and Ortigosa for the obtention of the tobacco-tgz transplastomic plants.

Author Contributions

Conceived and designed the experiments: SV. Performed the experiments: AVP OLJG. Analyzed the data: RS JMT MS SV. Wrote the paper: MS SV.

References

- Chiti F, Dobson CM (2006) Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem* 75: 333–366.
- Balch WE, Morimoto RI, Dillin A, Kelly JW (2008) Adapting proteostasis for disease intervention. *Science* 319: 916–919.
- Pepys MB (2006) Amyloidosis. *Annu Rev Med* 57: 223–241.
- Wickner S, Maurizi MR, Gottesman S (1999) Posttranslational quality control: folding, refolding, and degrading proteins. *Science* 286: 1888–1893.
- Nelson R, Eisenberg D (2006) Recent atomic models of amyloid fibril structure. *Curr Opin Struct Biol* 16: 260–265.
- Fernandez-Busquets X, de Groot NS, Fernandez D, Ventura S (2008) Recent structural and computational insights into conformational diseases. *Curr Med Chem* 15: 1336–1349.
- Dobson CM (2003) Protein folding and misfolding. *Nature* 426: 884–890.
- Ritter C, Maddelein ML, Siemer AB, Luhrs T, Ernst M, et al. (2005) Correlation of structural elements and infectivity of the HET-s prion. *Nature* 435: 844–848.
- Wickner RB, Edskes HK, Shewmaker F, Nakayashiki T (2007) Prions of fungi: inherited structures and biological roles. *Nat Rev Microbiol* 5: 611–618.
- Barnhart MM, Chapman MR (2006) Curli biogenesis and function. *Annu Rev Microbiol* 60: 131–147.
- Wang L, Maji SK, Sawaya MR, Eisenberg D, Riek R (2008) Bacterial inclusion bodies contain amyloid-like structure. *PLoS Biol* 6: e195.
- Morell M, Bravo R, Espargaro A, Sisquella X, Aviles FX, et al. (2008) Inclusion bodies: specificity in their aggregation process and amyloid-like structure. *Biochim Biophys Acta* 1783: 1815–1825.
- Oh J, Kim JG, Jeon E, Yoo CH, Moon JS, et al. (2007) Amyloidogenesis of type III-dependent harpins from plant pathogenic bacteria. *J Biol Chem* 282: 13601–13609.
- Boehm R (2007) Bioproduction of therapeutic proteins in the 21st century and the role of plants and plant cells as production platforms. *Annals of the New York Academy of Sciences* 1102: 121–134.
- Fernandez-San Millan A, Ortigosa SM, Hervás-Stubbs S, Corral-Martínez P, Seguí-Simarro JM, et al. (2008) Human papillomavirus L1 protein expressed in tobacco chloroplasts self-assembles into virus-like particles that are highly immunogenic. *Plant Biotechnology Journal* 6: 427–441.
- Lorand L (1996) Neurodegenerative diseases and transglutaminase. *Proceedings of the National Academy of Sciences of the United States of America* 93: 14310–14313.
- Villalobos E, Santos M, Talavera D, Rodríguez-Falcon M, Torne JM (2004) Molecular cloning and characterization of a maize transglutaminase complementary DNA. *Gene* 336: 93–104.
- Carvajal-Vallejos PK, Campos A, Fuentes-Prior P, Villalobos E, Almeida AM, et al. (2007) Purification and in vitro refolding of maize chloroplast transglutaminase over-expressed in *Escherichia coli*. *Biotechnology Letters* 29: 1255–1262.
- Ioannidis NE, Ortigosa SM, Veramendi J, Pinto-Marijuan M, Fleck I, et al. (2009) Remodeling of tobacco thylakoids by over-expression of maize plastidial transglutaminase. *Biochimica et Biophysica Acta* 1787: 1215–1222.
- Conchillo-Sole O, de Groot NS, Aviles FX, Vendrell J, Daura X, et al. (2007) AGGRESCAN: a server for the prediction and evaluation of “hot spots” of aggregation in polypeptides. *BMC Bioinformatics* 8: 65.
- Fernandez-Escamilla AM, Rousseau F, Schymkowitz J, Serrano L (2004) Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat Biotechnol* 22: 1302–1306.
- Ivanova MI, Sawaya MR, Gingery M, Attinger A, Eisenberg D (2004) An amyloid-forming segment of {beta}2-microglobulin suggests a molecular model for the fibril. *PNAS* 101: 10584–10589.
- Ventura S, Zurdo J, Narayanan S, Parreno M, Manges R, et al. (2004) Short amino acid stretches can mediate amyloid formation in globular proteins: the Src homology 3 (SH3) case. *Proc Natl Acad Sci U S A* 101: 7258–7263.
- de Groot NS, Sabate R, Ventura S (2009) Amyloids in bacterial inclusion bodies. *Trends Biochem Sci* 34: 408–416.
- Bravo R, Arimon M, Valle-Delgado JJ, Garcia R, Durany N, et al. (2008) Sulfated polysaccharides promote the assembly of amyloid beta(1-42) peptide into stable fibrils of reduced cytotoxicity. *Journal of Biological Chemistry* 283: 32471–32483.
- Vendruscolo M, Tartaglia GG (2008) Towards quantitative predictions in cell biology using chemical properties of proteins. *Mol Biosyst* 4: 1170–1175.
- Sabate R, de Groot NS, Ventura S (2010) Protein folding and aggregation in bacteria. *Cellular and Molecular Life Sciences* 67: 2695–2715.
- Kopito RR (2000) Aggresomes, inclusion bodies and protein aggregation. *Trends in Cell Biology* 10: 524–530.
- Kisilevsky R (2000) Amyloids: tombstones or triggers? *Nat Med* 6: 633–634.
- Wang L (2009) Towards revealing the structure of bacterial inclusion bodies. *Prion* 3: 139–145.
- Ortigosa S, Diaz-Vivancos P, Clemente-Moreno M, Pintó-Marijuan M, Fleck I, et al. (2010) Oxidative stress induced in tobacco leaves by chloroplast over-expression of maize plastidial transglutaminase. *Planta* In press.
- Meredith SC (2005) Protein denaturation and aggregation: Cellular responses to denatured and aggregated proteins. *Ann N Y Acad Sci* 1066: 181–221.
- Querfurth HW, LaFerla FM Alzheimer's disease. *New England Journal of Medicine* 362: 329–344.
- Oono Y, Wakasa Y, Hirose S, Yang L, Sakuta C, et al. (2010) Analysis of ER stress in developing rice endosperm accumulating beta-amyloid peptide. *Plant Biotechnology Journal* 8: 691–718.
- Takaiwa F, Hirose S, Takagi H, Yang L, Wakasa Y (2009) Deposition of a recombinant peptide in ER-derived protein bodies by retention with cysteine-rich prolamins in transgenic rice seed. *Planta* 229: 1147–1158.
- Castelli S, Vitale A (2005) The phaseolin vacuolar sorting signal promotes transient, strong membrane association and aggregation of the bean storage protein in transgenic tobacco. *J Exp Bot* 56: 1379–1387.
- Saito Y, Kishida K, Takata K, Takahashi H, Shimada T, et al. (2009) A green fluorescent protein fused to rice prolamin forms protein body-like structures in transgenic rice. *J Exp Bot* 60: 615–627.
- Gonzalez-Montalban N, Villaverde A, Aris A (2007) Amyloid-linked cellular toxicity triggered by bacterial inclusion bodies. *Biochem Biophys Res Commun* 355: 637–642.
- Bucciantini M, Giannoni E, Chiti F, Baroni F, Formigli L, et al. (2002) Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases. *Nature* 416: 507–511.
- Westermarck P, Lundmark K, Westermarck GT (2009) Fibrils from designed non-amyloid-related synthetic peptides induce AA-amyloidosis during inflammation in an animal model. *PLoS ONE* 4: e6041.

3.5 Publication V

Modeling amyloids in bacteria

Anna Villar-Piqué and Salvador Ventura

Microb. Cell Fact. 11: 166-168.
28 December 2012

COMMENTARY

Open Access

Modeling amyloids in bacteria

Anna Villar-Piqué¹ and Salvador Ventura^{1,2*}

Abstract

An increasing number of proteins are being shown to assemble into amyloid structures, self-seeding fibrillar aggregates that may lead to pathological states or play essential biological functions in organisms. Bacterial cell factories have raised as privileged model systems to understand the mechanisms behind amyloid assembly and the cellular fitness cost associated to the formation of these aggregates. In the near future, these bacterial systems will allow implementing high-throughput screening approaches to identify effective modulators of amyloid aggregation.

The aggregation of proteins into amyloid structures is the triggering event on the onset of a growing number of human disorders, from neurodegenerative diseases as Alzheimer, Parkinson, Huntington or transmissible spongiform encephalopathies to non-neurodegenerative systemic and localized amyloidosis, as senile systemic amyloidosis or type II diabetes [1]. In addition, it is now clear that different organisms, from virus to humans, exploit the special architecture of amyloids for functional purposes, such as cellular invasion or hormone storage [2]. Therefore, protein aggregation has emerged from a neglected area of protein science to a central issue in biology and biomedicine.

Many biochemical pathways, from DNA replication to protein degradation, have been modeled first in bacteria. However, despite it has been long recognized that heterologous protein expression in bacterial cell factories results often in the formation of insoluble deposits composed essentially by the target protein [3,4], only recently some groups have dared to exploit this well-characterized phenomena to model amyloid formation [5,6]. This delay resulted mainly from the fact that these aggregates, known as inclusion bodies (IBs), were traditionally considered unstructured protein particles only useful to obtain denatured protein for refolding purposes. This old framework has changed into a new scenario where intracellular aggregation in bacteria is providing important clues on the molecular determinants of amyloid formation and its remediation [7,8].

A first inflection in the field came along with early reports describing a selective molecular structure inside IBs. It has been found a substantial similitude between the properties of these bacterial aggregates and the pathological fibrils linked to amyloidosis. On one side, IBs generally bind thioflavin T and Congo Red, the typical amyloid dyes, and display seeding ability in the fibrillar assembly of homologous monomers [9,10]. On the other side, low resolution techniques, such as infrared spectroscopy, circular dichroism or X-ray diffraction, denote the presence of signals corresponding to tightly packed intermolecular β -sheets, similar to those in amyloid fibrils [9-11]. Importantly, these findings come from independent studies using completely different protein models, not related in sequence or structure, thus suggesting that the amyloid signature might be a generic feature of bacterial aggregates. Furthermore, high resolution approaches, such as hydrogen/deuterium exchange by NMR or solid-state NMR have been used to study different bacterial IBs, defining their molecular structure at the residue level. These analysis prove that, at least for amyloidogenic proteins, bacterial IBs and fibrils share the same amyloid core. However, they also show that part of the polypeptide sequence or, alternatively, a fraction of the molecules remain disordered and/or in native-like conformations inside these aggregates [11,12]. Contrary to the previous assumption that IBs were totally inactive, the presence of native-like structure endorse IBs with a certain degree of biological activity [13-17]. This observation has opened the door to the use of bacteria as small factories to produce promising functional materials and catalysts, boosting the investigation of the structural and functional properties of IBs [18-23].

* Correspondence: salvador.ventura@uab.es

¹Departament de Bioquímica i Biologia Molecular, Facultat de Biociències, Universitat Autònoma de Barcelona, Bellaterra E-08193, Spain

²Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, Bellaterra E-08193, Spain

It is now clear that they are the oligomeric assemblies populating the fibrillation pathway and not the mature fibrils that exert the main cytotoxic effect in conformational disorders [2]. The structures of these oligomeric intermediates states have been the subject of debate for many years. The similarity between IBs and amyloids has arisen a critical question: do IBs exert a cytotoxic effect analogous to that recurrently observed for fibrils? The answer is: Yes, the addition of purified bacterial IBs to neuronal cultured cells produces a loss in cell viability equal to that promoted by the same concentration of amyloid material [24,25]. Bacterial IBs contain small heat shock proteins (sHSPs), which are highly homologous to those found in the aggregates of the brains of patients suffering different neuronal pathologies. It has been proposed that in the brain these sHSPs might break down amyloid fibril structure, resulting in the accumulation of toxic oligomeric species. The observation that the neurotoxicity of IBs correlates with the amount of oligomeric assemblies and chaperones in these aggregates and the possibility to identify at the residue level the determinants of this effect [25] are expected to provide new molecular insights on the structures of the deleterious species in amyloid assemblies. Thus, protein overexpression in bacterial cell factories, by mimicking the conditions in the cell under stress, will likely allow to address aspects of amyloid biology that are otherwise technically impossible to study in more complex contexts.

Prion proteins are a particularly interesting and dangerous type of amyloids, since their aggregated states have self-perpetuating ability and thus become infectious. Het-s, from the fungus *Podospora anserina*, was the first prion protein whose bacterial IBs were shown to display amyloid-like properties [12,26]. The differential trait of these aggregates emerged when they were transfected into prion-free fungal strains, as they promoted prionic conversion at levels comparable to those induced by homologous amyloid fibrils [12]. This result has been later corroborated in the case of the yeast prion Sup35. The IBs of this protein have been used to induce the prion phenotype in prion-free yeast strains, with the novel evidence that the infectivity rate can be easily modulated by tuning the environmental conditions during the formation of IBs [27]. When instead of being expressed intracellularly, this protein is directed to the secretory pathway, the aggregates are formed in the cell surface of bacteria, but they are also able to template the conformational prionic change [28]. These recent observations provide perhaps the best confirmation that the IBs molecular structure highly resembles to the fine architecture of amyloid fibrils, in such a way that even the infectious properties of

amyloids, which depend on very specific conformational properties, are conserved in the two type of aggregates. This evidence bears important implications for the use of bacteria to model amyloids, since prion-like behaviour is currently receiving preferential attention in the field, due to the growing realisation that protein-based infection may be behind frequently occurring neurodegenerative disorders such Alzheimer's and Parkinson's diseases [29].

The increasing medical and economic impact of aggregation-linked diseases in our society has fueled the development of methods to identify chemical compounds that can interfere with amyloidogenic pathways, having thus therapeutic potential to treat or prevent these disorders. Generally, these assays, used by many biopharma companies, are cumbersome, lack reproducibility, use expensive synthetic peptides and are performed in physiologically non-relevant contexts. Several labs are focusing their efforts towards bacterial systems to overcome these limitations. In this context, fluorescent tag reporters of aggregation have been employed in bacteria to measure in a straightforward manner the amyloid assembly rate, as the final fluorescence of the aggregate is the result of a kinetic competition between folding and aggregation [15,30,31]. Any compound that enhances or inhibits one of these two competing reactions can be easily detected by spectrofluorometry. This property has been exploited recently both in living bacteria and *in vitro*, using purified IBs, to implement high-throughput, 96-well-plate based, assays able to identify and characterize novel amyloid modulators in large compound libraries [32,33]. As stated above, the amyloid nature of bacterial aggregates can be assessed using dyes such as Thioflavin-S (Th-S), whose spectroscopic properties change upon binding to amyloid structures. This characteristic, together with the ability of the dye to enter intact cells can be used to detect *in vivo* the formation of amyloid structures inside bacteria. The application of flow cytometry to detect (Th-S) fluorescence has been shown to be a fast, robust, quantitative, non-invasive method to screen for the presence of *in vivo* intracellular amyloid-like aggregates in bacteria as well as for monitoring the effect of amyloid inhibitors in intact cells, skipping the need for a genetically encoded reporter [34]. Although still in an early stage, it is clear that, apart from its academic interest, modeling amyloid formation in bacteria might render important economic revenues. The above examples illustrate how bacterial cell factories can be easily adapted to develop screening tools for amyloid aggregation inhibitors that will outperform the conventional screening procedures used by the industry.

Acknowledgements

Work in our lab is supported by grants BFU2010-14901 from Ministerio de Ciencia e Innovación (Spain) and 2009-SGR 760 from AGAUR (Generalitat de Catalunya). SV has been granted an ICREA ACADEMIA award (ICREA).

Received: 23 December 2012 Accepted: 23 December 2012

Published: 28 December 2012

References

1. Fernandez-Busquets X, de Groot NS, Fernandez D, Ventura S: **Recent structural and computational insights into conformational diseases.** *Curr Med Chem* 2008, **15**:1336–1349.
2. Invernizzi G, Papaleo E, Sabate R, Ventura S: **Protein aggregation: mechanisms and functional consequences.** *Int J Biochem Cell Biol* 2012, **44**:1541–1554.
3. Sorensen HP, Mortensen KK: **Soluble expression of recombinant proteins in the cytoplasm of Escherichia coli.** *Microb Cell Fact* 2005, **4**:1.
4. Jurgen B, Breitenstein A, Urlacher V, Buttner K, Lin H, Hecker M, Schweder T, Neubauer P: **Quality control of inclusion bodies in Escherichia coli.** *Microb Cell Fact* 2010, **9**:41.
5. Ventura S: **Sequence determinants of protein aggregation: tools to increase protein solubility.** *Microb Cell Fact* 2005, **4**:11.
6. Ventura S, Villaverde A: **Protein quality in bacterial inclusion bodies.** *Trends Biotechnol* 2006, **24**:179–185.
7. de Groot NS, Sabate R, Ventura S: **Amyloids in bacterial inclusion bodies.** *Trends Biochem Sci* 2009, **34**:408–416.
8. Garcia-Fruitos E, Sabate R, de Groot NS, Villaverde A, Ventura S: **Biological role of bacterial inclusion bodies: a model for amyloid aggregation.** *FEBS J* 2011, **278**:2419–2427.
9. Carrio M, Gonzalez-Montalban N, Vera A, Villaverde A, Ventura S: **Amyloid-like properties of bacterial inclusion bodies.** *J Mol Biol* 2005, **347**:1025–1037.
10. Morell M, Bravo R, Espargaro A, Sisquella X, Aviles FX, Fernandez-Busquets X, Ventura S: **Inclusion bodies: specificity in their aggregation process and amyloid-like structure.** *Biochim Biophys Acta* 2008, **1783**:1815–1825.
11. Wang L, Maji SK, Sawaya MR, Eisenberg D, Riek R: **Bacterial inclusion bodies contain amyloid-like structure.** *PLoS Biol* 2008, **6**:e195.
12. Wasmer C, Benkemoun L, Sabate R, Steinmetz MO, Couly-Salin B, Wang L, Riek R, Sauppe SJ, Meier BH: **Solid-state NMR spectroscopy reveals that E. coli inclusion bodies of HET-s(218–289) are amyloids.** *Angew Chem Int Ed Engl* 2009, **48**:4858–4860.
13. Garcia-Fruitos E, Gonzalez-Montalban N, Morell M, Vera A, Ferraz RM, Aris A, Ventura S, Villaverde A: **Aggregation as bacterial inclusion bodies does not imply inactivation of enzymes and fluorescent proteins.** *Microb Cell Fact* 2005, **4**:27.
14. Garcia-Fruitos E, Aris A, Villaverde A: **Localization of functional polypeptides in bacterial inclusion bodies.** *Appl Environ Microbiol* 2007, **73**:289–294.
15. de Groot NS, Ventura S: **Protein activity in bacterial inclusion bodies correlates with predicted aggregation rates.** *J Biotechnol* 2006, **125**:110–113.
16. Peternel S, Grdadolnik J, Gaberc-Porekar V, Komel R: **Engineering inclusion bodies for non denaturing extraction of functional proteins.** *Microb Cell Fact* 2008, **7**:34.
17. Wu W, Xing L, Zhou B, Lin Z: **Active protein aggregates induced by terminally attached self-assembling peptide ELK16 in Escherichia coli.** *Microb Cell Fact* 2011, **10**:9.
18. Garcia-Fruitos E: **Inclusion bodies: a new concept.** *Microb Cell Fact* 2010, **9**:80.
19. Garcia-Fruitos E, Vazquez E, Diez-Gil C, Corchero JL, Seras-Franzoso J, Ratera I, Veciana J, Villaverde A: **Bacterial inclusion bodies: making gold from waste.** *Trends Biotechnol* 2012, **30**:65–70.
20. Peternel S, Komel R: **Isolation of biologically active nanomaterial (inclusion bodies) from bacterial cells.** *Microb Cell Fact* 2010, **9**:66.
21. Rodriguez-Carmona E, Cano-Garrido O, Seras-Franzoso J, Villaverde A, Garcia-Fruitos E: **Isolation of cell-free bacterial inclusion bodies.** *Microb Cell Fact* 2010, **9**:71.
22. Vazquez E, Corchero JL, Burgueno JF, Seras-Franzoso J, Kosoy A, Bosser R, Mendoza R, Martinez-Lainez JM, Rinas U, Fernandez E, et al: **Functional inclusion bodies produced in bacteria as naturally occurring nanopills for advanced cell therapies.** *Adv Mater* 2012, **24**:1742–1747.
23. Villaverde A, Garcia-Fruitos E, Rinas U, Seras-Franzoso J, Kosoy A, Corchero JL, Vazquez E: **Packaging protein drugs as bacterial inclusion bodies for therapeutic applications.** *Microb Cell Fact* 2012, **11**:76.
24. Gonzalez-Montalban N, Villaverde A, Aris A: **Amyloid-linked cellular toxicity triggered by bacterial inclusion bodies.** *Biochem Biophys Res Commun* 2007, **355**:637–642.
25. Dasari M, Espargaro A, Sabate R, del Amo JM L, Fink U, Grelle G, Bieschke J, Ventura S, Reif B: **Bacterial inclusion bodies of Alzheimer's disease beta-amyloid peptides can be employed to study native-like aggregation intermediate states.** *ChemBiochem* 2011, **12**:407–423.
26. Sabate R, Espargaro A, Sauppe SJ, Ventura S: **Characterization of the amyloid bacterial inclusion bodies of the HET-s fungal prion.** *Microb Cell Fact* 2009, **8**:56.
27. Espargaro A, Villar-Pique A, Sabate R, Ventura S: **Yeast prions form infectious amyloid inclusion bodies in bacteria.** *Microb Cell Fact* 2012, **11**:89.
28. Sivanathan V, Hochschild A: **Generating extracellular amyloid aggregates using E. coli cells.** *Genes Dev* 2012, **26**:2659–2667.
29. Aguzzi A, Calella AM: **Prions: protein aggregation and infectious diseases.** *Physiol Rev* 2009, **89**:1105–1152.
30. Waldo GS, Standish BM, Berendzen J, Terwilliger TC: **Rapid protein-folding assay using green fluorescent protein.** *Nat Biotechnol* 1999, **17**:691–695.
31. Villar-Pique A, de Groot NS, Sabate R, Acebron SP, Celaya G, Fernandez-Busquets X, Muga A, Ventura S: **The effect of amyloidogenic peptides on bacterial aging correlates with their intrinsic aggregation propensity.** *J Mol Biol* 2012, **421**:270–281.
32. Kim W, Kim Y, Min J, Kim DJ, Chang YT, Hecht MH: **A high-throughput screen for compounds that inhibit aggregation of the Alzheimer's peptide.** *ACS Chem Biol* 2006, **1**:461–469.
33. Villar-Pique A, Espargaro A, Sabate R, de Groot NS, Ventura S: **Using bacterial inclusion bodies to screen for amyloid aggregation inhibitors.** *Microb Cell Fact* 2012, **11**:55.
34. Espargaro A, Sabate R, Ventura S: **Thioflavin-S staining coupled to flow cytometry. A screening tool to detect in vivo protein aggregation.** *Mol Biosyst* 2012, **8**:2839–2844.

doi:10.1186/1475-2859-11-166

Cite this article as: Villar-Piqué and Ventura: Modeling amyloids in bacteria. *Microbial Cell Factories* 2012 **11**:166.

Submit your next manuscript to BioMed Central
and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit



3.6 Publication VI

Inclusion bodies in the study of amyloid aggregation

Anna Villar-Piqué and Salvador Ventura

Book chapter included in Protein Aggregation in Bacteria: Functional and Structural Properties of Inclusion Bodies in Bacterial Cells.

Editor: M. Lotti.

John Wiley and Sons, Inc. Hoboken, New Jersey, USA.

In press

Inclusion bodies in the study of amyloid aggregation

Anna Villar-Piqué¹ and Salvador Ventura^{1,2}§

1 Departament de Bioquímica i Biologia Molecular, Facultat de Biociències, Universitat Autònoma de Barcelona, E-08193 Bellaterra (Spain).

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

§Corresponding author

Email addresses:

AVP: anna.villar@uab.cat

SV: salvador.ventura@uab.es

Abstract

Protein misfolding and aggregation into amyloid fibrils are the underlying causes of a growing number of highly debilitating human diseases. The failure to attain a native structure is also the main responsible for the aggregation of eukaryotic proteins in bacterial cells as inclusion bodies during biotechnological production. We discuss here how the recognition of the amyloid-like architecture of inclusion bodies and the use of bacterial cells to model intracellular protein aggregation is providing dramatic insights on the mechanism of protein self-assembly into toxic and/or infective amyloid conformers in cellular environments and opening new ways to identify therapeutic compounds for amyloid-linked disorders.

Keywords: amyloid fibrils, bacterial models, conformational disorders, inclusion bodies, protein aggregation.

1. Introduction

After their synthesis at the ribosomes, most proteins need to attain a globular, compact and specific three-dimensional conformation in a biologically relevant time frame in order to perform their physiological functions, since only properly folded protein structures can interact selectively with their molecular targets (Daggett and Fersht 2009). Misfolded species or folding intermediates often display an intrinsic tendency to interact aberrantly, which results in their self assembly into insoluble aggregates that in many cases become toxic and interfere with important cellular pathways. Accordingly, protein misfolding and deposition into human tissues is linked to the pathology of an increasing number of disorders, from neurodegenerative diseases such as Alzheimer, Parkinson or the transmissible spongiform encephalopathies to non-neurodegenerative systemic and localized amyloidosis, as senile systemic amyloidosis, type II diabetes or even some types of cancer (Chiti and Dobson 2006). Irrespective of the nature of the proteins involved in these disorders, the aggregates usually adopt the form of amyloid fibrils, all sharing a common cross- β sheet structure in which the β -strands are oriented perpendicularly to the axis of the fibril (Fernandez-Busquets, de Groot et al. 2008). Importantly, the formation of amyloid-like structures is not restricted to a subset of pathological proteins but rather it appears to be a generic property of polypeptides across all kingdoms of life, including bacteria (Dobson 2003; Sabate, de Groot et al. 2010).

Protein aggregation is also a major bottleneck during the biotechnological production and purification of proteins, hindering the commercialization of protein-based drugs (Ventura and Villaverde 2006). Despite it has been recurrently observed that recombinant protein expression in bacterial hosts often leads to the aggregation of the target protein into compact aggregated structures known as inclusion bodies (IBs), only

recently this phenomena has been shown to share striking similarities with the process of amyloid fibril formation by pathogenic proteins. As we discuss in the present chapter, this evidence is paving the way to use bacteria as fascinating model systems to study the protein interactions leading to the formation of toxic and infective amyloids as well as their use to screen for molecules able to interfere these contacts.

2. Structure of IBs

The IBs formed during the recombinant expression of proteins unrelated in sequence, secondary structure content, three-dimensional conformation, size or origin all appear, upon isolation, as refractile, porous, hydrated and very dense particles. They normally have a spherical-like shape (in some cases appearing as short cylinders and paracrystalline forms) with a diameter range between 0.2 and 1.2 μm , probably determined by the shape and size of the bacterial cell, the protein intrinsic characteristics and the micro-environmental conditions during the expression process (Wang 2009), but overall they display poor morphological variety. It has been shown that certain genes, like the protease clpP, play an active role on the control of IBs shape. Accordingly, clpP deletion results in unusual tear-shaped particles displaying enhanced surface–volume ratios (Garcia-Fruitos, Seras-Franzoso et al. 2010). IBs are located either in the periplasmic space, when a signal peptide is fused to the target protein (Arie, Miot et al. 2006), or in the cytoplasm of the expression hosts. In this last case, the aggregates are usually localized at the cell poles (Rokney, Shagan et al. 2009). DAPI staining of bacterial DNA has revealed that IBs are in fact interspersed with the bacterial nucleoid (Villar-Pique, de Groot et al. 2012). Under electronic microscopy, IBs seem to have an amorphous electro-dense structure without any visible element of regularity (Carrio and Villaverde

2005). This visual information has sustained the traditional idea that IBs are protein deposits lacking of any other functionality than acting as “dust balls” to harvest misfolded polypeptides and avoid their interaction with other cellular components. Thus, for a long time IBs have been considered as unstructured aggregates and remained poorly characterized. Nevertheless, recently, this classical vision has changed and accumulating evidences of an organized molecular structure allow describing the IBs formed by many different polypeptides as amyloid-like aggregates (de Groot, Sabate et al. 2009) (Garcia-Fruitos, Sabate et al. 2011).

2.1. Amyloid-like nature of IBs

In the last years, the existence of amyloid-like features in bacterial aggregates has been reported by several independent groups (Wang, Maji et al. 2008) (Carrio, Gonzalez-Montalban et al. 2005) (Ono, Kubota et al. 2006). Our group, using an aggregation-prone version of β -galactosidase, whose overexpression in *E. coli* resulted in both soluble and aggregated species, was the first one to suggest this possibility. We showed that the aggregated material was able to seed specifically the deposition of the monomeric and soluble species, which become increasingly incorporated in the aggregated seeds, showing that protein deposition in bacteria is a selective process and that specific interactions rather than the previously proposed unspecific contacts should be at the basis of IBs formation. To explore the degree of molecular organization of these IBs, we used infrared spectroscopy with a resulting spectrum indicative of extended intermolecular β -sheet structures. Furthermore, specific amyloid dyes were used (Thioflavin-T and Congo Red) and the typical increase in the emission signal for amyloid structures was detected in those IBs, as well. Importantly, the IBs formed by different and unrelated proteins

shared similar conformational properties; therefore suggesting for the first time that the presence of amyloid-like features could be a common characteristic of bacterial protein deposits (Carrio, Gonzalez-Montalban et al. 2005). Afterwards, the Riek group provided a conclusive demonstration for the generic presence of amyloid-like conformations in IBs by characterizing the structural properties of the bacterial aggregates formed by three proteins representative of the different native folds: all α -helix, α -helix/ β -sheet and all β -sheet. Using alternative techniques, such as X-ray diffraction, circular dichroism or NMR, they could demonstrate the presence of specific, well protected, β -sheets in the aggregates surrounded by less structured protein regions (Wang, Maji et al. 2008).

The recognition of the amyloid-like nature of IBs by the scientific community was hindered by their apparent amorphous appearance and the lack of any fibrillar structure. It was assumed that this was due to the inner localization of the amyloid cross- β sheets in the core of the aggregate, while the rest of the protein remained unstructured or even displayed a native-like conformation at the surface of the IB. Fortunately, proteolytic treatment with proteinase K has permitted to visualize the existence of fibrillar regions of IBs with morphology and dimensions compatibles with those of the amyloid fibrils formed by the same protein. This enzyme displays a high proteolytic activity in front of globular or disordered conformations but is poorly active against densely packed cross- β regions (Wilson, Mok et al. 2007). Hence, the addition of proteinase K to purified IBs has become a useful technique to visualize fibrillar structure inside bacterial aggregates, either by electronic (Wang, Maji et al. 2008) or atomic force microscopy (Morell, Bravo et al. 2008), representing the ultimate and unequivocally proof of amyloid-like content in IBs. Therefore, nowadays it seems clear that protein overexpression in bacteria results in the formation of non-native intermolecular contacts that glue the polypeptide chains to form IBs. The amyloid-like nature of IBs relies essentially on the fact that, as it happens

during the formation of fibrillar assemblies, the construction of these bacterial aggregates depends on the establishment of interactions occurring through homologous protein patches in a nucleation-dependent manner. (Garcia-Fruitos, Sabate et al. 2011). Hence, IBs are mainly composed by the recombinant protein although other proteins can also be embedded in them, such as molecular chaperones (Carrio and Villaverde 2005). The conformational quality of IBs varies depending on the protein expressed, as its intrinsic aggregation propensity determines both the assembly rate and the conformational and stability features of the aggregate (de Groot and Ventura 2006). In this way, despite the main secondary structure element in many IBs are non-native β -sheets, which form the core of the aggregate and likely act sequestering misfolded species or slow-folding intermediates into the IBs, it has been reported that these regions might coexist with disordered fragments and native-like conformations. This last observation has a prominent importance in the field of biotechnology as the range of possible utilities of IBs containing active proteins expands from their use as biocatalysts to the *in vivo* engineering of inert nanostructured materials (Garcia-Fruitos, Vazquez et al. 2012).

2.2. Detection and characterization of amyloid conformations inside IBs

The list of techniques available to characterize the amyloid-like nature of IBs is in continuous expansion. Only the most widely used are briefly addressed in this section and summarized in Figure 1. Moreover, in Table 1, we list the different approaches and protein models for which positive results have been reported. On one hand, the typical amyloid dyes constitute a fast method to report amyloid-like properties in bacterial aggregates. Among them, Thioflavin-T (Th-T), Thioflavin-S (Th-S) and Congo Red are the most employed. Upon binding to amyloid-like structures, they undergo spectroscopic

changes induced by the presence of the highly repetitive and regular cross- β sheet motif. These dyes can be used in purified IBs (Carrio, Gonzalez-Montalban et al. 2005) or cell extracts (Ono, Kubota et al. 2006). On the other hand, low-resolution biophysical approaches have been employed in numerous studies to confirm the amyloid content of IBs. One of the first structural techniques applied to IBs was infrared spectroscopy. The resultant spectra in the amide I region is usually dominated by a main band around 1621-1628 cm^{-1} indicative of a prominent intermolecular β -sheet content (Doglia, Ami et al. 2008) (Wang 2009). Besides, the use of circular dichroism spectroscopy in the far-UV region is very useful to confirm the presence of this secondary structure element, detected as a minimum around 217 nm, coexisting with signals correspondent to unstructured conformations (Wang, Maji et al. 2008). Perhaps, X-ray diffraction constitutes the most robust technique to unequivocally demonstrate the presence of β -sheet arranged in an amyloid-like fashion in IBs. The diffraction patterns of different IBs have been shown to consist of two rings, typical for the cross- β structure in amyloid fibrils, with a major reflection at 4.7 Å corresponding to the spacing between the strands in the β -sheet and a diffused reflection at ~ 10 Å interpreted as the distance between adjacent β -sheets. The two reflections are circular and not orthogonal as in the cross- β structure in highly ordered amyloids, indicating a lower alignment of the cross- β components in IBs, which is consistent with the presence of alternative conformations in these aggregates. More recent studies have also applied high-resolution techniques to accurately map the structured core of bacterial aggregates. Some examples are ^{13}C - ^{13}C solid-state NMR used to define the structure of HET-s prion IBs (Wasmer, Benkemoun et al. 2009) or the hydrogen/deuterium exchange based solution-state NMR, which permits to identify the solvent-protected backbone amide protons involved in intermolecular hydrogen bonds (Dasari, Espargaro et al. 2011) (Wang, Maji et al. 2008).

Finally, the kinetic and thermodynamic stabilities of IBs can be analysed adapting available physico-chemical methods aimed to study protein folding and stability. Among them, IBs denaturation experiments with urea (Upadhyay, Murmu et al. 2012) or guanidine hydrochloride (Espargaro, Sabate et al. 2008), whose solubilisation curves reflect their resistance to chemical unfolding, provide insights on the cooperativity of the interactions sustaining the aggregate structure. Proteinase K digestion curves are also useful to measure the aggregates compactness and stability (Upadhyay, Murmu et al. 2012); the kinetic solubilisation constants derived from proteolytic experiments usually correlate with the stabilities calculated from chemical denaturation assays (de Groot and Ventura 2006). Finally, atomic force microscopy tip indentation is a methodology that evaluates the mechanical properties of IBs, and allows to identify the presence and quantify the stability of the amyloid core, as the fibrillar material inside them is the main responsible of the resistance exerted by the IBs in front of this rupture force (Villar-Pique, de Groot et al. 2012).

3. Formation of IBs

Protein expression in bacteria is a fine tuned process where subtle imbalances in protein homeostasis can trigger massive protein misfolding and aggregation. Although living cells are endowed with powerful protein quality machinery composed mainly by molecular chaperones and proteases, protein deposition is a common event during heterologous overexpression, especially when eukaryotic proteins are produced in microbial cell factories. This effect can be attributed to two main factors. First, the absence of adequate protein posttranslational modifications, like glycosilations, in the bacterial cytosol. Importantly, the discovery that bacteria possess, in fact, both N-linked and O-linked

glycosylation pathways displaying certain similarity to the eukaryotic counterparts is opening new opportunities to manipulate these pathways in order to engineer proteins with improved solubility features (Nothaft and Szymanski 2010). Second, the fast and high transcription levels attained during recombinant protein expression. On one hand, the accumulation of folding intermediates saturates the protein quality machinery. On the other hand, it is now clear that it exists a correlation between the abundance of a protein in its natural environment and its intrinsic aggregation propensity, in such a way that, because protein aggregation is a second or higher order reaction and therefore strongly dependent on protein concentrations, abundant proteins are endorsed with higher intrinsic solubility than poorly expressed ones, which can support a higher intrinsic aggregation tendency (Tartaglia and Vendruscolo 2009; de Groot and Ventura 2010; Castillo, Grana-Montes et al. 2011). Protein production tends to maximize the expression levels, bypassing this evolutive constraint, resulting in most cases in proteins being expressed at much higher concentrations than the physiological ones. This imbalance leads indefectibly to protein aggregation.

3.1. *In vivo* formation kinetics

During recombinant protein expression, the formation of IBs starts when a threshold in the levels of misfolded or unfolded species in the cytoplasm of the bacterial cell is exceeded (Ignatova and Gierasch 2004). This process has been suggested to occur in two steps. First, misfolded or folding intermediate species aggregate in multiple small foci randomly localized in the cytoplasm and afterwards, they are selectively directed to the poles of the cell, where they are fused to form a large aggregate tethered to the cell pole. It has been suggested that both processes occur in an energy-dependent manner

(Rokney, Shagan et al. 2009), the last step requiring the proton motive force, and DnaK and DnaJ chaperones. Nonetheless, this last point is controversial, as another work indicates that the polar localization of intracellular IBs is circumstantial, in such a way that aggregation occurs in DNA-free space and therefore that nucleoid occlusion is necessary and sufficient for the polar localization of IBs, suggesting thus that the deposition of protein aggregates at the poles of *E. coli* occurs through a passive mechanism (Winkler, Seybert et al. 2010). Despite these discrepancies, what it seems clear is that a bacillar bacterial cell contains one or two IBs, at the most, which are distributed between the two daughter cells after division. Actually, the presence of more than two IBs in a cell is indicative of a reproductive defect and the impairment of the binary fission (Lindner, Madden et al. 2008).

During *in vivo* IB assembly, it takes place a kinetic competition between protein folding and aggregation, which reflects a balance between the formation of intramolecular contacts leading to the protein native state and the establishment of intermolecular contacts that lead to protein deposition. A recent wide human proteome analysis has revealed that regions relevant to folding are also important to aggregation and, more interestingly, that aggregation propensity is anticorrelated with folding propensity (Tartaglia and Vendruscolo 2010). This competition between folding and aggregation has been demonstrated by fusing a fluorescent reporter to aggregation-prone peptides in such a way that the final aggregate fluorescence can be directly correlated with the *in vivo* aggregation rate (de Groot and Ventura 2006). In a recent work, we have been able to visualize in real time IBs formation *in vivo* under confocal microscopy for proteins displaying differential intrinsic aggregation propensity and confirm this theory (Villar-Pique, de Groot et al. 2012).

3.2. Molecular determinants of IBs aggregation

Nowadays, it is widely accepted that the ability to self-assemble into amyloid-like structures is an inherent property of proteins (Chiti and Dobson 2006). However, the tendency to aggregate dramatically varies depending on the intrinsic characteristics of the polypeptide chains and the environmental conditions, such as pH, ionic strength or protein concentration (DuBay, Pawar et al. 2004). The amyloid-like nature of IBs permits to study the molecular determinants in their formation in a manner that is analogous to that used for the study of the determinants involved in the *in vitro* assembly of amyloid fibrils (Winkelmann, Calloni et al. 2010). In this sense, several mutagenesis studies have revealed the essential role played by specific side-chains properties, such as hydrophobicity, secondary structure propensity or net charge in amyloid formation (Polverino de Laureto, Taddei et al. 2003). A deeply studied molecule is the amyloid- β peptide, whose central hydrophobic cluster, composed by 5 residues (LVFFA), plays a crucial role in its non-native deposition (Esler, Stimson et al. 1996). Substitutions in this stretch with polar or charged amino acids decrease the *in vitro* aggregation propensity of the complete 42-residues sequence. The same applies for IBs formation, where it exists an absolute correlation between the intrinsic physico-chemical properties of the side chains in the central position of this sequence stretch (Phe19 in the wild type sequence) and the intracellular aggregation propensity of the mutated peptide in bacteria (de Groot, Aviles et al. 2006).

Usually, the cross- β sheet core of the amyloid fibrils or amyloid-like IBs is composed by short hydrophobic stretches, named “hot spots” (HSs) with high aggregation propensity (Conchillo-Sole, de Groot et al. 2007), confirming the accepted idea that not all of the polypeptide sequence is equally relevant for protein aggregation (Lopez de la Paz and

Serrano 2004; Ventura, Zurdo et al. 2004). These amyloidogenic regions modulate the aggregation propensity of the entire polypeptide chain thus regulating the *in vivo* IB formation rate (de Groot and Ventura 2006). However, the effect of HSs driving to aggregation can be counteracted by the presence of “gatekeepers” (Figure 2).. These short regions are enriched in residues that disfavour the formation of β structure, such as proline or charged residues (Reumers, Maurer-Stroh et al. 2009). They usually flank HS and are thought to promote the sequence specific binding of several molecular chaperones (Rousseau, Serrano et al. 2006). These sequential determinants of aggregation are absolutely crucial in the case of intrinsically disordered proteins, whose lack of a defined three-dimensional structure provokes the complete exposure of amyloidogenic regions, if they are present. Hence, their aggregation rate is basically modulated by the intrinsic aggregation propensity of the primary sequence. This effect underlies aggregation in yeast prions where unstructured regions in the soluble form and the prion forming domain embedded in the cross- β core of the amyloids overlap (Chiti and Dobson 2006). In globular proteins, non-structured regions are also relevant to aggregation, despite, in some cases, disordered and flexible loops can prevent it by means of the entropy penalty associated to the required conformational change (De Simone, Kitchen et al. 2012). Moreover, despite protein deposition is highly influenced by the aggregation propensity of the primary sequence, in the case of globular proteins is also regulated by their conformational stability (Chiti and Dobson 2006). In natively folded proteins, HSs or amyloidogenic stretches are usually buried in the hydrophobic core, due to their functional role in the hydrophobic collapse during folding, which acts also as an evolutionary mechanism to prevent protein self-assembly (Lin and Zewail 2012). Therefore, the population of at least partial unfolded conformers is required in order to uncover HSs and trigger aggregation. A favourable scenario for that reaction occurs

during recombinant expression (synthesis, transport and folding) when the nascent polypeptide is still partially folded and the bulk of protein input saturates the protein quality machinery of the cell (Hendrick and Hartl 1993), leaving aggregation-prone regions free to establish non-native intermolecular contacts. Nevertheless, even when the protein has attained its native state, the strong correlation observed between the aggregation propensity of globular proteins and their conformational stability suggests that local fluctuations of the native state might also leave aggregation-prone regions transiently exposed and prone to self-assembly (Calloni, Zoffoli et al. 2005) (Espargaro, Castillo et al. 2008) (Chiti and Dobson 2009). In this sense, the stabilization of the native state has been proposed as a therapeutic approach in conformational diseases (Liu and Bitan 2012) and has been shown as a successful strategy to increase the solubility in the recombinant production of globular proteins (Castillo, Espargaro et al. 2010).

3.3. Sequence specificity in IBs formation

The sequence specific nature of the contacts leading to *in vivo* IBs formation has been long discussed (Hart, Rinas et al. 1990) (Speed, Wang et al. 1996). One of the first evidences came along with the simultaneous coexpression of two different proteins in *E. coli*. It resulted in the appearance of two types of IBs in a single cell differing in morphology and composition and separable by differential centrifugation (Hart, Rinas et al. 1990). More recently, our group reported the spatial distribution inside IBs when two different polypeptides are expressed in the same cell. Using two amyloidogenic proteins fused to fluorescent reporters, we made use of FRET measurements to determine the lack of significant co-localization of both proteins in IBs. Interestingly, we demonstrated that, although the aggregates were made up of both proteins, their relative aggregation

propensities determined their relative positions in the IB, being the most aggregation prone one, the one embedded in the inner core (Morell, Bravo et al. 2008). This kinetic segregation of protein aggregation confirmed that interactions leading to the formation of the aggregate occur preferentially between homologous sequences. This property has also been demonstrated by means of seeding experiments. In independent works, purified bacterial IBs have been used to seed amyloid fibrils assembly (Carrio, Gonzalez-Montalban et al. 2005) (Fernandez-Tresguerres, de la Espina et al. 2010) (Dasari, Espargaro et al. 2011), demonstrating the specific amyloid-like nature of IBs (Figure 1M). This specificity was best visualized using cross-seeding reactions. As an example, the IBs formed by amyloid peptides A β 40 and A β 42 were able to seed any of the two soluble monomers. However, the IBs formed by the fungal prion HET-s did not affect either the nucleation or elongation rate of the amyloid formation processes of A β peptides (Dasari, Espargaro et al. 2011). Again, the cross-seeding between both A β peptides suggested that their aggregation reactions are nucleated through common regions, in agreement with many data demonstrating that this is the case for the corresponding amyloid fibrils (Kuperstein, Broersen et al. 2010).

4. IBs as the simplest model for *in vivo* amyloid toxicity

The similarity between bacterial IBs and amyloid fibrils suggests the possibility to exploit these cheap and easy obtainable protein aggregates as a model in the study of amyloid aggregation, especially for that related to human disorders where protein deposits are pathogenic. Thus, it is likely that bacteria, despite being a prokaryotic and simple organism, might provide us with a valuable tool to dissect the mechanisms associated to aggregation associated toxicity (Garcia-Fruitos, Sabate et al. 2011). The intrinsic toxicity

of IBs is still an open question with evidences pointing to opposite theories. A detoxification role of IBs has been proposed based on the idea that oligomers and prefibrillar aggregates are the toxic agents in the cellular environment (Kirkkitadze, Bitan et al. 2002). Hence, a mature protein deposit may act as an aggregation nucleus sequestering the toxic, still soluble species. However, when the IBs formed by amyloid proteins are added to neurons they turn to be highly toxic (de Groot, Sabate et al. 2009). The analysis of the fitness cost of protein aggregation for living bacterial cells has allowed explaining this apparent contradiction.

4.1 The fitness cost of amyloid aggregation

Albeit several examples of functional amyloid structures have been described (Fowler, Koulov et al. 2007), to a large extent, protein deposits are assumed to be deleterious for the cell. Accordingly, protein aggregation is evolutionary constrained. Organisms have developed different strategies to cope with this deleterious reaction and, at the same time, polypeptide chains have been shaped to diminish the risk of non-native self-assembly (Monsellier and Chiti 2007) (Pontarotti, Villar-Pique et al. 2012). Still, since aggregation-prone regions are also essential for the folding, the stability of the native state or the formation of quaternary structures, they cannot be completely avoided in protein sequences. As a consequence, most proteins should cope with a basal intrinsic aggregation risk.

Although the effect exerted by protein aggregates in bacterial physiology is still unclear, there exist evidences that point out to a toxic gain of function of proteins upon aggregation in the bacterial cytosol. One of the first data supporting this view was provided by a study addressing the assembly of the yeast Sup35 prion protein in the cytoplasm of bacteria. It described the formation of amyloid-like aggregates, whose

presence results in a decreased colony-forming ability once the bacterial population reaches the stationary phase (Ono, Kubota et al. 2006). It has been demonstrated that during bacterial division, protein aggregates are asymmetrically segregated giving rise to two cells with differential intracellular content (Stewart, Madden et al. 2005). Asymmetric division results in an old pole daughter cell inheriting the majority of deposits and displaying an aged phenotype which is reflected in a much lower division rate (Lindner, Madden et al. 2008) (Winkler, Seybert et al. 2010) (Fernandez-Tresguerres, de la Espina et al. 2010). Whether aggregation burden is the cause or the consequence of this aged phenotype is still a question to be answered. If protein aggregation impairs cell fitness in bacteria, then asymmetrically division would act as an evolutionary mechanism to keep the damage out of the new cells, promoting the rejuvenation of the bacterial population. Supporting this view, using a large set of amyloidogenic peptides with different intrinsic aggregation tendencies, we have recently reported a highly significant correlation between the aging effect in bacteria and the aggregation propensity of the peptide. We have observe that the peptides with higher intrinsic aggregation tendency form more compact and resistant IBs as a consequence of their faster assembly rate, promoting division defects to a larger extent (Villar-Pique, de Groot et al. 2012). These results indicate that the toxic effect promoted by protein aggregates in bacteria depends on their formation kinetics and/or conformational quality, despite the mechanism underlying this damage is yet to be discovered. A recent work has converged to demonstrate the existence of a fitness cost for aggregation in bacteria by means of mutagenesis experiments in gatekeeper residues. As explained above, these stretches counteract the aggregation propensity of the HSs (Figure 2). Using a recombinant fusion of a short aggregation-prone peptide with GFP, the authors demonstrate that amino acid substitutions in the gatekeeper residues flanking the aggregation moiety severely impact the aggregation

burden of the cell, promoting differences in cell fitness and affecting the proteostatic response. (Beerten, Jonckheere et al. 2012).

4.2. Cytotoxicity of amyloid IBs

The addition of amyloid material to mammalian cells results in cell metabolism impairment and death, demonstrating the toxic nature of the protein structures embedded in these aggregates. It is now clear that the most toxic conformations correspond to the prefibrillar amyloid oligomers rather than the mature fibrils, independently if they are assemblies of disease related proteins or not (Sakono and Zako 2010 ; Bucciantini, Giannoni et al. 2002; Baglioni, Casamenti et al. 2006). Therefore, the amyloid-like nature of IBs has arisen the question of their possible cytotoxicity for mammalian cells and tissues. We addressed this question using the Alzheimer's disease related peptides A β 40 and A β 42. These peptides were expressed in bacteria resulting in the formation of insoluble aggregates displaying physico-chemical features similar to those of the corresponding amyloid fibrils. Their addition into mammalian cell cultures induced metabolic damage, as measured by means of the cell proliferation MTT assay. Interestingly enough, this cytotoxicity was higher for the IBs formed by A β 42, which display higher amyloidogenicity and contains more oligomeric structures than A β 40, demonstrating that the amyloid propensity correlates with *in vivo* toxicity (Dasari, Espargaro et al. 2011). Furthermore, another work carried out with β -galactosidase fused to an amyloidogenic protein of foot-and-mouth disease virus (VP1) reported that *in vitro* thermal aggregates are not toxic, in contrast to purified IBs produced in bacteria (Gonzalez-Montalban, Villaverde et al. 2007). As the first ones are more compact and structured than IBs, these results converge with the idea that prefibrillar aggregates are

more toxic than mature fibrils. A recent work using different ataxin variants links the cell fitness cost caused by the different recombinant proteins with the amount of oligomeric species, suggesting thus that the mechanisms of toxicity in bacteria and eukaryotic cells are somehow evolutionarily conserved (Invernizzi, Aprile et al. 2012).

4.3. Infectious properties of IBs

Amyloid fibrils of prion proteins are characterized by their self-perpetuating and infectivity nature. The expression of these proteins in bacterial cells results in the formation of IBs with amyloid-like properties similar to those of their amyloid fibrils in their natural environment (Espargaro, Villar-Pique et al. 2012 ; Sabate, Espargaro et al. 2009). Regardless of the relevance of this structural information, the most interesting and new property of these IBs is their infective nature. This is the case of the HET-s, a *Podospora anserina* prion, whose IBs were used to transfect prion-free *P. anserina* strains, resulting in the acquisition of the prion phenotype at levels comparable to that induced by the HET-s amyloid fibrils (Wasmer, Benkemoun et al. 2009). The infectious properties of bacterial IBs have also been demonstrated for the yeast prion domain of Sup35 (Sup35-NM) in two independent studies. In the first study, the expressed protein was a variant of Sup35-NM displaying an increased propensity to convert into the prion conformation (Garrity, Sivanathan et al. 2010) whereas in the second study the wild type Sup35-NM was used instead (Figure 3). In both cases, the aggregates were used to infect yeast prion-free strains and induced the prion phenotype (Espargaro, Villar-Pique et al. 2012). Interestingly enough, the infectivity of these bacterial aggregates can be easily tuned by controlling the conditions during protein production. These experiments unequivocally demonstrate that the bacterial cytoplasm allows the formation of prion

aggregates with infectious and transmissible capabilities and, because the infectious properties of amyloids depend on highly specific structural characteristics, provide perhaps the most concluding data in favour of the amyloid nature of IBs.

5. Using IBs to screen for amyloid inhibitors

The increasing impact of amyloid-linked neurodegenerative diseases like Alzheimer's or Parkinson's disorders in our aging society is pushing the discovery of new approaches to identify molecules able to interfere with aggregation pathways in order to halt the progress of these devastating pathologies or even prevent their apparition. Screening methods for amyloid inhibitors are traditionally performed employing synthetic peptides. However, their use presents several disadvantages: the peptides are expensive, they tend to be highly insoluble causing reproducibility problems in the assays and, more importantly, the inhibitors are selected under physiologically non-relevant conditions. Also, the early detection of soluble prefibrillar aggregates is difficult as these assays usually measure aggregation by means of turbidity or using amyloid dyes and early aggregates are invisible to these tests.

A considerable advance in this area has come along with the implementation of *in vivo* systems based on the use of bacteria models. The amyloid-like structure of bacterial aggregates has allowed their exploitation as a cheap and powerful tool for screening assays. In this way, the demonstration that the assembly rate of amyloidogenic proteins into IBs is modulated by external factors provides a means to test the effect of small compounds on the aggregation of these polypeptides.

Waldo and coworkers pioneered the use of the GFP as a folding reporter, demonstrating that the activity of a fluorescent tag fused to a protein depends on the solubility of the

latter. Therefore, the fluorescence of cells expressing such protein chimera correlates with the aggregation propensity of the protein (Waldo, Standish et al. 1999). This approach has been exploited for the high-throughput screen of amyloid peptide aggregation inhibitors using 96 wells plates. On one hand, it has been employed using living bacterial cells (Kim, Kim et al. 2006). The main advantage of this system is the possibility to study the effect of the tested compounds in a biological relevant environment. Nonetheless, this method is limited by the ability of these compounds to enter the cells, introducing a bias on the properties of the screened library. Recently we have overcome this problem using purified IBs instead of complete cells. As it happens *in vivo*, we have shown that the *in vitro* refolding of these aggregates is sensible to the external conditions, such as the presence of aggregation promoters or inhibitors, which can be easily detected reading the final recovered GFP fluorescence after refolding (Villar-Pique, Espargaro et al. 2012) (Figure 4). Despite the simplicity of this assay, it has been demonstrated to distinguish between aggregation pathways and to provide a qualitative assessment of amyloid assembly modulators, like metallic ions. Beyond fluorescent tag based methods, intracellular aggregation can also be measured using other engineered proteins. One interesting example is the system based on the Tat protein export pathway, which uses the β -lactamase enzyme as a reporter. A ternary fusion between this enzyme, the Tat signal peptide and the target protein is able to cross the inner membrane only if the latter is properly folded, thus conferring ampicillin resistance to the cell (Fisher, Kim et al. 2006). On the contrary, the aggregation of the target protein prevents the periplasmatic localization of the β -lactamase, sensitizing cells to the antibiotic. This genetic selection method can be used as well as a bacterial cell-based screening assay for aggregation inhibitors (Lee, Ha et al. 2009). Finally, our group has recently developed a Thioflavin-S based tool to detect amyloid-like conformations *in vivo*.

Bacterial cells containing IBs are stained with this dye, which experiments spectroscopic changes upon binding to the amyloid structures inside bacterial aggregates. The coupling of this procedure with the high-speed, multi-parametric data acquisition and analysis properties of Flow Cytometry (FC) permits its use in wide scale analysis of large bacterial populations (Espargaro, Sabate et al. 2012). This fast, quantitative and high-throughput method has the potential to be used in screening assays for amyloid inhibitors, being an *in vivo*, non-invasive tool that overcomes the use of fusions to protein reporters, allowing thus to study natural, unmodified, amyloidogenic proteins.

6. Conclusions

In the present chapter, we illustrate how an increasing number of studies are taking advantage of the amyloid-like nature of the intracellular aggregates formed in bacteria for a number of different purposes that range from the mechanistic analysis of the process of protein aggregation in an *in vivo*-like environment to the identification of novel anti-amyloidogenic compounds. In the next future, it is likely that we will witness the use of bacterial systems to understand and model the evolutionary constraint imposed by aggregation on polypeptidic sequences or the use of proteomic tools to understand the cellular response to specific amyloid conformers. Overall, bacteria provide us with a golden lab tube for the study of amyloids.

Arie, J. P., M. Miot, et al. (2006). "Formation of active inclusion bodies in the periplasm of *Escherichia coli*." Mol Microbiol **62**(2): 427-37.

- Baglioni, S., F. Casamenti, et al. (2006). "Prefibrillar amyloid aggregates could be generic toxins in higher organisms." J Neurosci **26**(31): 8160-7.
- Beerten, J., W. Jonckheere, et al. (2012). "Aggregation gatekeepers modulate protein homeostasis of aggregating sequences and affect bacterial fitness." Protein Eng Des Sel **25**(7): 357-66.
- Bucciantini, M., E. Giannoni, et al. (2002). "Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases." Nature **416**(6880): 507-11.
- Calloni, G., S. Zoffoli, et al. (2005). "Investigating the effects of mutations on protein aggregation in the cell." J Biol Chem **280**(11): 10607-13.
- Carrio, M., N. Gonzalez-Montalban, et al. (2005). "Amyloid-like properties of bacterial inclusion bodies." J Mol Biol **347**(5): 1025-37.
- Carrio, M. M. and A. Villaverde (2005). "Localization of chaperones DnaK and GroEL in bacterial inclusion bodies." J Bacteriol **187**(10): 3599-601.
- Castillo, V., A. Espargaro, et al. (2010). "Deciphering the role of the thermodynamic and kinetic stabilities of SH3 domains on their aggregation inside bacteria." Proteomics **10**(23): 4172-85.
- Castillo, V., R. Grana-Montes, et al. (2011). "The aggregation properties of Escherichia coli proteins associated with their cellular abundance." Biotechnology Journal **6**(6): 752-60.
- Chiti, F. and C. M. Dobson (2006). "Protein misfolding, functional amyloid, and human disease." Annu Rev Biochem **75**: 333-66.
- Chiti, F. and C. M. Dobson (2009). "Amyloid formation by globular proteins under native conditions." Nat Chem Biol **5**(1): 15-22.

- Conchillo-Sole, O., N. S. de Groot, et al. (2007). "AGGRESCAN: a server for the prediction and evaluation of "hot spots" of aggregation in polypeptides." BMC Bioinformatics **8**: 65.
- Daggett, V. and A. R. Fersht (2009). "Protein folding and binding: moving into uncharted territory." Curr Opin Struct Biol **19**(1): 1-2.
- Dasari, M., A. Espargaro, et al. (2011). "Bacterial inclusion bodies of Alzheimer's disease beta-amyloid peptides can be employed to study native-like aggregation intermediate states." Chembiochem **12**(3): 407-23.
- de Groot, N. S., F. X. Aviles, et al. (2006). "Mutagenesis of the central hydrophobic cluster in Abeta42 Alzheimer's peptide. Side-chain properties correlate with aggregation propensities." Febs J **273**(3): 658-68.
- de Groot, N. S., R. Sabate, et al. (2009). "Amyloids in bacterial inclusion bodies." Trends Biochem Sci **34**(8): 408-16.
- de Groot, N. S. and S. Ventura (2006). "Effect of temperature on protein quality in bacterial inclusion bodies." FEBS Lett **580**(27): 6471-6.
- de Groot, N. S. and S. Ventura (2006). "Protein activity in bacterial inclusion bodies correlates with predicted aggregation rates." J Biotechnol **125**(1): 110-3.
- de Groot, N. S. and S. Ventura (2010). "Protein aggregation profile of the bacterial cytosol." PLoS One **5**(2): e9383.
- De Simone, A., C. Kitchen, et al. (2012). "Intrinsic disorder modulates protein self-assembly and aggregation." Proc Natl Acad Sci U S A **109**(18): 6951-6.
- Dobson, C. M. (2003). "Protein folding and misfolding." Nature **426**(6968): 884-90.
- Doglia, S. M., D. Ami, et al. (2008). "Fourier transform infrared spectroscopy analysis of the conformational quality of recombinant proteins within inclusion bodies." Biotechnol J **3**(2): 193-201.

- DuBay, K. F., A. P. Pawar, et al. (2004). "Prediction of the absolute aggregation rates of amyloidogenic polypeptide chains." J Mol Biol **341**(5): 1317-26.
- Esler, W. P., E. R. Stimson, et al. (1996). "Point substitution in the central hydrophobic cluster of a human beta-amyloid congener disrupts peptide folding and abolishes plaque competence." Biochemistry **35**(44): 13914-21.
- Espargaro, A., V. Castillo, et al. (2008). "The in vivo and in vitro aggregation properties of globular proteins correlate with their conformational stability: the SH3 case." J Mol Biol **378**(5): 1116-31.
- Espargaro, A., R. Sabate, et al. (2008). "Kinetic and thermodynamic stability of bacterial intracellular aggregates." FEBS Lett **582**(25-26): 3669-73.
- Espargaro, A., R. Sabate, et al. (2012). "Thioflavin-S staining coupled to flow cytometry. A screening tool to detect in vivo protein aggregation." Mol Biosyst **8**(11): 2839-44.
- Espargaro, A., A. Villar-Pique, et al. (2012). "Yeast prions form infectious amyloid inclusion bodies in bacteria." Microb Cell Fact **11**(1): 89.
- Fernandez-Busquets, X., N. S. de Groot, et al. (2008). "Recent structural and computational insights into conformational diseases." Current Medicinal Chemistry **15**(13): 1336-49.
- Fernandez-Tresguerres, M. E., S. M. de la Espina, et al. (2010). "A DNA-promoted amyloid proteinopathy in Escherichia coli." Mol Microbiol **77**(6): 1456-69.
- Fisher, A. C., W. Kim, et al. (2006). "Genetic selection for protein solubility enabled by the folding quality control feature of the twin-arginine translocation pathway." Protein Sci **15**(3): 449-58.
- Fowler, D. M., A. V. Koulov, et al. (2007). "Functional amyloid--from bacteria to humans." Trends Biochem Sci **32**(5): 217-24.

- Garcia-Fruitos, E., R. Sabate, et al. (2011). "Biological role of bacterial inclusion bodies: a model for amyloid aggregation." Febs J **278**(14): 2419-27.
- Garcia-Fruitos, E., J. Seras-Franzoso, et al. (2010). "Tunable geometry of bacterial inclusion bodies as substrate materials for tissue engineering." Nanotechnology **21**(20): 205101.
- Garcia-Fruitos, E., E. Vazquez, et al. (2012). "Bacterial inclusion bodies: making gold from waste." Trends Biotechnol **30**(2): 65-70.
- Garrity, S.J., V. Sivanathan, et al. (2010). "Conversion of a yeast prion protein to an infectious form in bacteria." Proc Natl Acad Sci U S A **107**(23): 10596-601.
- Gonzalez-Montalban, N., A. Villaverde, et al. (2007). "Amyloid-linked cellular toxicity triggered by bacterial inclusion bodies." Biochem Biophys Res Commun **355**(3): 637-42.
- Hart, R. A., U. Rinas, et al. (1990). "Protein composition of Vitreoscilla hemoglobin inclusion bodies produced in Escherichia coli." J Biol Chem **265**(21): 12728-33.
- Hendrick, J. P. and F. U. Hartl (1993). "Molecular chaperone functions of heat-shock proteins." Annu Rev Biochem **62**: 349-84.
- Ignatova, Z. and L. M. Gierasch (2004). "Monitoring protein stability and aggregation in vivo by real-time fluorescent labeling." Proc Natl Acad Sci U S A **101**(2): 523-8.
- Invernizzi, G., F. A. Aprile, et al. (2012). "The Relationship between Aggregation and Toxicity of Polyglutamine-Containing Ataxin-3 in the Intracellular Environment of Escherichia coli." PLoS One **7**(12): e51890.
- Kim, W., Y. Kim, et al. (2006). "A high-throughput screen for compounds that inhibit aggregation of the Alzheimer's peptide." ACS Chem Biol **1**(7): 461-9.

- Kirkitadze, M. D., G. Bitan, et al. (2002). "Paradigm shifts in Alzheimer's disease and other neurodegenerative disorders: the emerging role of oligomeric assemblies." J Neurosci Res **69**(5): 567-77.
- Kuperstein, I., K. Broersen, et al. (2010). "Neurotoxicity of Alzheimer's disease Abeta peptides is induced by small changes in the Abeta42 to Abeta40 ratio." EMBO J **29**(19): 3408-20.
- Lee, L. L., H. Ha, et al. (2009). "Discovery of amyloid-beta aggregation inhibitors using an engineered assay for intracellular protein folding and solubility." Protein Sci **18**(2): 277-86.
- Lin, M. M. and A. H. Zewail (2012). "Hydrophobic forces and the length limit of foldable protein domains." Proc Natl Acad Sci U S A **109**(25): 9851-6.
- Lindner, A. B., R. Madden, et al. (2008). "Asymmetric segregation of protein aggregates is associated with cellular aging and rejuvenation." Proc Natl Acad Sci U S A **105**(8): 3076-81.
- Liu, T. and G. Bitan (2012). "Modulating self-assembly of amyloidogenic proteins as a therapeutic approach for neurodegenerative diseases: strategies and mechanisms." ChemMedChem **7**(3): 359-74.
- Lopez de la Paz, M. and L. Serrano (2004). "Sequence determinants of amyloid fibril formation." Proc Natl Acad Sci U S A **101**(1): 87-92.
- Monsellier, E. and F. Chiti (2007). "Prevention of amyloid-like aggregation as a driving force of protein evolution." EMBO Rep **8**(8): 737-42.
- Morell, M., R. Bravo, et al. (2008). "Inclusion bodies: specificity in their aggregation process and amyloid-like structure." Biochim Biophys Acta **1783**(10): 1815-25.
- Nothaft, H. and C. M. Szymanski (2010). "Protein glycosylation in bacteria: sweeter than ever." Nat Rev Microbiol **8**(11): 765-78.

- Ono, B., M. Kubota, et al. (2006). "Production of a polymer-forming fusion protein in *Escherichia coli* strain BL21." Biosci Biotechnol Biochem **70**(12): 2813-23.
- Polverino de Laureto, P., N. Taddei, et al. (2003). "Protein aggregation and amyloid fibril formation by an SH3 domain probed by limited proteolysis." J Mol Biol **334**(1): 129-41.
- Pontarotti, P., A. Villar-Pique, et al. (2012). Protein Aggregation Acts as Strong Constraint During Evolution. Evolutionary Biology: Mechanisms and Trends, Springer Berlin Heidelberg: 103-120.
- Reumers, J., S. Maurer-Stroh, et al. (2009). "Protein sequences encode safeguards against aggregation." Hum Mutat **30**(3): 431-7.
- Rokney, A., M. Shagan, et al. (2009). "E. coli transports aggregated proteins to the poles by a specific and energy-dependent process." J Mol Biol **392**(3): 589-601.
- Rousseau, F., L. Serrano, et al. (2006). "How evolutionary pressure against protein aggregation shaped chaperone specificity." J Mol Biol **355**(5): 1037-47.
- Sabate, R., N. S. de Groot, et al. (2010). "Protein folding and aggregation in bacteria." Cellular and Molecular Life Sciences **67**(16): 2695-715.
- Sabate, R., A. Espargaro, et al. (2009). "Characterization of the amyloid bacterial inclusion bodies of the HET-s fungal prion." Microb Cell Fact **8**: 56.
- Sakono, M. and T. Zako (2010). "Amyloid oligomers: formation and toxicity of Abeta oligomers." Febs J **277**(6): 1348-58.
- Speed, M. A., D. I. Wang, et al. (1996). "Specific aggregation of partially folded polypeptide chains: the molecular basis of inclusion body composition." Nat Biotechnol **14**(10): 1283-7.
- Stewart, E. J., R. Madden, et al. (2005). "Aging and death in an organism that reproduces by morphologically symmetric division." PLoS Biol **3**(2): e45.

- Tartaglia, G. G. and M. Vendruscolo (2010). "Proteome-level interplay between folding and aggregation propensities of proteins." J Mol Biol **402**(5): 919-28.
- Tartaglia, G. G. and M. Vendruscolo (2009). "Correlation between mRNA expression levels and protein aggregation propensities in subcellular localisations." Mol Biosyst **5**(12): 1873-6.
- Upadhyay, A. K., A. Murmu, et al. (2012). "Kinetics of inclusion body formation and its correlation with the characteristics of protein aggregates in Escherichia coli." PLoS One **7**(3): e33951.
- Ventura, S. and A. Villaverde (2006). "Protein quality in bacterial inclusion bodies." Trends Biotechnol **24**(4): 179-85.
- Ventura, S., J. Zurdo, et al. (2004). "Short amino acid stretches can mediate amyloid formation in globular proteins: the Src homology 3 (SH3) case." Proc Natl Acad Sci U S A **101**(19): 7258-63.
- Villar-Pique, A., N. S. de Groot, et al. (2012). "The effect of amyloidogenic peptides on bacterial aging correlates with their intrinsic aggregation propensity." J Mol Biol **421**(2-3): 270-81.
- Villar-Pique, A., A. Espargaro, et al. (2012). "Using bacterial inclusion bodies to screen for amyloid aggregation inhibitors." Microb Cell Fact **11**: 55.
- Waldo, G. S., B. M. Standish, et al. (1999). "Rapid protein-folding assay using green fluorescent protein." Nat Biotechnol **17**(7): 691-5.
- Wang, L. (2009). "Towards revealing the structure of bacterial inclusion bodies." Prion **3**(3): 139-45.
- Wang, L., S. K. Maji, et al. (2008). "Bacterial inclusion bodies contain amyloid-like structure." PLoS Biol **6**(8): e195.

- Wasmer, C., L. Benkemoun, et al. (2009). "Solid-state NMR spectroscopy reveals that E. coli inclusion bodies of HET-s(218-289) are amyloids." Angew Chem Int Ed Engl **48**(26): 4858-60.
- Wilson, L. M., Y. F. Mok, et al. (2007). "A structural core within apolipoprotein C-II amyloid fibrils identified using hydrogen exchange and proteolysis." J Mol Biol **366**(5): 1639-51.
- Winkelmann, J., G. Calloni, et al. (2010). "Low-level expression of a folding-incompetent protein in Escherichia coli: search for the molecular determinants of protein aggregation in vivo." J Mol Biol **398**(4): 600-13.
- Winkler, J., A. Seybert, et al. (2010). "Quantitative and spatio-temporal features of protein aggregation in Escherichia coli and consequences on protein quality control and cellular ageing." Embo J **29**(5): 910-23.

Table 1. Proteins forming amyloid-like IBs in bacteria. Regardless their native conformations, all *in vivo* formed aggregates display amyloid properties as demonstrated by different techniques.

Figure 1. Physico-chemical techniques to detect and characterize the amyloid-like structure in bacterial IBs formed by *Podospora anserina* prion HET-s. A-D: Visualization of IBs under transmission electronic microscopy. Proteolytic digestion with pK enables the apparition of fibrillar structure (D). E-H: Congo Red binding to IBs. Spectroscopic changes promoted by the IBs amyloid structure and analyzed by absorbance (E-F) and Congo Red staining of IBs visualized under wide-field (G) and fluorescence microscopy (H). I: ^{13}C - ^{13}C solid-state NMR, showing that HET-s IBs display a similar spectrum of that of HET-s *in vitro* amyloid fibrils. J-L: Secondary structure analysis of IBs, confirming the presence of intermolecular β -sheets by means of infrared spectroscopy (J-K) and circular dichroism (L). M: Seeding capacity of HET-s IBs in the *in vitro* amyloid fibrils formation. This figure has been reproduced with permission from (Garcia-Fruitos, Sabate et al. 2011).

Figure 2. Representation of hot spots and gatekeepers. Above: conceptual schema of protein aggregation driven by solvent exposed HS (represented as dotted residues). Below: the presence of gatekeepers (represented as black star-shaped residues) favours protein folding against aggregation reaction. The acquisition of the native state prevents detrimental self-assembly since HSs remain buried in the inner protein core.

Figure 3. The prion domain of the yeast prion Sup35 (Sup35-NM) is expressed in *E. coli* cells forming amyloid-like intracellular IBs. After cell lysis, soluble and insoluble fractions

are separated by centrifugation and both are used to transform spheroplasts of prion-free yeast strains (phenotype [psi-]). Upon growing on 1/4 YPD medium plates, cells converted to prion conformation (phenotype [PSI+]) appear as white or pink, whereas [psi-] remaining cells appear as red. Transformation with the insoluble fraction results in a higher prion conversion efficiency demonstrating that, at least, part of the protein embedded in these bacterial IBs acquires a transmissible prion conformation.

Figure 4. Schematic representation of a screening assay for amyloid modulators based on A β 42-GFP IBs. IBs recovered and purified from bacterial cells are chemically denatured and, subsequently, *in vitro* refolded. The method is based on the assumption that the kinetic competition between A β 42 aggregation and GFP folding occurring during IBs formation in the cell, can be reproduced in an *in vitro* environment. Thus, the presence of molecules modulators of amyloid aggregation in the refolding buffer is straightforward detected by means of the GFP fluorescence recovered after the reaction: aggregation enhancers decrease the recovered fluorescence while aggregation inhibitors promote high fluorescence recovery. This figure has been reproduced with permission from (Villar-Pique, Espargaro et al. 2012).

Table 1

	Disordered		fold		and fold		fold	Prion proteins		
	A peptide	VP1	ESAT-6	Human growth hormone	BMP2 (13-74)	L-asparaginase II	MOG (ECD)	Het-s	Ure2p	Sup35
Amyloid dyes	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Infra red spectroscopy	✓	✓		✓		✓		✓	✓	✓
Circular dichroism	✓		✓					✓		
Seeding ability	✓	✓						✓	✓	✓
Solid-state NMR	✓							✓		
H/D exchange solution-state NMR	✓		✓		✓		✓			
AFM tip indentation	✓									
Denaturation curves	✓	✓		✓		✓		✓	✓	✓
Proteolytic curves	✓	✓		✓		✓				
X-ray diffraction			✓		✓		✓			
Atomic force microscopy	✓									
Electronic microscopy	✓	✓	✓	✓	✓	✓		✓		✓
Reference*	1 2 3 4 5	6 4 7	8	9	8	9	8	10 11	12	13 12

*

1. M. Dasari, A. Espargaro, R. Sabate et al., *Chembiochem* **12** (3), 407.
2. N. S. de Groot and S. Ventura, *FEBS Lett* **580** (27), 6471 (2006).
3. A. Espargaro, R. Sabate, and S. Ventura, *FEBS Lett* **582** (25-26), 3669 (2008).
4. M. Morell, R. Bravo, A. Espargaro et al., *Biochim Biophys Acta* **1783** (10), 1815 (2008).
5. A. Villar-Pique, N. S. de Groot, R. Sabate et al., *J Mol Biol* **421** (2-3), 270.
6. M. Carrio, N. Gonzalez-Montalban, A. Vera et al., *J Mol Biol* **347** (5), 1025 (2005).
7. O. Cano-Garrido, E. Rodriguez-Carmona, C. Diez-Gil et al., *Acta Biomater.*
8. L. Wang, S. K. Maji, M. R. Sawaya et al., *PLoS Biol* **6** (8), e195 (2008).
9. A. K. Upadhyay, A. Murmu, A. Singh et al., *PLoS One* **7** (3), e33951.
10. C. Wasmer, L. Benkemoun, R. Sabate et al., *Angew Chem Int Ed Engl* **48** (26), 4858 (2009).
11. R. Sabate, A. Espargaro, S. J. Saupé et al., *Microb Cell Fact* **8**, 56 (2009).
12. A. Espargaro, A. Villar-Pique, R. Sabate et al., *Microb Cell Fact* **11** (1), 89.
13. B. Ono, M. Kubota, H. Kimiduka et al., *Biosci Biotechnol Biochem* **70** (12), 2813 (2006).

Figure 1

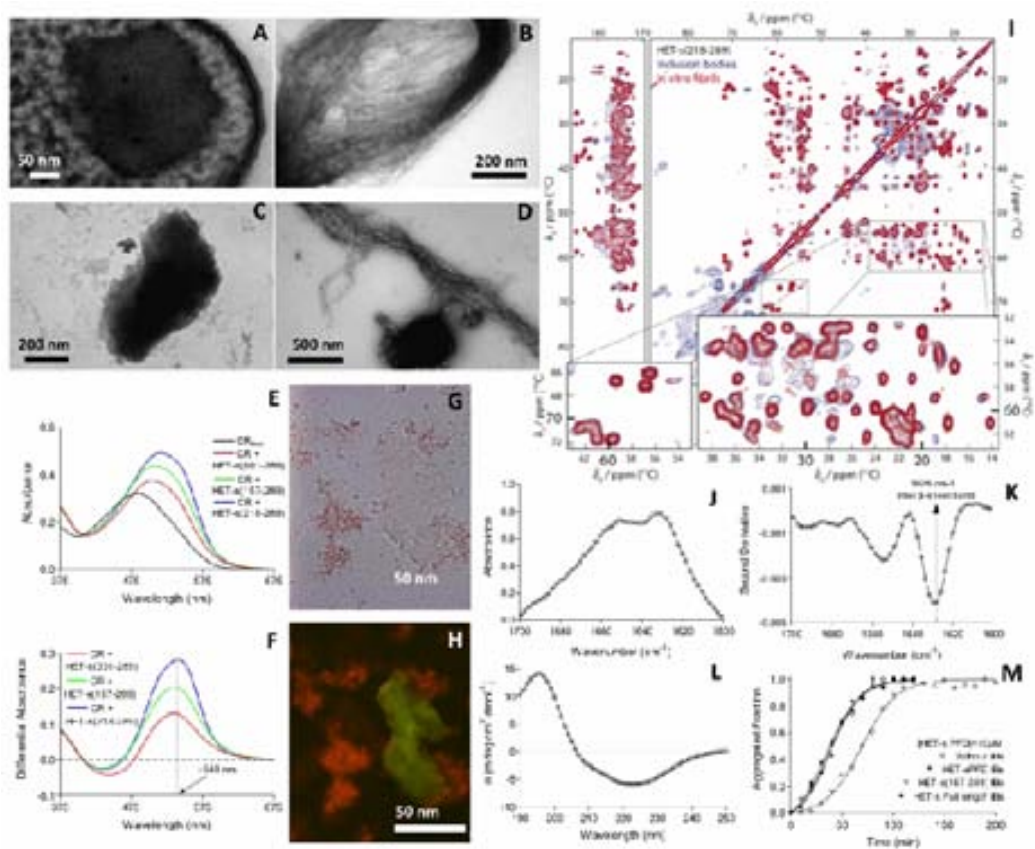


Figure 2

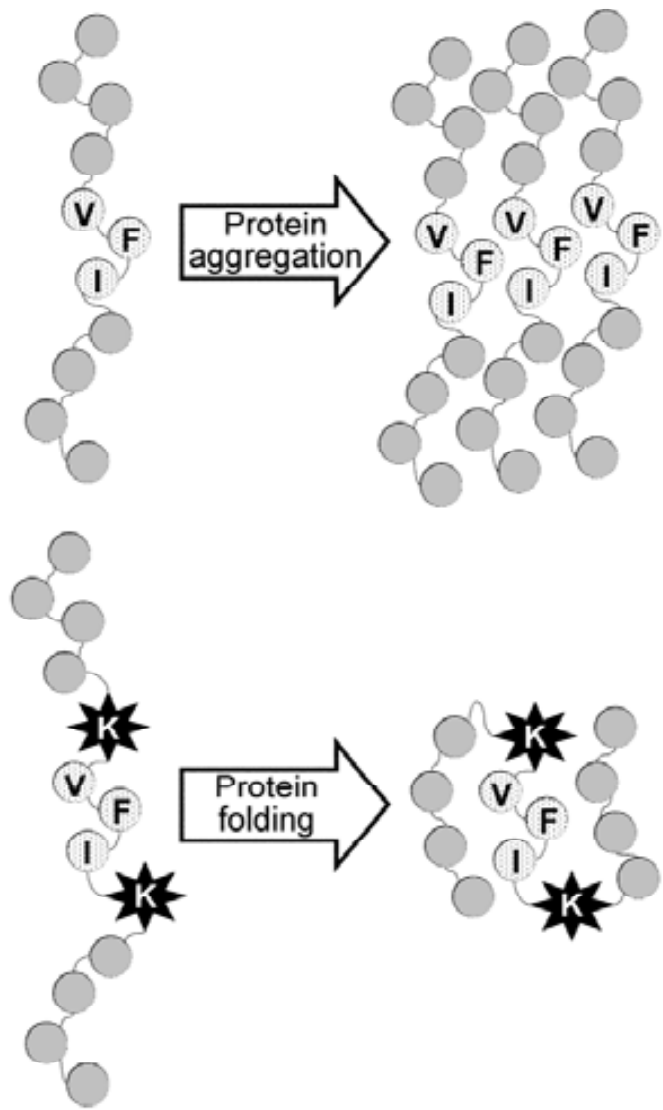


Figure 3

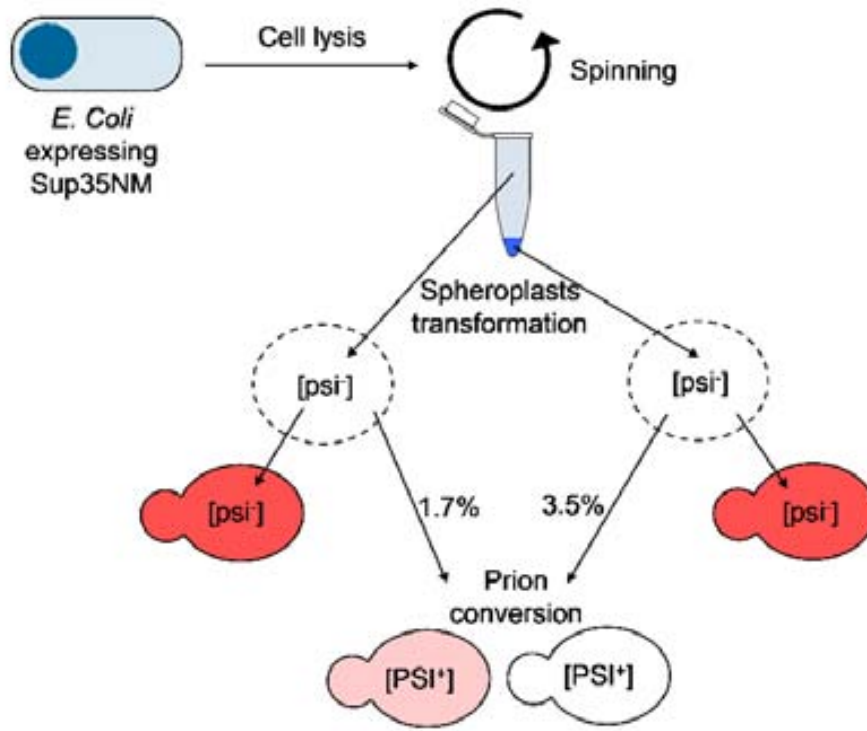
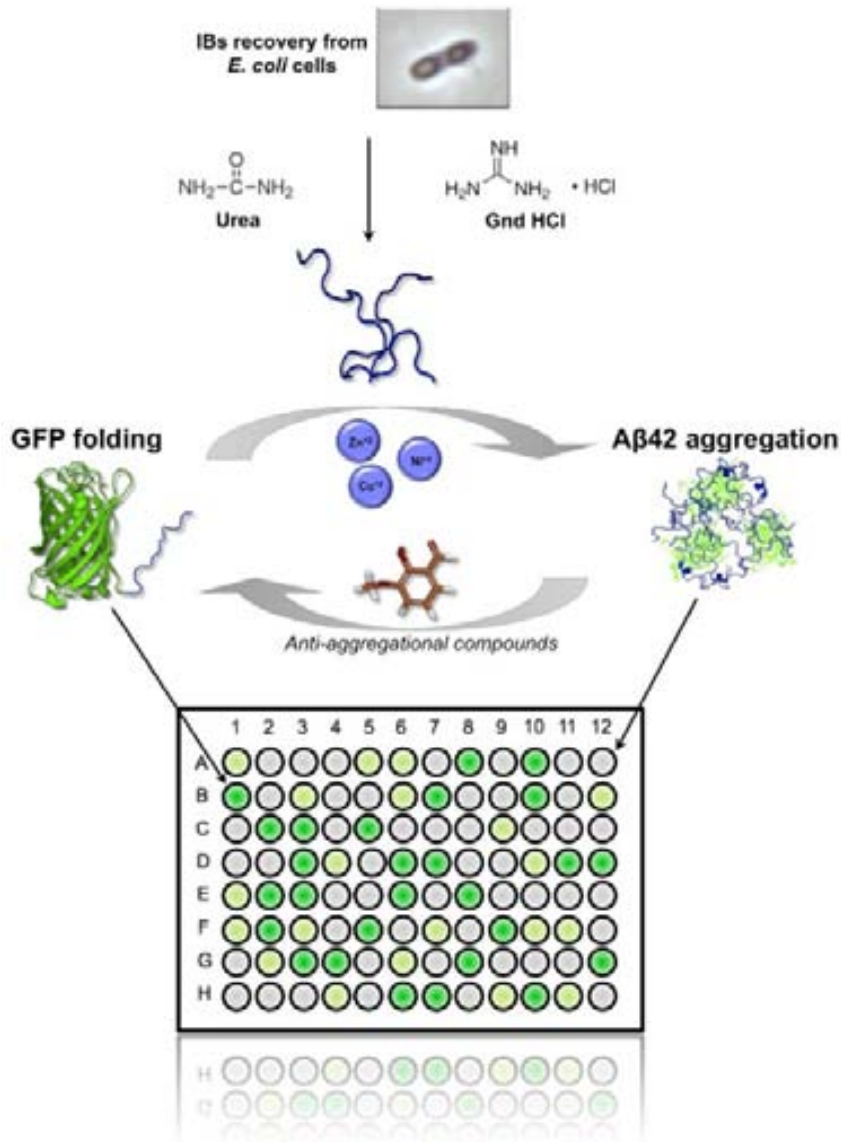


Figure 4



Discussion

4.1 Protein aggregates in unicellular organisms

Protein misfolding and aggregation are biological events linked to human conformational disorders and, thus, animal models have been employed to put the data in a biological context [116, 117]. However, due to the high complexity of these models, the molecular mechanisms underlying aggregation have been traditionally analyzed by performing *in vitro* studies where synthetic peptides or purified proteins are assembled in non-physiological conditions [118, 119].

During the last years, we and others have also addressed the study of intracellular protein aggregates by employing unicellular models, such as yeast and bacteria, which are simple but well characterized and easy to handle organisms. The discovery that protein deposits in bacterial cells display amyloid-like features [73, 75] paved the way for using this prokaryotic organism as model in the study of amyloid deposition. On this regard, a bulk of studies in our group have been focused in the AD related peptide, A β 42, fused to the GFP (A β 42-GFP). This fusion protein was exploited to build a library of proteins with different aggregation propensities, based on point mutations in the 19th residue of the amyloid peptide, a position highly relevant for the amyloidogenicity of the entire sequence [23]. Waldo and collaborators developed the fusion to a fluorescent tag as a reporter system to assess the productive folding of a protein of interest, whose solubility and, thus, functionality depended on the GFP folding [90]. Hence, the fluorescence of the host expression cell was indicative of the proper folding of the tagged protein. This tagging strategy was also implemented in our group; however, in our case, the GFP fluorescence was used as a reporter of the aggregational state of the amyloid peptide, rather than the solubility [84, 101]. Due to the high amyloidogenicity of A β 42, practically all the recombinant protein is driven to the formation of IBs, which exhibit green fluorescence. The IBs fluorescence depends on the specific introduced mutation, a property that was exploited to develop an aggregation predictor, AGGRESCAN, based on aggregation propensity values ascribed to each of the twenty natural amino acids [34]. Importantly, two relevant conclusions were extracted from those experiments: i) bacterial IBs contain native-like species, being partially functional and ii) the activity of the IBs correlates with the predicted aggregation propensity of the recombinant polypeptide. This last conclusion indicates that, in the intracellular space, there is

a kinetic competition between folding and aggregation of nascent polypeptides determining the conformational quality of the formed IB, which reflects the relation between native and non-native contacts.

In the present thesis, using A β 42-GFP mutants representing the extremes in terms of aggregation propensity, we have been able to visualize their different *in vivo* assembly rate in *E. coli*, which determines the final IB activity. Therefore, we provide an ultimate and unequivocal proof of the mentioned competition, where native intramolecular contacts leading to the GFP folding compete with the establishment of aberrant intermolecular interactions that drive the protein to the IBs. Moreover, in the same work we observed that high aggregation prone A β 42-GFP variants are assembled into more compact and resistant IBs, which are more damaging for the cell, as it will be explained below. Thus, the conformational quality of an IB is not only reflected in its activity but also in its physico-chemical properties. Hence, this latter parameter is also dependent on the aggregation propensity and, therefore, encoded in the primary sequence (Fig. 4.1).

In an independent work constituting this thesis, we expressed A β 42-GFP in the baker's yeast *S. cerevisiae*, replicating in an eukaryotic organism the experiments previously carried out in *E. coli*. Yeast has been widely employed to model conformational disorders because, despite its simplicity, it shares many of the essential metabolic pathways of complex eukaryotes and recapitulates several pathological hallmarks of neurodegenerative diseases [111]. For instance, the expression of the amyloid peptide in yeast results to be toxic when it is targeted to the secretory pathway [112]. However, in our group, the expression of A β 42 in yeast was used to implement an assay to screen for aggregation inhibitors compounds based on cell survival dependent on the A β 42 aggregation [120]. Because the aim of that work was to develop such system, it lacked of a deep analysis of protein deposition in yeast and, hence, the work presented in this thesis covers this research issue, pretending to obtain new insights in the knowledge achieved by the group about amyloid aggregation.

In order to compare amyloid aggregation occurring in an eukaryotic organism to that taking place in bacteria, we expressed the complete set of A β 42-GFP variants in yeast cytoplasm. A β 42-GFP mutants carrying amino acids in the 19th position with aromatic and hydrophobic side chains exhibited higher deposition levels when compared to mutants with polar and charged amino acids in the substituted position. We observed that the self-assembly of these fusion proteins in yeast correlates with their aggregation propensity values, as obtained from the previous *E. coli* studies. This suggests that the amino acid composition of a polypeptide sequence acts as a strong determinant of aggregation intrinsic to the protein regardless the expression conditions and the host cell. Nevertheless, the expression of the A β 42-GFP mutants in yeast also revealed some remarkable differences from that in *E. coli*, including an increase in the solubility and different A β 42-GFP levels between mutants, which are caused, at least partially, by proteolytic degradation as it will be explained below.

4.2 Amyloid aggregation promotes aging

One of the main objectives of the present work constituted the analysis of the effect of intracellular protein aggregates on cell division. At the beginning of this thesis, two independent publications appeared [80, 107] that, together with an earlier one

[82], reported that *E. coli* cells carrying protein deposits display an aged phenotype, which is reflected by a decrease in the replication rate. The binary fission in a bacillus like *E. coli* generates two different offspring cells, one derived from the new pole and another one derived from the old pole. Among the differences between both cells, the old pole offspring burdens IBs and exhibits a slower division rate. Despite ignoring whether the inheritance of protein aggregates is the cause or the consequence of aging, this asymmetric division of *E. coli* cells can be regarded as a mechanism of damage purification, since it enables the apparition of a rejuvenated population (the new pole offspring). This view was in good agreement with posterior data revealing that misfolded and aggregated proteins impose a fitness cost for the host cell [121], adding experimental evidences to the theory that protein aggregation acts as a strong evolutive constraint [45].

Here, we present a work where we took advantage of our well characterized system of A β 42-GFP variants to study aggregation-promoted aging, since we fortuitously observed that *E. coli* cells expressing these amyloidogenic proteins exhibited a filamentous phenotype. An in depth analysis of those cells revealed that they contained three or more IBs interspersed with cellular nucleoids, which indicated an impairment of the binary fission and thereby, after chromosome replication, cells elongated without dividing. Interestingly, we observed that this defect in cell division correlated with the tendency to self-assemble of the expressed A β 42-GFP. Together with the data obtained from the physico-chemical characterization of the same IBs, we demonstrated that the higher aggregation prone A β 42-GFP variants produce more resistant IBs that are more deleterious for the cell (Fig. 4.1). Our results suggest that this reproductive inability might rather be related to specific amyloid features of the IBs than to a generic response of cellular metabolism to cope with protein aggregation. Although the molecular mechanisms underlying this cytotoxicity are not understood and require further studies, taking into consideration our data, we propose two feasible explanations. On one side, the relation between the structural resistance and hardness of IBs and cell division defects suggests a possible steric impediment exerted by the protein deposits. On the other side, the difference in the assembly rate might be the cause of a differential sequestration of endogenous proteins by the nascent IB, predisposing the cellular metabolism to an abnormal functionality. In this sense, it has been reported the localization of molecular chaperones like DnaK and GroEL embedded inside IBs [122]. More interestingly, imbalance in the chaperones concentration has been described to impair the FtsZ-ring formation, indispensable for the bacterial binary fission [123].

4.3 Cellular response to protein aggregation

The interrelationship between aging and protein aggregation arises the question of how cells cope with this damaging burden. Among possible cellular responses, we focused our studies in molecular chaperones. Actually, the saturation of the protein quality machinery is one of the most important causes of protein deposition during recombinant expression [124, 125]. Thus, we addressed, *in vitro* and *in vivo*, the role of the main chaperones complexes in our model system based on A β 42-GFP. Firstly, we demonstrated that, *in vitro*, purified A β 42-GFP IBs are disaggregated and refolded by the combination of two complementary chaperone activities: i) chaperone ClpB, which acts on extracting ensembled polypeptides from IBs and

ii) the chaperone complex DnaK-DnaJ, which assists folding of nascent polypeptides and misfolded or partially unfolded species [9]. Interestingly, by performing these experiments with the *wt* protein and the F19D mutant, displaying high and low aggregation propensities, respectively; we observed a lower effect in the case of *wt* IBs. Thus, we can conclude that a lower conformational quality of IBs (which means higher compacting and resistance) enables less efficient aggregates reactivation work by molecular chaperones (Fig. 4.1).

These *in vitro* results were complemented with *in vivo* studies, using the same A β 42-GFP mutants. On this regard, the simultaneous overexpression of GroEL-GroES and DnaK-DnaJ systems, both assisting protein folding, resulted in an increase in the conformational quality of the IBs and, more interestingly, in a decrease in the levels of cell division impairment. The activity of molecular chaperones depended on the aggregation propensity of the mutant, corroborating the data obtained from *in vitro* experiments. On the contrary, the *in vivo* overexpression of ClpB did not produce any effect. Albeit coexpression with chaperones represents a common strategy to increase the solubility of the recombinant protein [125], in our system and probably due to the high amyloidogenicity of the A β 42-GFP, molecular chaperones alleviate the deleterious effect of protein aggregates by remodeling their conformational quality without a significant increase in the solubility. On one hand, this is in good agreement with the correlation between physico-chemical properties and cell division impairment described above. On the other hand, the fact that ClpB does not exhibit any significant effect in our system model probably indicates that folding assistance is not exerted in polypeptides already aggregated but rather in soluble misfolded or unfolded species, which will aggregate in the IBs with a higher proportion of native contacts. Chaperones have been described to buffer destabilizing mutations in polypeptide sequences thus preventing aggregation and contributing to protein evolution and genetic diversity [126, 127]. Our results suggest that the chaperones effect is not restricted to an increase of mutated proteins solubility, but might also extend to a buffering activity of the fitness cost promoted by protein deposits.

As aforementioned, the expression of A β 42-GFP variants in yeast, contrarily to that in bacteria, resulted in different levels of the recombinant protein between mutants. Importantly, we detected a significant correlation between protein levels and aggregation propensity, where high aggregation prone variants presented less amount of total A β 42-GFP. By genetically and chemically modulating cellular proteolytic machinery, we observed that: i) A β 42-GFP is cleared by distinct protein degradation pathways (autophagy, vacuolar proteases and proteasome complex) and ii) the aggregation propensity of A β 42-GFP modulates the proteolytic activity of the cell on the protein target. Interestingly, the contribution of the proteasome to this activity is small, in agreement with the theory that this complex tends to act on soluble misfolded species rather than on irreversible large aggregates [128]. The data obtained from these studies suggest that, in *S. cerevisiae*, aggregates clearance depends on the protein embedded in these deposits. Proteolysis is more active against proteins with a higher tendency to self-assemble, which generate aggregates displaying higher stability. Considering that, in bacteria, these aggregates are also more harmful for the cells in terms of reproductive ability, it is feasible that in yeast they also contribute more to the fitness cost associated to protein deposition. In this sense, proteolytic degradation of α -syn aggregates in yeast has been reported to be a detoxification process, increasing cell survival [129]. Therefore, it seems that cel-

lular proteolytic machinery might detect damaging aggregates and thus clear them as safeguard mechanism.

Overall, the data obtained from these works corroborate that cellular protein quality machinery, by means of molecular chaperones and proteolytic pathways, is sensible to the apparition of protein deposits. Moreover, our results also indicate that cellular response against these aggregates depends on their amyloid nature and conformational quality, which, in turn, are determined by the intrinsic aggregation propensity of the protein (Fig. 4.1). Taking into account that these protein deposits display different degree of cytotoxicity, it is tempting to propose that the cellular response to protein aggregation has evolved to be able to recognize the most dangerous particles. In the same line of evidence, using a mutagenesis strategy in gatekeepers residues, Rousseau and coworkers confirmed that polypeptides aggregation propensity conditions the expression profile of chaperones and proteases as well as the fitness cost imposed by the presence of the aggregates [29].

4.4 Bacterial IBs in the screening of amyloid modulators

Another part of the present thesis was aimed to develop a screening system for modulator compounds of amyloid deposition. The broad knowledge acquired by the group in the field of intracellular protein aggregation and the wide characterization of the IBs formed by the A β 42-GFP fusion protein paved the way for implementing them in such screening system. The system developed was based on the following *in vivo* data: i) the kinetic competition between folding and aggregation during IBs formation and, ii) the demonstration that GFP fluorescence activity is a reflection of such competition. Additionally, the implementation of the methodology required the confirmation that, similarly to *in vivo* IBs assembly, the refolding of these aggregates in *in vitro* conditions is led by the same competition between the aggregation of the A β 42 moiety and the folding of the fluorescent tag. Hence, the method consists of two consecutive steps applied to purified A β 42-GFP IBs: first, denaturation of the IBs and afterwards, refolding of the protein, which implies the recovery of a certain degree of GFP activity. The first step guarantees that the refolding process starts *de novo* from unfolded and isolated protein species. Thereby, in the subsequent refolding process, the self-assembly reaction reproduces *in vitro* the folding-aggregation competition during protein synthesis in the cell. However, the protein concentration used in the assay is much lesser than that present in the crowded cellular cytoplasm, which favours the GFP folding, extending the dynamic range of the method.

Protein deposition is not only determined by intrinsic factors associated to the polypeptide sequence but also by external conditions, thereupon, we proposed to include modulators of aggregation in the refolding buffer and detect their effect by means of the final fluorescence in a simple and cost-effective assay. Thus, any compound promoting or inhibiting amyloid aggregation would reduce or increase, respectively, the recovered fluorescence after the refolding step, as a reflection of the conformational quality of the formed aggregate. To provide the assay with a proof of concept, we based on the AD metal hypothesis and demonstrated that the method is capable to detect the aggregation enhancer ability of metallic cations, whose role in the AD pathogenicity has been widely described [130]. Importantly, we were able to detect the effect of such metals in the range of physiological concentrations. Furthermore, we evidenced that the system is also sensible to the presence of inhibitors

compounds of metal-promoted aggregation, thus, validating its use for fast screening of amyloid aggregation inhibitors.

Overall, this assay represents a relevant advancement in the field of screening methodologies for amyloid inhibitors, which are traditionally carried out using synthetic peptides. Our system overcomes some of the problems associated to these peptide-based *in vitro* assays, such as high cost and cumbersome performing. With suitable expression conditions, theoretically, any protein can form IBs, which are easily obtained from cheap and fast-growing bacteria. Furthermore, the assay is implemented using 96-well plates, allowing for high-throughput application.

4.5 Intracellular amyloids in plants

Protein deposition into amyloid conformation has been traditionally described in animals and related to pathological conditions. An important turning point in the field was, on one hand, the discovery that not only disease-related proteins can adopt this conformation, but also many other proteins not associated to pathogenicity [131]. Moreover, amyloid assemblies have been reported for proteins completely distinct in sequence and native fold [75]. On the other hand, the description of amyloid-like aggregates in fungi and bacteria was highly relevant, as well [78, 132]. Therefore, due to the increasing number of evidences suggesting that amyloid aggregation is a generic process of polypeptides occurring regardless the host organism, we purposed to characterize protein aggregates in the vegetable kingdom in order to assess their putative amyloid nature.

Recombinant expression in plants can result in the accumulation of the heterologous protein in insoluble deposits. Similarly to the scenario taking place in other expression hosts, the bulk of protein input saturates the cellular protein quality machinery and the accumulation of unfolded or misfolded species makes them to eventually aggregate. In the work herein presented, we studied the deposition of maize transglutaminase (TGZ) in the chloroplasts of tobacco transplastomic plants. Firstly, we demonstrated that this protein possesses a highly amyloidogenic region within its primary sequence and that it is capable to form amyloid-like aggregates *in vitro*. Afterwards, we analyzed the TGZ aggregates formed in the chloroplasts of transgenic tobacco demonstrating that they display amyloid-like properties. Interestingly, we also observed that the presence of TGZ aggregates provoked an aberrant leaf phenotype, including pigment deficiencies and thylakoid appression abnormalities. These deleterious effects might be the consequence of an imbalance in the TGZ activity or steric impediments provoked by the aggregate. However, it should not be discarded the possibility of a toxic gain of function caused by the amyloid structure embedded in the aggregates. Indeed, it has been previously described how proteins not related to conformational diseases become toxic when they are assembled into amyloid-like aggregates [133]. Thus, we can suggest, from our data, that recombinant proteins can adopt cytotoxic amyloid structures in transgenic plants. A dramatic question arising from this work is whether these aggregates might be toxic to other organisms, analogously to the toxicity displayed by other amyloid-like aggregates when applied to mammalian cultured cells [134]; a question that should be further addressed since the presence of those deposits might compromise the safety of transgenic plants.

Finally, the results obtained from this study are striking as it represents the first report, to the best of our knowledge, of amyloid deposits in plants. Therefore, together with the presence of amyloids in animals, fungi and bacteria, the description of amyloid-like aggregates in the vegetable kingdom suggests that this structure might be omnipresent across all the phylogenetic tree.

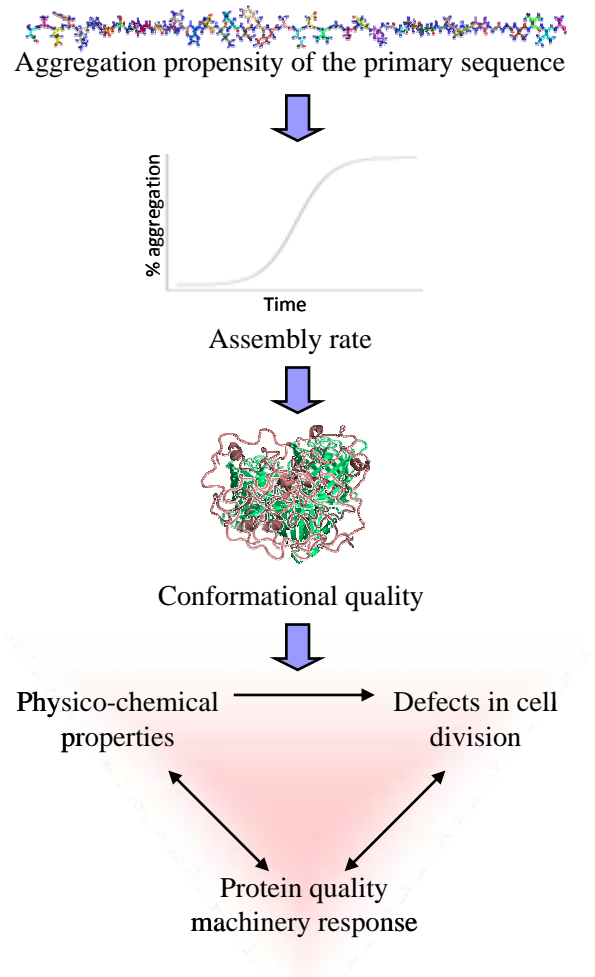


Fig. 4.1. The amino acid composition of polypeptides acts as a strong intrinsic determinant of deposition. The aggregation propensity of a primary sequence modulates the *in vivo* self-assembly rate. High aggregation prone sequences tend to rapidly aggregate and, thus, the formed deposit exhibit a low conformational quality, which is the reflection of the relation between native intramolecular contacts and non-native intermolecular ones. The conformational quality of an intracellular aggregate is a key factor determining its physico-chemical properties and modulating its cytotoxic effects on cell division and the activity of the cellular protein quality machinery.

Conclusions

Publication I: The effect of amyloidogenic peptides on bacterial aging correlates with their intrinsic aggregation propensity.

- Protein aggregation propensity determines IBs formation rate in the intracellular space.
- Protein aggregation propensity correlates with the IBs conformational quality and, thus, with their physico-chemical properties, since a higher proportion of aberrant intermolecular contacts provides resistance and hardness to the aggregate.
- Amyloid IBs impair the binary fission of the host cell, evidencing that protein deposition promotes an aging phenotype in bacteria, which correlates with the aggregation propensity of the protein assembled.
- Molecular chaperones increase the conformational quality of the IBs and alleviate the aging effect. The activity of this protein quality machinery is dependent on the aggregation propensity of the expressed polypeptide.

Publication II: Using bacterial inclusion bodies to screen for amyloid aggregation inhibitors.

- The *in vivo* competition between folding and aggregation during IB formation can be reproduced *in vitro* in a refolding buffer.
- A screening assay for amyloid aggregation modulators is developed. The methodology requires two consecutive steps: i) denaturation of the IBs and ii) refolding of the protein *in vitro*.
- The method detects the amyloid enhancer activity of previously described metal cations in the range of physiological concentrations. It is also sensible to the presence of inhibitor compounds of metal-promoted aggregation.
- The assay results cost-effective and allows high-throughput application, enabling its putative use in the biotechnology industry.

Publication III: Linking protein aggregation propensity and clearance in yeast.

- Aggregation of amyloidogenic proteins in the intracellular space of yeast resembles to that in bacteria, strengthening the importance of the intrinsic sequential determinants of protein deposition.

- Protein deposits in the yeast cytoplasm are cleared by distinct proteolytic pathways, including autophagy, vacuolar proteases and proteasome.
- The activity of the cellular proteolysis machinery correlates with the aggregation propensity of the protein embedded in the deposits.

Publication IV: Amyloid-like protein inclusions in tobacco transgenic plants.

- Maize transglutaminase (TGZ) possesses a highly amyloidogenic region within its sequence and forms amyloid-like aggregates when assembled *in vitro*.
- Tobacco transplastomic plants expressing TGZ display a pathogenic leaf phenotype.
- Recombinant TGZ is found in insoluble deposits in the chloroplasts of the host tobacco plants. These protein aggregates exhibit amyloid-like properties.
- This report represents the first evidence that a plant protein can self-assemble into amyloid-like deposits in a vegetable transgenic organism.

Publications V & VI: Modeling amyloids in bacteria. & Inclusion bodies in the study of amyloid aggregation.

- Protein overexpression in bacteria commonly results in the formation of IBs constituted mainly by the recombinant protein.
- Despite the first assumption that these deposits were unstructured, an increasing body of studies evidence that they are glued by sequential specific contacts and contain a wide range of conformations, including native-like species and cross- β structure, which provide them with a certain degree of activity and amyloid-like properties.
- The amyloid-like nature of IBs has been demonstrated by a battery of complementary techniques and has enabled their use as model in the study of amyloid aggregation.
- During IBs formation, there exists a kinetic competition between aggregation and folding.
- Aggregation in IBs highly resembles to the formation of amyloid fibrils linked to human diseases, including the establishment of aberrant intermolecular contacts, similar intrinsic and external determinants and sequence specificity.
- Bacterial IBs can be used as model for toxicity associated to protein deposition: loss in cell fitness, amyloid-triggered toxicity and prion-like infectivity.
- Bacterial IBs can be implemented in screening assays for amyloid aggregation inhibitor compounds.

References

- [1] F. Crick. Central dogma of molecular biology. *Nature*, 227(5258):561–563, 1970.
- [2] C. B. Anfinsen. Principles that govern the folding of protein chains. *Science*, 181(4096):223–230, 1973.
- [3] K. A. Dill and H. S. Chan. From Levinthal to pathways to funnels. *Nat. Struct. Biol.*, 4(1):10–19, 1997.
- [4] E. Herczenik and M. F. Gebbink. Molecular and cellular aspects of protein misfolding and disease. *FASEB J.*, 22(7):2115–2133, 2008.
- [5] F. U. Hartl and M. Hayer-Hartl. Converging concepts of protein folding in vitro and in vivo. *Nat. Struct. Mol. Biol.*, 16(6):574–581, 2009.
- [6] F. U. Hartl, A. Bracher, and M. Hayer-Hartl. Molecular chaperones in protein folding and proteostasis. *Nature*, 475(7356):324–332, 2011.
- [7] T. R. Jahn and S. E. Radford. Folding versus aggregation: polypeptide conformations on competing pathways. *Arch. Biochem. Biophys.*, 469(1):100–117, 2008.
- [8] H. E. Smith. The transcriptional response of *Escherichia coli* to recombinant protein insolubility. *J. Struct. Funct. Genomics*, 8(1):27–35, 2007.
- [9] R. Sabate, N. S. de Groot, and S. Ventura. Protein folding and aggregation in bacteria. *Cell. Mol. Life Sci.*, 67(16):2695–2715, 2010.
- [10] J. Tyedmers, A. Mogk, and B. Bukau. Cellular strategies for controlling protein aggregation. *Nat. Rev. Mol. Cell Biol.*, 11(11):777–788, 2010.
- [11] P. Tompa. Intrinsically disordered proteins: a 10-year recap. *Trends Biochem. Sci.*, 37(12):509–516, 2012.
- [12] P. Radivojac, L. M. Iakoucheva, C. J. Oldfield, Z. Obradovic, V. N. Uversky, and A. K. Dunker. Intrinsic disorder and functional proteomics. *Biophys. J.*, 92(5):1439–1456, 2007.
- [13] V. N. Uversky. Intrinsically disordered proteins and novel strategies for drug discovery. *Expert Opin. Drug Discov.*, 7(6):475–488, 2012.
- [14] A. K. Dunker, I. Silman, V. N. Uversky, and J. L. Sussman. Function and structure of inherently disordered proteins. *Curr. Opin. Struct. Biol.*, 18(6):756–764, 2008.
- [15] F. Chiti and C. M. Dobson. Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.*, 75:333–366, 2006.

- [16] M. M. Babu, R. van der Lee, N. S. de Groot, and J. Gsponer. Intrinsically disordered proteins: regulation and disease. *Curr. Opin. Struct. Biol.*, 21(3):432–440, 2011.
- [17] R. J. Ellis and A. P. Minton. Cell biology: join the crowd. *Nature*, 425(6953):27–28, 2003.
- [18] F. Chiti, M. Stefani, N. Taddei, G. Ramponi, and C. M. Dobson. Rationalization of the effects of mutations on peptide and protein aggregation rates. *Nature*, 424(6950):805–808, 2003.
- [19] K. F. DuBay, A. P. Pawar, F. Chiti, J. Zurdo, C. M. Dobson, and M. Vendruscolo. Prediction of the absolute aggregation rates of amyloidogenic polypeptide chains. *J. Mol. Biol.*, 341(5):1317–1326, 2004.
- [20] V. Vetri, F. Librizzi, M. Leone, and V. Militello. Thermal aggregation of bovine serum albumin at different pH: comparison with human serum albumin. *Eur. Biophys. J.*, 36(7):717–725, 2007.
- [21] M. Lopez de la Paz and L. Serrano. Sequence determinants of amyloid fibril formation. *Proc. Natl. Acad. Sci. U.S.A.*, 101(1):87–92, 2004.
- [22] S. Ventura, J. Zurdo, S. Narayanan, M. Parreno, R. Mangués, B. Reif, F. Chiti, E. Giannoni, C. M. Dobson, F. X. Aviles, and L. Serrano. Short amino acid stretches can mediate amyloid formation in globular proteins: the Src homology 3 (SH3) case. *Proc. Natl. Acad. Sci. U.S.A.*, 101(19):7258–7263, 2004.
- [23] N. Sanchez de Groot, I. Pallares, F. X. Aviles, J. Vendrell, and S. Ventura. Prediction of "hot spots" of aggregation in disease-linked polypeptides. *BMC Struct. Biol.*, 5:18, 2005.
- [24] E. Monsellier, M. Ramazzotti, N. Taddei, and F. Chiti. Aggregation propensity of the human proteome. *PLoS Comput. Biol.*, 4(10):e1000199, 2008.
- [25] V. Castillo and S. Ventura. Amyloidogenic regions and interaction surfaces overlap in globular proteins related to conformational diseases. *PLoS Comput. Biol.*, 5(8):e1000476, 2009.
- [26] A. Espargaro, V. Castillo, N. S. de Groot, and S. Ventura. The in vivo and in vitro aggregation properties of globular proteins correlate with their conformational stability: the SH3 case. *J. Mol. Biol.*, 378(5):1116–1131, 2008.
- [27] K. H. Lim, H. J. Dyson, J. W. Kelly, and P. E. Wright. Localized Structural Fluctuations Promote Amyloidogenic Conformations in Transthyretin. *J. Mol. Biol.*, 2013.
- [28] J. Reumers, S. Maurer-Stroh, J. Schymkowitz, and F. Rousseau. Protein sequences encode safeguards against aggregation. *Hum. Mutat.*, 30(3):431–437, 2009.
- [29] J. Beerten, W. Jonckheere, S. Rudyak, J. Xu, H. Wilkinson, F. De Smet, J. Schymkowitz, and F. Rousseau. Aggregation gatekeepers modulate protein homeostasis of aggregating sequences and affect bacterial fitness. *Protein Eng. Des. Sel.*, 25(7):357–366, 2012.
- [30] F. Rousseau, L. Serrano, and J. W. Schymkowitz. How evolutionary pressure against protein aggregation shaped chaperone specificity. *J. Mol. Biol.*, 355(5):1037–1047, 2006.
- [31] A. M. Szalkowski and M. Anisimova. Markov models of amino acid substitution to study proteins with intrinsically disordered regions. *PLoS ONE*, 6(5):e20488, 2011.

- [32] A. De Simone, C. Kitchen, A. H. Kwan, M. Sunde, C. M. Dobson, and D. Frenkel. Intrinsic disorder modulates protein self-assembly and aggregation. *Proc. Natl. Acad. Sci. U.S.A.*, 109(18):6951–6956, 2012.
- [33] D. Guidolin, L. F. Agnati, G. Albertin, C. Tortorella, and K. Fuxe. Bioinformatics aggregation predictors in the study of protein conformational diseases of the human nervous system. *Electrophoresis*, 33(24):3669–3679, 2012.
- [34] O. Conchillo-Sole, N. S. de Groot, F. X. Aviles, J. Vendrell, X. Daura, and S. Ventura. AGGRESKAN: a server for the prediction and evaluation of "hot spots" of aggregation in polypeptides. *BMC Bioinformatics*, 8:65, 2007.
- [35] A. M. Fernandez-Escamilla, F. Rousseau, J. Schymkowitz, and L. Serrano. Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat. Biotechnol.*, 22(10):1302–1306, 2004.
- [36] S. Maurer-Stroh, M. Debulpaep, N. Kueemmerer, M. Lopez de la Paz, I. C. Martins, J. Reumers, K. L. Morris, A. Copland, L. Serpell, L. Serrano, J. W. Schymkowitz, and F. Rousseau. Exploring the sequence determinants of amyloid structure using position-specific scoring matrices. *Nat. Methods*, 7(3):237–242, 2010.
- [37] S. Zibae, O. S. Makin, M. Goedert, and L. C. Serpell. A simple algorithm locates beta-strands in the amyloid fibril core of alpha-synuclein, Abeta, and tau using the amino acid sequence alone. *Protein Sci.*, 16(5):906–918, 2007.
- [38] G. G. Tartaglia and M. Vendruscolo. The Zyggregator method for predicting protein aggregation propensities. *Chem. Soc. Rev.*, 37(7):1395–1401, 2008.
- [39] S. O. Garbuzynskiy, M. Y. Lobanov, and O. V. Galzitskaya. FoldAmyloid: a method of prediction of amyloidogenic regions from protein sequence. *Bioinformatics*, 26(3):326–332, 2010.
- [40] A. W. Bryan, M. Menke, L. J. Cowen, S. L. Lindquist, and B. Berger. BETASCAN: probable beta-amyloids identified by pairwise probabilistic analysis. *PLoS Comput. Biol.*, 5(3):e1000333, 2009.
- [41] V. Castillo, R. Grana-Montes, R. Sabate, and S. Ventura. Prediction of the aggregation propensity of proteins from the primary sequence: aggregation properties of proteomes. *Biotechnol. J.*, 6(6):674–685, 2011.
- [42] M. Belli, M. Ramazzotti, and F. Chiti. Prediction of amyloid aggregation in vivo. *EMBO Rep.*, 12(7):657–663, 2011.
- [43] S. J. Hamodrakas. Protein aggregation and amyloid fibril formation prediction software from primary sequence: towards controlling the formation of bacterial inclusion bodies. *FEBS J.*, 278(14):2428–2435, 2011.
- [44] A. C. Tsois, N. C. Papandreou, V. A. Iconomidou, and S. J. Hamodrakas. A consensus method for the prediction of 'aggregation-prone' peptides in globular proteins. *PLoS ONE*, 8(1):e54175, 2013.
- [45] E. Monsellier and F. Chiti. Prevention of amyloid-like aggregation as a driving force of protein evolution. *EMBO Rep.*, 8(8):737–742, 2007.
- [46] C. Debes, M. Wang, G. Caetano-Anolles, and F. Gräter. Evolutionary optimization of protein folding. *PLoS Comput. Biol.*, 9(1):e1002861, 2013.
- [47] D. Fass. Disulfide bonding in protein biophysics. *Annu. Rev. Biophys.*, 41:63–79, 2012.
- [48] R. Grana-Montes, N. S. de Groot, V. Castillo, J. Sancho, A. Velazquez-Campoy, and S. Ventura. Contribution of disulfide bonds to stability, folding, and amyloid fibril formation: the PI3-SH3 domain case. *Antioxid. Redox Signal.*, 16(1):1–15, 2012.

- [49] A. L. Watters, P. Deka, C. Corrent, D. Callender, G. Varani, T. Sosnick, and D. Baker. The highly cooperative folding of small naturally occurring proteins is likely the result of natural selection. *Cell*, 128(3):613–624, 2007.
- [50] S. Rauscher, S. Baud, M. Miao, F. W. Keeley, and R. Pomes. Proline and glycine control protein self-organization into elastomeric or amyloid fibrils. *Structure*, 14(11):1667–1676, 2006.
- [51] P. Jiang, W. Xu, and Y. Mu. Amyloidogenesis abolished by proline substitutions but enhanced by lipid binding. *PLoS Comput. Biol.*, 5(4):e1000357, 2009.
- [52] C. Parrini, N. Taddei, M. Ramazzotti, D. Degl’Innocenti, G. Ramponi, C. M. Dobson, and F. Chiti. Glycine residues appear to be evolutionarily conserved for their ability to inhibit aggregation. *Structure*, 13(8):1143–1151, 2005.
- [53] G. G. Tartaglia, S. Pechmann, C. M. Dobson, and M. Vendruscolo. Life on the edge: a link between gene expression levels and aggregation rates of human proteins. *Trends Biochem. Sci.*, 32(5):204–206, 2007.
- [54] G. G. Tartaglia, S. Pechmann, C. M. Dobson, and M. Vendruscolo. A relationship between mRNA expression levels and protein solubility in *E. coli*. *J. Mol. Biol.*, 388(2):381–389, 2009.
- [55] V. Castillo, R. Grana-Montes, and S. Ventura. The aggregation properties of *Escherichia coli* proteins associated with their cellular abundance. *Biotechnol. J.*, 6(6):752–760, 2011.
- [56] G. De Baets, J. Reumers, J. Delgado Blanco, J. Dopazo, J. Schymkowitz, and F. Rousseau. An evolutionary trade-off between protein turnover rate and protein aggregation favors a higher aggregation propensity in fast degrading proteins. *PLoS Comput. Biol.*, 7(6):e1002090, 2011.
- [57] G. B. Irvine, O. M. El-Agnaf, G. M. Shankar, and D. M. Walsh. Protein aggregation in the brain: the molecular basis for Alzheimer’s and Parkinson’s diseases. *Mol. Med.*, 14(7-8):451–464, 2008.
- [58] L. C. Serpell, M. Sunde, M. D. Benson, G. A. Tennent, M. B. Pepys, and P. E. Fraser. The protofilament substructure of amyloid fibrils. *J. Mol. Biol.*, 300(5):1033–1039, 2000.
- [59] D. Eisenberg and M. Jucker. The amyloid state of proteins in human diseases. *Cell*, 148(6):1188–1203, 2012.
- [60] T. Stromer and L. C. Serpell. Structure and morphology of the Alzheimer’s amyloid fibril. *Microsc. Res. Tech.*, 67(3-4):210–217, 2005.
- [61] G. Bhak, Y. J. Choe, and S. R. Paik. Mechanism of amyloidogenesis: nucleation-dependent fibrillation versus double-concerted fibrillation. *BMB Rep.*, 42(9):541–551, 2009.
- [62] G. Invernizzi, E. Papaleo, R. Sabate, and S. Ventura. Protein aggregation: mechanisms and functional consequences. *Int. J. Biochem. Cell Biol.*, 44(9):1541–1554, 2012.
- [63] N. Carulla, G. L. Caddy, D. R. Hall, J. Zurdo, M. Gairi, M. Feliz, E. Giralt, C. V. Robinson, and C. M. Dobson. Molecular recycling within amyloid fibrils. *Nature*, 436(7050):554–558, 2005.
- [64] S. Mankar, A. Anoop, S. Sen, and S. K. Maji. Nanomaterials: amyloids reflect their brighter side. *Nano. Rev.*, 2, 2011.
- [65] A. Dusa, J. Kaylor, S. Edridge, N. Bodner, D. P. Hong, and A. L. Fink. Characterization of oligomers during alpha-synuclein aggregation using intrinsic tryptophan fluorescence. *Biochemistry*, 45(8):2752–2760, 2006.

- [66] M. Fandrich. Oligomeric intermediates in amyloid formation: structure determination and mechanisms of toxicity. *J. Mol. Biol.*, 421(4-5):427–440, 2012.
- [67] J. Brouillette, R. Caillierez, N. Zommer, C. Alves-Pires, I. Benilova, D. Blum, B. De Strooper, and L. Buee. Neurotoxicity and memory deficits induced by soluble low-molecular-weight amyloid- β 1-42 oligomers are revealed in vivo by using a novel animal model. *J. Neurosci.*, 32(23):7852–7861, 2012.
- [68] M. J. Diogenes, R. B. Dias, D. M. Rombo, H. Vicente Miranda, F. Maiolino, P. Guerreiro, T. Nasstrom, H. G. Franquelim, L. M. Oliveira, M. A. Castanho, L. Lannfelt, J. Bergstrom, M. Ingelsson, A. Quintas, A. M. Sebastiao, L. V. Lopes, and T. F. Outeiro. Extracellular alpha-synuclein oligomers modulate synaptic transmission and impair LTP via NMDA-receptor activation. *J. Neurosci.*, 32(34):11750–11762, 2012.
- [69] F. Bemporad and F. Chiti. Protein misfolded oligomers: experimental approaches, mechanism of formation, and structure-toxicity relationships. *Chem. Biol.*, 19(3):315–327, 2012.
- [70] H. Jang, L. Connelly, F. T. Arce, S. Ramachandran, B. L. Kagan, R. Lal, and R. Nussinov. Mechanisms for the Insertion of Toxic, Fibril-like β -Amyloid Oligomers into the Membrane. *J. Chem. Theory. Comput.*, 9(1):822–833, 2013.
- [71] M. M. Carrio, J. L. Corchero, and A. Villaverde. Dynamics of in vivo protein aggregation: building inclusion bodies in recombinant bacteria. *FEMS Microbiol. Lett.*, 169(1):9–15, 1998.
- [72] L. Wang. Towards revealing the structure of bacterial inclusion bodies. *Prion*, 3(3):139–145, 2009.
- [73] M. Carrio, N. Gonzalez-Montalban, A. Vera, A. Villaverde, and S. Ventura. Amyloid-like properties of bacterial inclusion bodies. *J. Mol. Biol.*, 347(5):1025–1037, 2005.
- [74] B. Ono, M. Kubota, H. Kimiduka, H. Kawaminami, T. Ueto, S. Yokosawa, M. Iseda, Y. Yamamoto, Y. Murakami, and S. Yokota. Production of a polymer-forming fusion protein in *Escherichia coli* strain BL21. *Biosci. Biotechnol. Biochem.*, 70(12):2813–2823, 2006.
- [75] L. Wang, S. K. Maji, M. R. Sawaya, D. Eisenberg, and R. Riek. Bacterial inclusion bodies contain amyloid-like structure. *PLoS Biol.*, 6(8):e195, 2008.
- [76] S. M. Doglia, D. Ami, A. Natalello, P. Gatti-Lafranconi, and M. Lotti. Fourier transform infrared spectroscopy analysis of the conformational quality of recombinant proteins within inclusion bodies. *Biotechnol. J.*, 3(2):193–201, 2008.
- [77] M. Morell, R. Bravo, A. Espargaro, X. Sisquella, F. X. Aviles, X. Fernandez-Busquets, and S. Ventura. Inclusion bodies: specificity in their aggregation process and amyloid-like structure. *Biochim. Biophys. Acta*, 1783(10):1815–1825, 2008.
- [78] N. S. de Groot, R. Sabate, and S. Ventura. Amyloids in bacterial inclusion bodies. *Trends Biochem. Sci.*, 34(8):408–416, 2009.
- [79] A. Rokney, M. Shagan, M. Kessel, Y. Smith, I. Rosenshine, and A. B. Oppenheim. *E. coli* transports aggregated proteins to the poles by a specific and energy-dependent process. *J. Mol. Biol.*, 392(3):589–601, 2009.
- [80] J. Winkler, A. Seybert, L. Konig, S. Pruggnaller, U. Haselmann, V. Sourjik, M. Weiss, A. S. Frangakis, A. Mogk, and B. Bukau. Quantitative and spatiotemporal features of protein aggregation in *Escherichia coli* and consequences on protein quality control and cellular ageing. *EMBO J.*, 29(5):910–923, 2010.

- [81] R. A. Hart, U. Rinas, and J. E. Bailey. Protein composition of Vitreoscilla hemoglobin inclusion bodies produced in *Escherichia coli*. *J. Biol. Chem.*, 265(21):12728–12733, 1990.
- [82] A. B. Lindner, R. Madden, A. Demarez, E. J. Stewart, and F. Taddei. Asymmetric segregation of protein aggregates is associated with cellular aging and rejuvenation. *Proc. Natl. Acad. Sci. U.S.A.*, 105(8):3076–3081, 2008.
- [83] E. Garcia-Fruitos, N. Gonzalez-Montalban, M. Morell, A. Vera, R. M. Ferraz, A. Aris, S. Ventura, and A. Villaverde. Aggregation as bacterial inclusion bodies does not imply inactivation of enzymes and fluorescent proteins. *Microb. Cell Fact.*, 4:27, 2005.
- [84] N. S. de Groot and S. Ventura. Protein activity in bacterial inclusion bodies correlates with predicted aggregation rates. *J. Biotechnol.*, 125(1):110–113, 2006.
- [85] W. Wu, L. Xing, B. Zhou, and Z. Lin. Active protein aggregates induced by terminally attached self-assembling peptide ELK16 in *Escherichia coli*. *Microb. Cell Fact.*, 10:9, 2011.
- [86] S. Y. Park, S. H. Park, and S. K. Choi. Active inclusion body formation using *Paenibacillus polymyxa* PoxB as a fusion partner in *Escherichia coli*. *Anal. Biochem.*, 426(1):63–65, 2012.
- [87] S. Peternel and R. Komel. Active Protein Aggregates Produced in *Escherichia coli*. *Int. J. Mol. Sci.*, 12(11):8275–8287, 2011.
- [88] E. Vazquez, J. L. Corchero, J. F. Burgueno, J. Seras-Franzoso, A. Kosoy, R. Bossier, R. Mendoza, J. M. Martinez-Lainez, U. Rinas, E. Fernandez, L. Ruiz-Avila, E. Garcia-Fruitos, and A. Villaverde. Functional inclusion bodies produced in bacteria as naturally occurring nanopills for advanced cell therapies. *Adv. Mater. Weinheim*, 24(13):1742–1747, 2012.
- [89] K. Talafova, E. Hrabarova, D. A. Chorvat, and J. Nahalka. Bacterial inclusion bodies as potential synthetic devices for pathogen recognition and a therapeutic substance release. *Microb. Cell Fact.*, 12(1):16, 2013.
- [90] G. S. Waldo, B. M. Standish, J. Berendzen, and T. C. Terwilliger. Rapid protein-folding assay using green fluorescent protein. *Nat. Biotechnol.*, 17(7):691–695, 1999.
- [91] W. Kim, Y. Kim, J. Min, D. J. Kim, Y. T. Chang, and M. H. Hecht. A high-throughput screen for compounds that inhibit aggregation of the Alzheimer’s peptide. *ACS Chem. Biol.*, 1(7):461–469, 2006.
- [92] A. Espargaro, R. Sabate, and S. Ventura. Thioflavin-S staining coupled to flow cytometry. A screening tool to detect in vivo protein aggregation. *Mol. Biosyst.*, 8(11):2839–2844, 2012.
- [93] C. Ballard, S. Gauthier, A. Corbett, C. Brayne, D. Aarsland, and E. Jones. Alzheimer’s disease. *Lancet*, 377(9770):1019–1031, 2011.
- [94] D. P. Purohit, N. O. Batheja, M. Sano, K. D. Jashnani, R. N. Kalaria, A. Karunamurthy, S. Kaur, A. S. Shenoy, K. Van Dyk, J. Schmeidler, and D. P. Perl. Profiles of Alzheimer’s disease-related pathology in an aging urban population sample in India. *J. Alzheimers Dis.*, 24(1):187–196, 2011.
- [95] Q. Guo, Z. Wang, H. Li, M. Wiese, and H. Zheng. APP physiological and pathophysiological functions: insights from animal models. *Cell Res.*, 22(1):78–89, 2012.
- [96] K. H. Ashe and K. R. Zahs. Probing the biology of Alzheimer’s disease in mice. *Neuron*, 66(5):631–645, 2010.

- [97] N. Khandekar, K. H. Lie, P. S. Sachdev, and K. S. Sidhu. Amyloid precursor proteins, neural differentiation of pluripotent stem cells and its relevance to Alzheimer's disease. *Stem Cells Dev.*, 21(7):997–1006, 2012.
- [98] W. Kim and M. H. Hecht. Mutations enhance the aggregation propensity of the Alzheimer's A beta peptide. *J. Mol. Biol.*, 377(2):565–574, 2008.
- [99] S. Zhang, N. Casey, and J. P. Lee. Residual structure in the Alzheimer's disease peptide: probing the origin of a central hydrophobic cluster. *Fold. Des.*, 3(5):413–422, 1998.
- [100] W. P. Esler, E. R. Stimson, J. R. Ghilardi, Y. A. Lu, A. M. Felix, H. V. Vinters, P. W. Mantyh, J. P. Lee, and J. E. Maggio. Point substitution in the central hydrophobic cluster of a human beta-amyloid congener disrupts peptide folding and abolishes plaque competence. *Biochemistry*, 35(44):13914–13921, 1996.
- [101] N. S. de Groot, F. X. Aviles, J. Vendrell, and S. Ventura. Mutagenesis of the central hydrophobic cluster in Abeta42 Alzheimer's peptide. Side-chain properties correlate with aggregation propensities. *FEBS J.*, 273(3):658–668, 2006.
- [102] S. N. Cohen, A. C. Chang, H. W. Boyer, and R. B. Helling. Construction of biologically functional bacterial plasmids in vitro. *Proc. Natl. Acad. Sci. U.S.A.*, 70(11):3240–3244, 1973.
- [103] J. F. Morrow, S. N. Cohen, A. C. Chang, H. W. Boyer, H. M. Goodman, and R. B. Helling. Replication and transcription of eukaryotic DNA in Escherichia coli. *Proc. Natl. Acad. Sci. U.S.A.*, 71(5):1743–1747, 1974.
- [104] F. R. Schmidt. Recombinant expression systems in the pharmaceutical industry. *Appl. Microbiol. Biotechnol.*, 65(4):363–372, 2004.
- [105] K. Terpe. Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biotechnol.*, 72(2):211–222, 2006.
- [106] E. Garcia-Fruitos, R. Sabate, N. S. de Groot, A. Villaverde, and S. Ventura. Biological role of bacterial inclusion bodies: a model for amyloid aggregation. *FEBS J.*, 278(14):2419–2427, 2011.
- [107] M. E. Fernandez-Tresguerres, S. M. de la Espina, F. Gasset-Rosa, and R. Giraldo. A DNA-promoted amyloid proteinopathy in Escherichia coli. *Mol. Microbiol.*, 77(6):1456–1469, 2010.
- [108] D. Porro, B. Gasser, T. Fossati, M. Maurer, P. Branduardi, M. Sauer, and D. Mattanovich. Production of recombinant proteins and metabolites in yeasts: when are these systems better than bacterial production systems? *Appl. Microbiol. Biotechnol.*, 89(4):939–948, 2011.
- [109] S. Tenreiro and T. F. Outeiro. Simple is good: yeast models of neurodegeneration. *FEMS Yeast Res.*, 10(8):970–979, 2010.
- [110] C. Pereira, C. Bessa, J. Soares, M. Leao, and L. Saraiva. Contribution of yeast models to neurodegeneration research. *J. Biomed. Biotechnol.*, 2012:941232, 2012.
- [111] V. Khurana and S. Lindquist. Modelling neurodegeneration in *Saccharomyces cerevisiae*: why cook with baker's yeast? *Nat. Rev. Neurosci.*, 11(6):436–449, 2010.
- [112] F. D'Angelo, H. Vignaud, J. Di Martino, B. Salin, A. Devin, C. Cullin, and C. Marchal. A yeast model for amyloid- β aggregation exemplifies the role

- of membrane trafficking and PICALM in cytotoxicity. *Dis. Model. Mech.*, 6(1):206–216, 2013.
- [113] R. Boehm. Bioproduction of therapeutic proteins in the 21st century and the role of plants and plant cells as production platforms. *Ann. N. Y. Acad. Sci.*, 1102:121–134, 2007.
- [114] J. Xu, M. C. Dolan, G. Medrano, C. L. Cramer, and P. J. Weathers. Green factory: plants as bioproduction platforms for recombinant proteins. *Biotechnol. Adv.*, 30(5):1171–1184, 2012.
- [115] L. R. Wilken and Z. L. Nikolov. Recovery and purification of plant-made recombinant proteins. *Biotechnol. Adv.*, 30(2):419–433, 2012.
- [116] L. Crews, E. Rockenstein, and E. Masliah. APP transgenic modeling of Alzheimer’s disease: mechanisms of neurodegeneration and aberrant neurogenesis. *Brain Struct Funct*, 214(2-3):111–126, 2010.
- [117] J. D. Wadsworth, E. A. Asante, and J. Collinge. Review: contribution of transgenic models to understanding human prion disease. *Neuropathol. Appl. Neurobiol.*, 36(7):576–597, 2010.
- [118] J. Stohr, N. Weinmann, H. Wille, T. Kaimann, L. Nagel-Steger, E. Birkmann, G. Panza, S. B. Prusiner, M. Eigen, and D. Riesner. Mechanisms of prion protein assembly into amyloid. *Proc. Natl. Acad. Sci. U.S.A.*, 105(7):2409–2414, 2008.
- [119] M. Fandrich, M. Schmidt, and N. Grigorieff. Recent progress in understanding Alzheimer’s β -amyloid structures. *Trends Biochem. Sci.*, 36(6):338–345, 2011.
- [120] M. Morell, N. S. de Groot, J. Vendrell, F. X. Aviles, and S. Ventura. Linking amyloid protein aggregation and yeast survival. *Mol. Biosyst.*, 7(4):1121–1128, 2011.
- [121] K. A. Geiler-Samerotte, M. F. Dion, B. A. Budnik, S. M. Wang, D. L. Hartl, and D. A. Drummond. Misfolded proteins impose a dosage-dependent fitness cost and trigger a cytosolic unfolded protein response in yeast. *Proc. Natl. Acad. Sci. U.S.A.*, 108(2):680–685, 2011.
- [122] M. M. Carrio and A. Villaverde. Localization of chaperones DnaK and GroEL in bacterial inclusion bodies. *J. Bacteriol.*, 187(10):3599–3601, 2005.
- [123] S. Sugimoto, K. Saruwatari, C. Higashi, and K. Sonomoto. The proper ratio of GrpE to DnaK is important for protein quality control by the DnaK-DnaJ-GrpE chaperone system and for cell division. *Microbiology*, 154(Pt 7):1876–1885, 2008.
- [124] S. Ventura and A. Villaverde. Protein quality in bacterial inclusion bodies. *Trends Biotechnol.*, 24(4):179–185, 2006.
- [125] A. de Marco, E. Deuerling, A. Mogk, T. Tomoyasu, and B. Bukau. Chaperone-based procedure to increase yields of soluble recombinant proteins produced in *E. coli*. *BMC Biotechnol.*, 7:32, 2007.
- [126] N. Tokuriki and D. S. Tawfik. Chaperonin overexpression promotes genetic variation and enzyme evolution. *Nature*, 459(7247):668–673, 2009.
- [127] T. A. Williams and M. A. Fares. The effect of chaperonin buffering on protein evolution. *Genome Biol. Evol.*, 2:609–619, 2010.
- [128] D. Kaganovich, R. Kopito, and J. Frydman. Misfolded proteins partition between two distinct quality control compartments. *Nature*, 454(7208):1088–1095, 2008.
- [129] D. Petroi, B. Popova, N. Taheri-Talesh, S. Irniger, H. Shahpasandzadeh, M. Zweckstetter, T. F. Outeiro, and G. H. Braus. Aggregate clearance of

- α -synuclein in *Saccharomyces cerevisiae* depends more on autophagosome and vacuole function than on the proteasome. *J. Biol. Chem.*, 287(33):27567–27579, 2012.
- [130] A. I. Bush and R. E. Tanzi. Therapeutics for Alzheimer’s disease based on the metal hypothesis. *Neurotherapeutics*, 5(3):421–432, 2008.
- [131] F. Chiti and C. M. Dobson. Amyloid formation by globular proteins under native conditions. *Nat. Chem. Biol.*, 5(1):15–22, 2009.
- [132] R. B. Wickner, H. K. Edskes, F. Shewmaker, and T. Nakayashiki. Prions of fungi: inherited structures and biological roles. *Nat. Rev. Microbiol.*, 5(8):611–618, 2007.
- [133] N. Gonzalez-Montalban, A. Villaverde, and A. Aris. Amyloid-linked cellular toxicity triggered by bacterial inclusion bodies. *Biochem. Biophys. Res. Commun.*, 355(3):637–642, 2007.
- [134] M. Dasari, A. Espargaro, R. Sabate, J. M. Lopez del Amo, U. Fink, G. Grelle, J. Bieschke, S. Ventura, and B. Reif. Bacterial inclusion bodies of Alzheimer’s disease β -amyloid peptides can be employed to study native-like aggregation intermediate states. *Chembiochem.*, 12(3):407–423, 2011.

Agraïments (Acknowledgments)

Aquesta tesi ha tingut molts protagonistes als quals vull agrair la seva participació:

Al Salva, el meu director, del qual he rebut un magnífic suport i seguiment. Gràcies per haver-me fet treballar de valent i per tantíssimes coses que he après tots aquests anys. Però sobretot gràcies per haver-me fet confiança i haver-me traspasat l'entusiasme per aquesta feina.

Al Pep, per oferir-me la beca PIF, així com també un lloc al seu laboratori.

Al Raimon, pel seu suport i la seva ajuda, especialment durant els primers temps.

A la Natàlia, per tot el que m'ha ensenyat i per haver estat la meva *padrina científica*. Perquè abans de començar, em va convèncer que això valia la pena i podia ser divertit. Per l'ajuda constant, tant físicament com *virtual*. Però més enllà, vull agrair-li la seva gran comprensió i el seu recolzament a nivell personal.

Al Madan, per l'estada a l'LMB de Cambridge, un lloc excepcional i fascinant. Especialment, vull agrair-li el seu tracte càlid i personal. Tampoc puc oblidar-me de la Natàlia (novament) i del Marc, que em van acollir afectuosament i van *aguantar* més d'alguna plorera...

Als meus companys de grup i a les meves companyes de laboratori. Per la convivència diària, i per compartir infinitat d'experiències i moments, des dels més distesos fins als més *intel·lectuals*, tots ells m'han enriquit enormement.

Als meus companys de departament. Moltes gràcies per tots el esmorzars i dinars compartits. Gràcies per fer de la torre de bioquímica un lloc tan familiar i agradable que les hores passen i passen i al cap del dia has creuat un bon grapat de somriures i de petites entrançables converses. És realment meravellós gaudir d'aquest ambient a la feina.

Al Salva, l'Helena, el Santi i la Magda, per la feina silenciosa però ben feta, present en cada un dels experiments d'aquesta tesi.

Als meus amics, per ser un petit oasi de felicitat. Gràcies per tot l'afecte, per fer-m'ho passar bé, i per ensenyar-me a relativitzar les coses de la feina. Especialment vull donar les gràcies a la Paula i la Sara, per una amistat autèntica, supervivent a gairebé 20 anys i també a molts kilòmetres de distància. També a l'Alba, per ser tot el que es pot esperar d'un amic. Per passar de companya de pràctiques i treballs, a compartir sopars, festes, vacances... però també els moments més durs, on sempre m'he sentit acompanyada.

I finalment a la meva petita gran família: lo papa, la mama, lo Roger, el Gerard, la Cristina, lo Franc i la petitona Laia, així com també als que ja no hi són. Gràcies perquè aquesta tesi ha tingut moments difícils i molt difícils i, malgrat tot, heu tingut una paciència infinita i ganes d'animar-me per seguir endavant. I sobretot, perquè m'heu ensenyat que ser doctora és important però no té cap sentit si abans no s'és persona. Gràcies per aportar-me pau i tranquil·litat i pel recolzament constant, moltes gràcies de tot cor. De forma molt especial vull agrair tot això a la Padrina: vostè va veure començar aquesta aventura i va saber que arribaria aquest dia, fins i tot molt abans de que jo ho sabés. Gràcies pels seus consells i la seva estimació, sempre presents en mi.

Appendices

Linking intracellular protein aggregation propensity and protein quality control degradation

Anna Villar-Piqué^{1,2} and Salvador Ventura^{1,2§}

1 Departament de Bioquímica i Biologia Molecular, Facultat de Biociències, Universitat Autònoma de Barcelona, E-08193 Bellaterra (Spain).

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

§Corresponding author

SV: salvador.ventura@uab.es

Running title: Aggregation and clearance of A β in yeast

Abstract

Protein deposition is linked to many pathological conditions, including several neurodegenerative diseases. An increasing body of evidences indicate that aggregation into amyloid-like structures is a general property of polypeptides encoded in the primary sequence in both eukaryotic and prokaryotic organisms. Using a large set of amyloid β peptide (A β 42) mutants, we have shown previously that, in bacteria, it exists a strict correlation between the sequential aggregation propensity and the intracellular deposition of these proteins. In the present work, we have analyzed the aggregation behavior of the same A β 42 variants when expressed in yeast. We demonstrate here that in this eukaryotic organism the capacity of the cellular proteolytic machinery to degrade these aggregation-prone peptides is significantly correlated with their intrinsic aggregation propensity, thus providing a quantitative evaluation of the ability of yeast cells to clear dangerous protein aggregates.

Key Words

Protein aggregation, amyloid peptide, fluorescent reporter, aggregation propensity, aggregates clearance, yeast.

Introduction

Protein aggregation and the long term cellular persistence of aggregates are pathological hallmarks of a large set of human disorders, the so-called conformational diseases, such as Alzheimer's (AD), Huntington's (HD) and Parkinson's (PD) diseases, diabetes type II or transmissible spongiform encephalopathies^{1; 2}. Moreover, protein deposition is also a common phenomenon during recombinant expression in simple organisms such as unicellular fungi or bacteria^{3; 4}. Interestingly, the aggregates formed in these microorganisms resemble those involved in the onset of the aforementioned disorders^{5; 6; 7}, indicating that protein self-assembly into β -sheet enriched amyloid-like structures is a generic process in direct competition with native protein folding^{8; 9; 10}, regardless of the considered host. Indeed, it has been shown that the large majority of polypeptides display at least one and often multiple aggregation-prone stretches, usually involved in the formation of the hydrophobic core^{11; 12}. As it occurs with folding, the aggregation propensity of proteins represents an essential property of the behaviour of polypeptides encoded in their primary sequences^{13; 14; 15}. Different bioinformatic algorithms have exploited this feature to predict protein deposition and amyloid formation by identifying and quantifying the aggregation-promoting regions within a given sequence^{16; 17; 18; 19}. Despite their success in predicting aggregation *in vitro*, the challenge is asserting if the predictive power of these programs also applies in a more complex biological context¹⁶. Addressing this question is not trivial, since quantitative evaluation of the aggregation propensities of polypeptides *in vivo* is technically complex. In a first step toward this direction, we exploited in previous works the competition between protein folding and aggregation in the bacterial cytosol to approximate intracellular aggregation rates. Essentially, we generated 19 variants of the A β peptide (A β 42) in which the original Phe residue in position 19 was substituted by the rest of natural proteinogenic amino acids. This residue is located in the central hydrophobic cluster (CHC), a short stretch comprising residues from Leu17 to Ala21 that controls, to a large extent, the aggregation propensity of the entire sequence. The 20 A β 42 mutants were fused to GFP, which acts as a reporter of the aggregation propensity of the A β moiety and were recombinantly expressed in *Escherichia coli*. The expression of the different peptide variants in *E. coli* resulted in the formation of cytoplasmic amyloid-like inclusion bodies whose GFP activity reflected the aggregation rate of the corresponding mutant^{20; 21}. These data constitute the core of AGGRESCAN, a method

implemented in our group to predict protein aggregation propensity^{22; 23}. We have shown recently that in this model system the cellular fitness cost induced by protein deposition is tightly regulated by the intrinsic properties of the polypeptide chain, linking thus phenotype and sequence^{24; 25}.

We have applied here the same approach in yeast to provide a quantitative assessment of the aggregation properties of proteins and how they correlate to sequential features in a more complex eukaryotic background. The budding yeast *Saccharomyces cerevisiae* is an excellent model system for the study of human cell biology in health and disease, since the basic features of eukaryotic cell biology evolved before the split between yeast and metazoans. This organism shares with higher eukaryotes numerous fundamental cellular pathways involved in neurodegeneration, such as protein quality control, membrane trafficking, autophagy or oxidative stress^{26; 27}. On this regard, humanized yeast models for AD, HD and PD have been successfully developed, recapitulating some of the pathological features associated with these disorders^{28; 29; 30; 31; 32; 33}. Here we analyze the aggregation properties of the above mentioned A β 42 mutants when expressed in yeast. In contrast to bacteria, where degradative pathways did not seem to affect the fate of these polypeptides, in yeast A β 42 variants are actively cleared from the cytosol by the yeast proteolytic machinery. We observe a significant correlation between the intrinsic properties of the peptides and both their intracellular aggregation and clearance *in vivo*. This model system should allow to rationalize and predict the impact of sequential changes in the deposition and degradation of amyloidogenic polypeptides in eukaryotic environments.

Material and Methods

Yeast strain and plasmids

Yeast strain BY4741 (*MAT a his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*) was transformed with pESC(-Ura) plasmids (Stratagene), encoding for the Aβ42-GFP fusion protein and the 19 variants differing in the 19th residue of Aβ42, as previously described³. Standard lithium/polyethylene glycol protocol was used for the transformation and selective synthetic complete medium deficient for uracil (SC-URA) was employed for plasmid selection.

Protein expression

Yeast cells were grown overnight in glucose SC-URA medium at 30°C and 100 μL were used to inoculate 5 mL of fresh medium. At an OD₆₀₀ of 0.5, cells were changed to a fresh raffinose SC-URA medium. After 30 min, cells were changed again to a fresh SC-URA medium containing 2% of galactose as a source of carbon to induce the recombinant protein expression. After 15 h at 30°C, cells were harvested, washed in sterile water and pellets were stored at -80°C for further analyses.

Microscopy

Cells were washed three times with sterile PBS and 5 μL were placed on top of microscopy glass slides covered with coverslips. Images were obtained at a 40-fold magnification using an emission filter for GFP under UV light excitation in a Leica fluorescence microscope (Leica DMBR, Heidelberg, Germany).

Fluorescence measurements

Cell pellets were resuspended in PBS to an OD₆₀₀ of 1. The emission spectra of GFP was recorded on a Cary Eclipse Spectrofluorometer (Agilent Technologies, Santa Clara, CA, USA) in the range 500-600 nm with a data interval of 1 nm and using an excitation wavelength of 488 nm.

Immunoblotting analysis

Cell pellets were resuspended in PBS. 200 μL of each mutant were prepared to an OD₆₀₀ of 20 and 5 μL were used for total fraction western-blot. For soluble/insoluble fraction analyses, 100 μL were centrifuged and resuspended in the same volume of Y-PER protein extraction reagent from Thermo Scientific (supplemented with a protease

inhibitor cocktail tablet from Roche) to induce cell lysis. After 20 min incubation at room temperature under mild agitation, mixtures were centrifuged at maximum speed for 30 min. Insoluble fractions were resuspended in 100 μ L of PBS containing a protease inhibitor cocktail tablet and used for western-blot analysis, together with 100 μ L of the soluble fraction. 5 μ L of each sample were loaded in a 14% SDS-PAGE and blotted onto a PVDF membrane. Immunodetection was performed using β -amyloid antibody 6E10 from Covance and membranes were developed with the ECL method. Densitometries were performed using ImageJ software.

Flow cytometry analysis

Cells expressing A β 42 $_{wt}$ -GFP and A β 42F19E-GFP were harvested and washed three times in 0.22 μ m filtered PBS. Flow cytometry measurements were performed using a FACSCanto flow cytometer (BD Biosciences, San Jose, CA, USA) equipped with a 488 nm blue argon laser. Gated cells (by means of FSC and SSC parameters) were analyzed for green emission measured on a 530/30 nm BP filter. Obtained data were analyzed using BD FACSDiva 4.0 software.

Cell viability assays

Overnight cultures of A β 42 $_{wt}$ -GFP and A β 42F19E-GFP grown in SC-URA medium supplied with glucose (or galactose for the colony-forming units assay) were washed with sterile water and diluted to an OD₆₀₀ of 0.8. The following experiments were carried out in triplicates.

For determining the colony-forming units per mL (cfu/mL), 100 μ L of serial dilutions (10^{-3} - 10^{-7}) were plated in medium containing glucose and galactose and incubated at 30 °C for 3 days. Only colonies in plates containing between 30 and 300 colonies were monitored.

For spotting assay, 8 μ L of serial dilutions (10^{-1} - 10^{-4}) were spotted in glucose and galactose plates, subsequently incubated at 30 °C for 2 days.

For growth curve, overnight cultures diluted in water were used to inoculate 250 μ L of fresh medium containing glucose and galactose to a final OD₆₀₀ of 0.15 in 96 well plates. Cultures growth was monitored overnight at 28 °C in a Victor 3 Plate Reader (Perkin-Elmer, Inc., Waltham, MA, USA) measuring the OD₅₉₅ every 15 min.

Proteolytic degradation analysis

Yeast strains specifically used in these assays were: BJ5459 (*MAT a ura3-52 trp1 lys2-801 leu2Δ1 his3Δ200 pep4Δ::HIS3 prb1Δ1.6R can1 GAL*), and Δ erg6 and Δ atg1, both in BY4741 background.

In experiments with chemical compounds, the cell wall permeable yeast strain, Δ erg6, transformed with pESC(-Ura) plasmids, was incubated overnight with galactose SC-URA medium. Prior to drug application, protein expression was arrested by changing medium from galactose to glucose supplemented. The incubation with drugs was performed for 4 h at 30°C. The chemical compounds and the final concentrations used in these assays were: Phenylmethanesulfonyl fluoride (PMSF) dissolved in ethanol (EtOH) at 1 mM, rapamycin dissolved in dimethyl sulfoxide (DMSO) at 100 nM and MG-132 dissolved in DMSO at 50 μ M; all of them from Sigma-Aldrich. EtOH and DMSO treatments at the corresponding concentrations were used as controls.

Results

Prediction of the aggregation propensity of the 20 A β 42 mutants.

The library of 20 A β 42 mutants used in this study differ only in the residue in position 19 of the A β peptide (Fig. S1)²⁰. This residue is located in the CHC of the peptide and has been shown to affect the folding, self-assembly, and fibril structure of A β ^{34; 35}. We analyzed the aggregation propensity of these variants with three different algorithms conceptually unrelated to AGGRESCAN²² to obtain an unbiased evaluation of their intrinsic aggregation properties previous to their experimental characterization in yeast. We employed: (i) Zyggregator³⁶, which uses a set of physico-chemical properties of amino acid residues such as hydrophobicity, charge, and the propensity to adopt α -helical or β -structural conformations, (ii) FoldAmyloid³⁷, which exploits the amino acids expected packing density and their propensity to establish hydrogen bonds and (iii) TANGO³⁸, which is based on the tendency of a sequence to form β -sheets that will remain buried in the structure of an amyloid. The aggregation properties predicted by these algorithms, as well as AGGRESCAN, for the 20 A β 42 mutants were normalized and they are compared in Fig. 1. All predictors converge on ascribing the higher propensity values to A β 42 variants displaying aromatic and hydrophobic residues (Ile, Leu, Met, Phe, Val, Trp and Tyr) in position 19.

Formation of aggregated foci by A β 42-GFP mutants in the yeast cytoplasm.

Following the same strategy that we used previously in bacteria, the 20 A β 42 mutants were fused to GFP and the resulting fusions (A β -GFP) expressed individually in the BY4741 yeast strain. We placed A β -GFP under control of the tightly regulated galactose-inducible *GALI* promoter to allow, rapid, strong and synchronous induction of protein expression in all cells. After 15 hours of protein expression, GFP fluorescence was visible in the cytoplasm of all the strains. However, fluorescent foci were only observable in cells expressing certain mutants, while the rest of the variants presented diffuse GFP fluorescence throughout the cytoplasm (Fig. 2). We monitored and quantified the presence of foci in 1000 yeast cells for each of the 20 A β -GFP variants (Fig. 3A). Ile, Phe, Tyr, Leu, Met, Trp and Val mutants displayed the higher proportion of cells containing one or more foci in their cytoplasm, followed by Cys, Ala and, to a lower extent, Thr. The rest of variants did not exhibit any detectable A β -GFP foci, which contrasts with mutants like Ile or Phe mutants, where more than 60% of the

cells contained visible aggregates (Fig. 3A). According to the predictions in the previous section, the mutants resulting in the formation of foci in the yeast cytoplasm correspond to those exhibiting a high intrinsic aggregation propensity. We explored if for those mutants there exists any correlation between the number of cells displaying visible aggregates and the previously calculated relative aggregation rates in the bacterial cytosol (Fig. 3B). A striking correlation between these two parameters was observed ($R= 0.95$, $p<0.00005$), suggesting that the formation of protein inclusions occurs under similar constraints in bacteria and yeast cytosols. The data also imply the presence of an aggregation threshold below which no foci are formed. The residues leading to the formation of foci correspond precisely with the 10 more aggregation-prone amino acids in the AGGRESCAN scale, in such a way that mutants with residues whose intrinsic aggregation propensities are below that of Thr, do not result in any foci formation. Fitting the data with the other aggregation propensity scales resulted in lower correlations (Fig. S2), indicating that the aggregation propensities in yeast, in terms of foci formation, essentially recapitulate those observed in bacteria.

Solubility of A β 42-GFP mutants in the yeast cytoplasm.

In order to further characterize the aggregation properties of the different variants, yeast cells were lysed and the resulting soluble and insoluble fractions were analyzed by immunoblotting (Fig. 4A). Only the fusion proteins were immunoreactive to 6E10, an A β -specific antibody. Quantification of the bands corresponding to soluble and insoluble protein showed that, despite the presence of diffuse fluorescence in the cytosol and the absence of observable foci in many mutants, A β -GFP was in most cases mainly located in the insoluble fraction (Fig. 4A), a property that was independent of the system used for cell lysis and consistent with previous data in a similar yeast model, where A β 42 was fused at the GFP C-terminus and in which both the *wild type*, forming cytoplasmic foci, and a double mutant Phe19Ser/Leu34Pro, displaying diffuse fluorescence, were found in the insoluble fraction³⁹. However, in our conditions, the presence of soluble A β 42-GFP was also observable for many of the mutants. This is in contrast to what was previously observed in *E. coli*, where, independent of the considered mutant, >95% of the protein was located in the insoluble fraction. To compare the solubility of the different variants in yeast we normalized the fraction of soluble protein relative to the sum of soluble and insoluble protein for each mutant (Fig. 4B). Consistent with their high intrinsic aggregation propensities Phe, Trp, Ile and Tyr

variants were almost exclusively located in the insoluble fractions while for polar residues like Glu, Asn and Gln > 40% of the protein fusion was located in the soluble fraction. A plot of the percentage of soluble protein present in each mutant relative to its AGGRESCAN predicted aggregation propensity renders a significant correlation ($R=0.81$, $p<0.00002$) suggesting that intrinsic properties are important determinants of the *in vivo* solubility of the mutants in the cytosol (Fig. 4C).

Fluorescence levels of A β 42-GFP mutants in the yeast cytoplasm.

In our protein fusions, the GFP moiety acts as a reporter of the aggregation state of the A β 42⁴⁰; thus the total fluorescence of intact cells was expected to reflect aggregation propensities, as it occurs in bacteria²⁰. Fluorescence emission curves were collected for cells expressing each A β -GFP variant (representative examples are shown in Fig. 5A) and the relative intensity at the GFP emission maximum (510 nm) was measured (Fig. 5B). Again the aromatic residues Trp, Phe and Tyr displayed the lowest fluorescence levels and the polar residues Glu, Gln and Asn, the highest, in such a way that the fluorescence emitted by cells expressing the Glu variant was nine-fold higher than that emitted by those expressing the Trp variant. This fluorescence dynamic range between the most and the least fluorescent mutant is more than two times higher than that observed in *E. coli*. We measured the correlation between fluorescence emission and the predicted aggregation propensity according to AGGRESCAN; despite being significant, the correlation is lower than expected ($R=0.78$ $p<0.00003$) (Fig. 5C). The correlations obtained using the other three predictors were even lower (Fig. S2). An inspection of Fig. 5B and 5C allows to note deviations from the correlation for mutants bearing Trp, Phe, that, despite being predicted as aggregation-prone, behave as outliers, exhibiting lower levels of fluorescence emission than expected, which suggests that in contrast to bacteria, other factors apart from the intrinsic aggregation propensity might influence the fate of A β -GFP proteins in the more complex yeast cytosol.

Protein levels of A β 42-GFP mutants in the cytoplasm of yeast.

From Fig. 4A it is evident that, apart from the distribution between the soluble and insoluble fraction, the mutants differ in the levels of total protein present in the cytosol. This is again in contrast with *E. coli* data, where protein levels were essentially identical between the different mutants²⁰. To quantify the cytosolic protein levels we performed immunoblotting analysis with equal amounts of cells expressing the 20 different variants. As shown in Fig. 6A, striking differences between mutants were observed.

Trp and Phe mutants display extremely low protein levels when compared with variants bearing polar residues like Glu or Gln. The correlation between expression levels and intrinsic aggregation propensity was significant but rather low ($R=0.71$, $p<0.0002$). Theoretically, a protein can fit into four different classes depending on its aggregation propensity and relative protein levels (Fig. 6B). Interestingly, the A β -GFP protein set fits only into two of the four possible groups: (i) proteins with low aggregation propensity and high protein levels and (ii) proteins with high aggregation propensity and low protein levels. We did not observe any protein in the theoretical classes having high aggregation propensity and high protein levels or *vice versa*, which indicates that these two features are always anti-correlated in our protein set.

Cytotoxicity of A β 42-GFP variants with high and low aggregation propensity

To ensure that the observed differences in protein levels were not caused by differential plasmid loss between mutants, and thus by differences in gene expression, cultures containing the same number of cells of the high and low abundant Glu and Phe mutants were plated in glucose-containing selective medium (repressing conditions) 15h after induction of protein expression and the number of resulting colonies quantified. No significant differences in the number of colonies generated by the two strains were observed (Fig. 7A); the same results were obtained in spotting assays (Fig. 7B). This allows to discard plasmid copy variability as a major contributor to the observed differences in protein levels. These results are in good agreement with previous data obtained for fusions of A β 42 variants displaying different aggregation propensity in frame with GFP²⁹ or with the functional domain of Sup35⁴¹. Also in line with these studies, no significant differences were observed in the number of colonies generated by the two mutants when plated in galactose-containing selective medium (inducing conditions) (Fig. 7A and 7B). We also monitored the growth kinetics of the two variants under repressing and inducing conditions. As expected, yeast growth in glucose-containing medium was faster than in the presence of galactose, however in any of the two conditions we could observe significant differences in growth rates between mutants (Fig. 7C). The data indicate that the two variants do not exhibit differential cytotoxicity and are consistent with previous evidences showing that A β 42 (expressed alone or stabilized by the GFP tag) produced in the cytoplasm does not significantly impair yeast growth²⁹.

Proteolytic clearance of A β 42-GFP variants in yeast.

Due to the often deleterious nature of aggregation-prone proteins, they are usually targeted to degradation to prevent their accumulation⁴². The ubiquitin–proteasome and autophagy–lysosome/vacuolar pathways are the two main routes of protein and organelle clearance in eukaryotic cells^{43; 44; 45}. Differential proteolysis of the A β -GFP variants might well account for the observed differences in protein steady state levels. In yeast, amyloidogenic proteins can be degraded both by the proteasome and by autophagy/vacuolar systems^{46; 47}. Thus, we analyzed the impact of blocking these pathways, either using chemical compounds or genetically modified strains, on the levels of both the high and low aggregating Glu and Phe mutants.

We first addressed the role of autophagy by expressing the two A β -GFP fusions in a *Δatg1* strain. The ATG1 gene (autophagy-specific gene 1) encodes for a serine/threonine kinase involved in autophagy regulation and essential for induction of the autophagic pathway⁴⁸. After overnight expression, the amount of total protein was found to increase for both mutants, compared to parental strains, as revealed by immunoblotting (Fig.8A). To further confirm that the autophagy/vacuolar pathways were involved in the degradation of these two proteins, they were expressed in the BJ5459 strain lacking the PEP4 gene encoding for the vacuolar protease A, which initiates the activation of different vacuolar hydrolases, including aminopeptidase I, carboxypeptidase Y and proteinase B⁴⁹. As shown in Fig. 8A, the levels of the two proteins were higher in the mutant strain than in the parental one.

We used a chemical approach to further confirm the contribution of the autophagy/vacuolar pathways to the observed steady state protein levels. In this case, the cell wall permeable *Δerg6* mutant yeast strain was used to facilitate drug uptake and increase the intracellular concentration of the drugs added to the culture⁵⁰: PMSF and rapamycin. PMSF is a serine proteinase inhibitor that blocks the activity of numerous vacuolar proteases⁵¹ without affecting the proteasome function. Rapamycin is an inhibitor of the Tor2 kinase, which negatively regulates the autophagic pathway; thus rapamycin acts as an autophagy-inducing drug⁵². Because these compounds are known to have pleiotropic effects in yeast cells, protein expression was arrested prior to pharmacological treatments. Immunoblotting analysis shows that for both mutants protein levels increase after PMSF treatment and are reduced after incubation with rapamycin, compared to the respective controls (Fig. 8B), confirming thus the role played by autophagy and vacuolar proteases in the degradation of these protein species.

We addressed the role of the proteosomal machinery in A β -GFP clearance using the potent proteasome inhibitor MG-132⁵³ in the $\Delta erg6$ strain background. Immunoblots of MG-132 treated cells after arresting protein expression showed an increase in the protein levels of both mutants, when compared to controls (Fig. 8C), indicating that proteasome-mediated pathway is also involved in the processing of A β -GFP fusions.

Importantly, the changes in protein levels observed after chemical or genetic modulation of the different proteolytic pathways are not identical in the two mutants (Fig. 8D). Inhibition of proteolytic activity always has a higher impact on the levels of the Phe variant, whereas the induction of autophagy by rapamycin preferentially reduces the levels of the Glu mutant, likely because those of the Phe variant is already under autophagy control in normal conditions.

We hypothesized that the higher proteolytic susceptibility of the Phe mutant should be reflected not only in overall lower protein levels in the cell population but also in a higher number of cells in which the protein has been completely degraded and thus fluorescence cannot be detected. To test this hypothesis, we compared the proportion of cells displaying and not displaying fluorescence in the Phe and Glu variants, as measured using flow cytometry, which allows to uncouple mean fluorescence emission and the number of cells in the population contributing to this fluorescence (Fig. 9). In agreement with the spectrofluorimetric data, the mean fluorescence emission of cells expressing the Glu mutant was clearly higher than this of cells expressing the Phe variant. More interestingly, whereas we could detect fluorescence in 68% of Glu expressing cells, only 46% of Phe expressing cells exhibited fluorescence, reflecting thus differences in the proteolytic processing of the two proteins.

Overall, the data suggest a preferential action of the cellular proteolytic machinery against variants with high aggregation propensity, explaining thus the association between aggregation propensity and protein levels observed in Fig. 6.

Discussion

Many late-onset neurodegenerative diseases are caused by aggregation-prone proteins⁵⁴. Extensive genetic and transgenic data suggest that many of the mutations causing these proteinopathies trigger disease by conferring an increasing aggregation propensity to the affected proteins that results in a toxic gain of function². The aggregation propensities of proteins are thought to be controlled to a large extent by the physicochemical properties encoded in the primary sequence¹⁵. It has been shown that this rule applies *in vitro*, however it is not clear that it will work also *in vivo*, where proteostatic mechanisms might introduce significant bias to this relationship. Thus, it is important to understand both the intrinsic properties and the pathways that regulate the fate of harmful aggregation-prone proteins inside living cells. In an attempt to provide new clues on this process, we have expressed in the yeast cytosol a library of 20 A β -GFP mutants exhibiting a continuous gradient of intrinsic aggregation propensities, as predicted by different algorithms (Fig. 1) and previously validated *in vivo* in bacteria²⁰. The characterization of the aggregative properties of this protein set provides one of the few available quantitative data to model protein deposition inside eukaryotic cells.

A first important observation is that only a half of the mutants formed detectable aggregated foci when expressed in yeast, indicating that in our particular experimental conditions it exists a threshold for intracellular inclusion formation. This contrasts with the data previously obtained in bacteria, where all the variants aggregated into inclusion bodies. In our set, the foci formation threshold is located between Ser and Thr, two polar uncharged residues differing only in a single methylene group, illustrating the exquisite control that the sequence exerts on *in vivo* aggregation propensities, specially if we take into account that the complete A β -GFP fusion consists of ~ 300 residues. In this way, if we only consider those variants for which foci formation could be observed, the correlation between the aggregation rates derived from *E. coli* experiments and the percentage of cells displaying inclusions in yeast is strikingly high, in such a way that the 10 mutants showing this capability match exactly with the 10 variants with the highest AGGRESCAN propensities, suggesting that, as it happens in bacteria¹⁶, this algorithm constitutes a useful tool to predict protein aggregation tendencies in eukaryotic cells. The observed correlation between yeast and bacteria data provides one of the few experimental evidences that sequential features tune protein aggregation in a

similar manner in different biological contexts, a concept that we all assume to be true when using aggregation predictors to forecast the impact of genetic mutations in human conformational disorders.

Although coherent with previous reports³⁹, it was surprising to observe that even for those mutants displaying only diffuse fluorescence, more than 50% of the recombinant protein was located in the insoluble fraction. This is consistent with the observation that in eukaryotic backgrounds, aggregation might be different from inclusion formation⁵⁵. Inclusions refer to abnormal intracellular structures observed microscopically, while the formation of aggregates describes a biochemical phenomenon. The interchangeable use of the two terms results from the observation that aggregation-prone proteins linked to conformational disorders are often found in inclusions under pathological conditions. Our data indicate that these processes can be in fact experimentally dissociated. In eukaryotes the formation of inclusions is thought to be an active process involving different cellular components, including microtubules⁵⁶. Accordingly, in a cellular model of HD it was shown that microtubule inhibition results in a marked reduction of inclusions, despite most of the poly-Q protein remained still aggregated⁵⁵. However, this reduction of inclusions formation was also associated with an increase in the steady-state level of soluble protein, an observation that is coherent with the fact that mutants unable to form foci in our dataset are also those exhibiting the highest levels of soluble protein, thus linking protein-aggregation propensity and inclusion formation in yeast, in such a way that a certain level of solubility would likely prevent the proteins to enter into the inclusion formation pathway. This is in contrast to what happens in bacteria, where formation of inclusions seems to depend on passive diffusion⁵⁷ and all A β -GFP variants, including those bearing polar residues in position 19, appear to be above the aggregation threshold; thus they differ in their kinetics of aggregation but not in their solubility at equilibrium²¹.

Despite we find significant correlations between AGGRESCAN predicted aggregation propensities and the distribution of the variants in the soluble and insoluble fractions or their overall fluorescence emission, we also observe important deviations for certain mutants, which make both correlations being lower than expected. Specifically, the most aggregation-prone mutants tend to behave as outliers, displaying less soluble protein and fluorescence levels than predicted, which suggests that the fate of these mutants might be under special surveillance by the protein quality control machinery

(PQC). The PQC consists of molecular chaperones and various proteases that recognize and repair damaged proteins or, alternatively, remove the aberrant proteins⁵⁸. Despite, these protein functions tend to be well conserved across species, the repertoire of proteins involved in PQC in bacteria and yeast are certainly divergent, which might contribute to the observed differences between *E. coli* and yeast models. The fact that, in contrast to bacteria, the protein steady levels differ between variants suggests that this is the case.

As a general trend, in our dataset proteins predicted to be soluble are present at higher levels than those with higher predicted aggregation propensity, in such a way that the cellular protein levels of the different mutants are anti-correlated with their intrinsic aggregation tendencies ($R=0.71$, $p<0.0002$). The aggregation propensity of a given protein sequence in a defined environment depends on different physicochemical properties, mainly on its hydrophobicity, secondary structure and overall charge¹⁵. Zyggregator is an algorithm that uses a parametrization of this properties to predict aggregation propensities. The correlation between Zyggregator predictions and protein levels is however low ($R= 0.47$, $p<0.02$) (Fig. S2), indicating that the specific parameters used to predict aggregation rates do not work properly to predict protein levels, which leads us to the question of what are the specific physicochemical properties recognized by the PQC resulting in different protein steady levels in our system. In this context, charge does not seem to be a determinant factor, since Gln and Asn exhibit similar or higher protein levels than the charged acidic Glu and Asp residues. The secondary structure propensity of the polypeptide chain does not either seem to play a major role in PQC recognition, since the correlations of protein levels with β -sheet ($R=0.60$, $p<0.003$) and α -helix ($R=0.34$, $p<0.07$) propensities are rather low. This leaves us with hydrophobicity, a property known to be recognized by the yeast PQC in other protein models⁵⁹. We selected hydrophobicity scales composed of experimentally determined transfer free energies for each individual amino acid since they include the contributions of the peptide bonds, an important feature if the protein region is expected to be unfolded and interacting with the components of the PQC. The correlation coefficient between protein levels and hydrophobicity are $R=0.65$ ($p<0.001$) and $R=0.72$ ($p<0.0002$) for the octanol⁶⁰ and interface⁶¹ scales, respectively (Fig. 9). Thus, considering only the interface hydrophobicity of the residue at position 19 provides a predictive power similar to that of AGGRESCAN. In fact, Trp, the main

outliner in the different plots, is the residue displaying the highest hydrophobicity in this scale.

The PQC degradation system should target misfolded species and not folded states, which implies that it needs to recognize a feature that reports selectively on conformational defects present in the misfolded state. Hydrophobicity fulfils this requirement better than secondary structure, which is characteristic of folded states, or charge, which is already exposed to solvent in native conformations. Exposed hydrophobicity constitutes in most cases a structural abnormality since hydrophobic residues are buried inside the core, involved in the formation of protein interfaces or located within membranes and, therefore, their exposure indicates defects in folding, assembly or membrane insertion. In our dataset, the mutated residues are expected to be located in an essentially disordered context and highly exposed to solvent, mimicking the conditions in a misfolded environment; thus our data provide strong support for a selective recognition of hydrophobicity by protein degradation systems. Accordingly, we show that autophagy/vacuolar pathways preferentially target *wild type* A β over the Glu mutant, since chemical or genetic inhibition of the pathway results in a higher relative increase of the protein levels in the *wild type* protein, whereas induction of autophagy has a more severe effect in the Glu variant; likely because under these specific conditions it becomes a substrate of the degradation pathway, whereas in normal conditions it is not. The same type of selectivity seems to apply for the proteasome mediated degradation of A β -GFP fusions. Interestingly enough, Phe, in the wild type, is the most hydrophobic residue in the interface scale after Trp, whereas Glu occupies the last position in this hydrophobicity ranking. The fact that both AGGRESCAN and hydrophobicity scales correlate similarly with protein levels suggests that the binding of protein substrates to PQC might require a degree of exposed hydrophobicity similar to that promoting self-assembly and aggregation of the substrates, allowing in this way the PQC to target specifically the degradation of dangerous insoluble sequences. This view is consistent with the fact that the binding sites of chaperones and ubiquitin ligases acting as upstream conformational sorters for degradative pathways are themselves hydrophobic^{62, 63}.

Our results demonstrate that autophagy is involved in the clearance of *wild type* A β -GFP in yeast and are fully consistent with those recently reported for a C-terminal fusion of A β to GFP, where induction of autophagy by the drug latrepirdine resulted in

an increased degradation³⁹, as well as with those obtained for a yeast model of α -synuclein aggregation, overall converging to point out autophagy as a critical and generic component of the cellular clearance of potentially toxic aggregation-prone proteins. Proteasomal degradation is also involved in the processing of A β -GFP, despite its contribution is moderate and its activity cannot compensate for the lack of autophagic function since genetic or chemical inhibition of these pathways results in a much less efficient protein clearance when compared with *wild-type* cells. In fact, it has been suggested that once aggregation-prone substrates cannot be efficiently handled by the proteasome, autophagy becomes the default clearance pathway⁶⁴. Nevertheless, the aim of the present work is not to provide a detailed description of the pathways involved in intracellular A β -GFP degradation, but rather to demonstrate that the decision on whether a substrate would become a target of not of these pathways depends on its intrinsic aggregation/hydrophobicity propensity. Flow cytometry data indicate that the PQC performs fairly well in protecting cells from the presence aggregation-prone *wild type* A β -GFP variant, since less than half of the cells exhibit detectable fluorescence and thus an intact fusion, consistent with the view that a robust PQC degradation system should target those misfolded species with a higher risk of forming insoluble aggregates. Nevertheless, in agreement with previous observations^{29; 41}, aggregation-prone mutants forming intracellular deposits are not more cytotoxic than those displaying diffuse cytosolic fluorescence. This might indicate that moderate amounts of A β -GFP aggregates would be below a threshold for toxicity in yeast, as previously suggested for α -synuclein⁴⁷, or, alternatively, they might play a cytoprotective role as shown for mammalian aggresomes, which serve as a cytoplasmic recruitment centers to facilitate degradation of toxic proteins by autophagic pathways⁵⁵. If this latter scenario also applies in yeast, it would be consistent with our observation that the most aggregation-prone variants both form inclusions and are preferentially targeted for protein degradation, since these two outcomes would respond to the same process.

It is clear, that a deeper knowledge of the mechanism by which proteins are targeted for degradation will be important for understanding pathologic processes. Our present demonstration that sorting to these pathways is strongly influenced by the physicochemical properties of the protein sequence should facilitate the development of rational therapeutic approaches for protein conformational diseases. The potency of these type of strategies is highlighted but the recent observation that proteins ‘tagged’

with small, synthetic, non-polar molecules that increase their surface hydrophobicity, a key property according to our study, are targeted for degradation by the PQC, allowing thus their selective removal⁶⁵.

Figure Legends

Figure 1. Comparison of the predicted aggregation propensities of the 20 A β 42 mutants. Bar graph representing the aggregation propensities obtained by means of four bioinformatic algorithms: AGGRESCAN (<http://bioinf.uab.es/aggrescan/>), Zyggregator (<http://www-vendruscolo.ch.cam.ac.uk/zyggregator.php>), FoldAmyloid (<http://bioinfo.protres.ru/fold-amyloid/oga.cgi>) and TANGO (<http://tango.crg.es/>). Values have been normalized for each predictor.

Figure 2. Expression of the A β 42-GFP mutants in the intracellular space of *S. cerevisiae*. Microscopy images of yeast cells expressing A β 42-GFP variants for 15 hours under UV light. Fluorescence is present either as small foci (red arrows) or diffused in the cytoplasm of the cells.

Figure 3. Monitoring cells with intracellular aggregation foci. (a) Graph bar representing the percentage of cells containing fluorescent foci for each mutant. Values were obtained by monitoring the number of cells with intracellular aggregates among ~1000 cells from two different cultures for each mutant. (b) Correlation between the percentage of cells containing visible intracellular aggregation foci and the intrinsic aggregation propensity in *E. coli*.

Figure 4. Amount of soluble and insoluble A β 42-GFP detected by immunoblotting. (a) Western-blot of protein fractions for each mutant. Yeast cells were lysed with a chemical reagent and soluble and insoluble fractions were separated and loaded onto the protein gels (For each mutant: left and right bands correspond to the soluble and insoluble protein fractions, respectively). (b) Bar graph with the values obtained from the western-blot quantification using ImageJ software. Each bar represents the ratio between soluble and total protein levels. (c) Correlation between the relative amount of soluble protein and the intrinsic aggregation propensity *E. coli*.

Figure 5. Fluorescence emission of A β 42-GFP mutants. (a) Examples of GFP fluorescence curves obtained from cultures after 15 h expressing A β 42-GFP. Emission curves of washed entire cells were recorded in a spectrofluorometer using an excitation wavelength of 488 nm. (b) Bar graph representing the fluorescence of the 20 mutants expressing A β 42-GFP obtained from the GFP emission pick at 510 nm. (c) Correlation between the GFP fluorescence and the intrinsic aggregation propensity obtained *E. coli*.

Figure 6. Total A β 42-GFP levels detected by immunoblotting. (a) Western-blot of the total protein fraction of each mutant. Equal amounts of cultures were loaded onto the protein gels, as well as a reference band (not shown) to compare both membranes. (b) Dot-plot representing the spatial distribution of A β 42-GFP mutants depending on the intrinsic aggregation propensity and the amount of total A β 42-GFP. (in blue:

mutants with low aggregation propensity presenting high A β 42-GFP levels; in red: mutants with high aggregation propensity presenting low A β 42-GFP levels).

Figure 7. Viability of cells expressing A β 42-GFP (F and E mutants) in presence of glucose and galactose (protein induction conditions). (a) Bar graph representing colony forming units (cfu/mL) obtained from plating serial dilutions of each culture. (b) Images of plates representative of the spotting assay performed for each mutant. (c) Growth curve obtained from monitoring the OD₅₉₅ of cultures in a 96 well plate.

Figure 8. Involvement of cellular proteolytic pathways affecting A β 42-GFP levels in A β 42-GFP*wt* (F) and A β 42F19E-GFP (E) mutant. (a) Western-blot against A β 42-GFP in the autophagy deficient Δ atg1 strain (left) and in the proteases deficient strain BJ5459 (right). Expression in BY4741 strain is used as control. (b) Western-blot against A β 42-GFP in the wall-permeable strain Δ erg6 in the presence of the autophagy enhancer rapamycin dissolved in DMSO (left) and in the presence of the proteases inhibitor PMSF dissolved in EtOH (right). Treatments with the corresponding solvents are used as controls. (c) Western-blot against A β 42-GFP in the wall permeable strain Δ erg6 in the presence of the proteasome inhibitor MG-132 dissolved in DMSO, which is used as control treatment. (d) Bar graph representing the variation in A β 42-GFP levels caused by the genetic or chemical modulation of proteolytic pathways. Values are obtained from the quantification of western-blot using ImageJ software and correspond to the ratio of protein after modulating the proteolytic activity and protein in their respective controls. In the case of rapamycin treatment, the inverse of this ratio is represented since this compound reduces the protein levels

Figure 9. Flow cytometry detection of cells expressing A β 42-GFP. Flow cytometry analysis of F and E mutants (upper and lower series, respectively). Left panels correspond to forward scatter (FCS) vs. side scatter (SSC) dot-plots showing P1 gate. Middle panels correspond to SSC vs. GFP fluorescence dot-plots showing P2 population of cells expressing A β 42-GFP. Right panels correspond to cell frequency histograms for the analysis of cells expressing A β 42-GFP, delimited in P3 population. P2 and P3 are redundant populations exhibiting similar percentages.

Supplementary material

Figure 1. Scheme of the fusion protein expressed in this work. The green fluorescent protein (green) was fused to the N-terminus of the A β 42 peptide (black) through a short linker (magenta) in between. The different GFP-A β 42 variants differ only in the 19th residue, whose Phe in the *wild-type* has been point mutated to the 19 other natural amino acids.

Figure 2. Correlation between fluorescence emission and different prediction algorithms. GFP fluorescence values for each mutant (represented in Fig. 5) are plotted

against aggregation propensity predicted using alternative bioinformatics algorithms (represented in Fig. 1): (a) Zygggregator, (b) FoldAmyloid and (c) Tango.

Acknowledgements

We thank Josep A. Biosca (Universitat Autònoma de Barcelona) for kindly providing us the BJ5459 yeast strain as well as Prof. Gerhard H. Braus (Georg-August-Universität Göttingen) for the Δ erg6 and Δ atg1 yeast strains. Work in our lab is supported by grants BFU2010-14901 from Ministerio de Ciencia e Innovación (Spain) and 2009-SGR 760 from AGAUR (Generalitat de Catalunya). SV has been granted an ICREA ACADEMIA award (ICREA).

References

1. Chiti, F. & Dobson, C. M. (2006). Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem* **75**, 333-66.
2. Invernizzi, G., Papaleo, E., Sabate, R. & Ventura, S. (2012). Protein aggregation: mechanisms and functional consequences. *International Journal of Biochemistry & Cell Biology* **44**, 1541-54.
3. Morell, M., de Groot, N. S., Vendrell, J., Aviles, F. X. & Ventura, S. (2011). Linking amyloid protein aggregation and yeast survival. *Molecular BioSystems* **7**, 1121-8.
4. Ventura, S. & Villaverde, A. (2006). Protein quality in bacterial inclusion bodies. *Trends Biotechnol* **24**, 179-85.
5. Carrio, M., Gonzalez-Montalban, N., Vera, A., Villaverde, A. & Ventura, S. (2005). Amyloid-like properties of bacterial inclusion bodies. *J Mol Biol* **347**, 1025 - 1037.
6. Wang, L., Maji, S. K., Sawaya, M. R., Eisenberg, D. & Riek, R. (2008). Bacterial inclusion bodies contain amyloid-like structure. *PLoS Biol* **6**, e195.
7. Morell, M., Bravo, R., Espargaro, A., Sisquella, X., Aviles, F. X., Fernandez-Busquets, X. & Ventura, S. (2008). Inclusion bodies: specificity in their aggregation process and amyloid-like structure. *Biochimica et Biophysica Acta* **1783**, 1815-25.
8. Dobson, C. M. (2001). The structural basis of protein folding and its links with human disease. *Philos Trans R Soc Lond B Biol Sci* **356**, 133-45.
9. Rousseau, F., Schymkowitz, J. & Serrano, L. (2006). Protein aggregation and amyloidosis: confusion of the kinds? *Curr Opin Struct Biol* **16**, 118-26.

10. Fitzpatrick, A. W., Knowles, T. P., Waudby, C. A., Vendruscolo, M. & Dobson, C. M. (2011). Inversion of the balance between hydrophobic and hydrogen bonding interactions in protein folding and aggregation. *PLoS Comput Biol* **7**, e1002169.
11. Linding, R., Schymkowitz, J., Rousseau, F., Diella, F. & Serrano, L. (2004). A comparative study of the relationship between protein structure and beta-aggregation in globular and intrinsically disordered proteins. *J Mol Biol* **342**, 345 - 353.
12. Espargaro, A., Castillo, V., de Groot, N. S. & Ventura, S. (2008). The in vivo and in vitro aggregation properties of globular proteins correlate with their conformational stability: the SH3 case. *J Mol Biol* **378**, 1116-31.
13. Ventura, S. (2005). Sequence determinants of protein aggregation: tools to increase protein solubility. *Microbial Cell Factories* **4**, 11.
14. Ventura, S., Zurdo, J., Narayanan, S., Parreno, M., Manges, R., Reif, B., Chiti, F., Giannoni, E., Dobson, C. M., Aviles, F. X. & Serrano, L. (2004). Short amino acid stretches can mediate amyloid formation in globular proteins: the Src homology 3 (SH3) case. *Proc Natl Acad Sci U S A* **101**, 7258-63.
15. Chiti, F., Stefani, M., Taddei, N., Ramponi, G. & Dobson, C. M. (2003). Rationalization of the effects of mutations on peptide and protein aggregation rates. *Nature* **424**, 805-8.
16. Belli, M., Ramazzotti, M. & Chiti, F. (2011). Prediction of amyloid aggregation in vivo. *EMBO Reports* **12**, 657-63.
17. Castillo, V., Grana-Montes, R., Sabate, R. & Ventura, S. (2011). Prediction of the aggregation propensity of proteins from the primary sequence: aggregation properties of proteomes. *Biotechnology Journal* **6**, 674-85.
18. Ahmed, A. B. & Kajava, A. V. (2013). Breaking the amyloidogenicity code: Methods to predict amyloids from amino acid sequence. *FEBS Lett* **587**, 1089-95.
19. Guidolin, D., Agnati, L. F., Albertin, G., Tortorella, C. & Fuxe, K. (2012). Bioinformatics aggregation predictors in the study of protein conformational diseases of the human nervous system. *Electrophoresis* **33**, 3669-79.
20. de Groot, N. S., Aviles, F. X., Vendrell, J. & Ventura, S. (2006). Mutagenesis of the central hydrophobic cluster in Abeta42 Alzheimer's peptide. Side-chain properties correlate with aggregation propensities. *Febs J* **273**, 658-68.
21. de Groot, N. S. & Ventura, S. (2006). Protein activity in bacterial inclusion bodies correlates with predicted aggregation rates. *J Biotechnol* **125**, 110-3.

22. Conchillo-Sole, O., de Groot, N. S., Aviles, F. X., Vendrell, J., Daura, X. & Ventura, S. (2007). AGGRESCAN: a server for the prediction and evaluation of "hot spots" of aggregation in polypeptides. *BMC Bioinformatics* **8**, 65.
23. de Groot, N. S., Castillo, V., Grana-Montes, R. & Ventura, S. (2012). AGGRESCAN: method, application, and perspectives for drug design. *Methods in Molecular Biology* **819**, 199-220.
24. Espargaro, A., Sabate, R. & Ventura, S. (2008). Kinetic and thermodynamic stability of bacterial intracellular aggregates. *FEBS Letters* **582**, 3669-73.
25. Villar-Pique, A., de Groot, N. S., Sabate, R., Acebron, S. P., Celaya, G., Fernandez-Busquets, X., Muga, A. & Ventura, S. (2012). The effect of amyloidogenic peptides on bacterial aging correlates with their intrinsic aggregation propensity. *Journal of Molecular Biology* **421**, 270-81.
26. Khurana, V. & Lindquist, S. (2010). Modelling neurodegeneration in *Saccharomyces cerevisiae*: why cook with baker's yeast? *Nature Reviews Neuroscience* **11**, 436-49.
27. Tenreiro, S. & Outeiro, T. F. (2010). Simple is good: yeast models of neurodegeneration. *FEMS Yeast Res* **10**, 970-9.
28. Treusch, S., Hamamichi, S., Goodman, J. L., Matlack, K. E., Chung, C. Y., Baru, V., Shulman, J. M., Parrado, A., Bevis, B. J., Valastyan, J. S., Han, H., Lindhagen-Persson, M., Reiman, E. M., Evans, D. A., Bennett, D. A., Olofsson, A., DeJager, P. L., Tanzi, R. E., Caldwell, K. A., Caldwell, G. A. & Lindquist, S. (2011). Functional links between A β toxicity, endocytic trafficking, and Alzheimer's disease risk factors in yeast. *Science* **334**, 1241-5.
29. D'Angelo, F., Vignaud, H., Di Martino, J., Salin, B., Devin, A., Cullin, C. & Marchal, C. (2013). A yeast model for amyloid-beta aggregation exemplifies the role of membrane trafficking and PICALM in cytotoxicity. *Dis Model Mech* **6**, 206-16.
30. Cooper, A. A., Gitler, A. D., Cashikar, A., Haynes, C. M., Hill, K. J., Bhullar, B., Liu, K., Xu, K., Strathearn, K. E., Liu, F., Cao, S., Caldwell, K. A., Caldwell, G. A., Marsischky, G., Kolodner, R. D., Labaer, J., Rochet, J. C., Bonini, N. M. & Lindquist, S. (2006). Alpha-synuclein blocks ER-Golgi traffic and Rab1 rescues neuron loss in Parkinson's models. *Science* **313**, 324-8.
31. Outeiro, T. F. & Lindquist, S. (2003). Yeast cells provide insight into alpha-synuclein biology and pathobiology. *Science* **302**, 1772-5.
32. Krobitsch, S. & Lindquist, S. (2000). Aggregation of huntingtin in yeast varies with the length of the polyglutamine expansion and the expression of chaperone proteins. *Proc Natl Acad Sci U S A* **97**, 1589-94.

33. Willingham, S., Outeiro, T. F., DeVit, M. J., Lindquist, S. L. & Muchowski, P. J. (2003). Yeast genes that enhance the toxicity of a mutant huntingtin fragment or alpha-synuclein. *Science* **302**, 1769-72.
34. Morimoto, A., Irie, K., Murakami, K., Masuda, Y., Ohigashi, H., Nagao, M., Fukuda, H., Shimizu, T. & Shirasawa, T. (2004). Analysis of the secondary structure of beta-amyloid (Abeta42) fibrils by systematic proline replacement. *J Biol Chem* **279**, 52781 - 52788.
35. Bitan, G., Vollers, S. S. & Teplow, D. B. (2003). Elucidation of primary structure elements controlling early amyloid beta-protein oligomerization. *J Biol Chem* **278**, 34882-9.
36. Tartaglia, G. G. & Vendruscolo, M. (2008). The Zyggregator method for predicting protein aggregation propensities. *Chem Soc Rev* **37**, 1395-401.
37. Garbuzynskiy, S. O., Lobanov, M. Y. & Galzitskaya, O. V. (2010). FoldAmyloid: a method of prediction of amyloidogenic regions from protein sequence. *Bioinformatics* **26**, 326-32.
38. Fernandez-Escamilla, A. M., Rousseau, F., Schymkowitz, J. & Serrano, L. (2004). Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat Biotechnol* **22**, 1302 - 1306.
39. Bharadwaj, P. R., Verdile, G., Barr, R. K., Gupta, V., Steele, J. W., Lachenmayer, M. L., Yue, Z., Ehrlich, M. E., Petsko, G., Ju, S., Ringe, D., Sankovich, S. E., Caine, J. M., Macreadie, I. G., Gandy, S. & Martins, R. N. (2012). Latrepirdine (Dimebon) enhances autophagy and reduces intracellular GFP-Abeta42 levels in yeast. *J Alzheimers Dis* **32**, 949-67.
40. Villar-Pique, A., Espargaro, A., Sabate, R., de Groot, N. S. & Ventura, S. (2012). Using bacterial inclusion bodies to screen for amyloid aggregation inhibitors. *Microbial Cell Factories* **11**, 55.
41. Bagriantsev, S. & Liebman, S. (2006). Modulation of Abeta42 low-n oligomerization using a novel yeast reporter system. *BMC Biol* **4**, 32.
42. Rubinsztein, D. C. (2006). The roles of intracellular protein-degradation pathways in neurodegeneration. *Nature* **443**, 780-6.
43. Tanaka, K., Mizushima, T. & Saeki, Y. (2012). The proteasome: molecular machinery and pathophysiological roles. *Biol Chem* **393**, 217-34.
44. Webb, J. L., Ravikumar, B., Atkins, J., Skepper, J. N. & Rubinsztein, D. C. (2003). Alpha-Synuclein is degraded by both autophagy and the proteasome. *The Journal of biological chemistry* **278**, 25009-13.
45. Rubinsztein, D. C., Codogno, P. & Levine, B. (2012). Autophagy modulation as a potential therapeutic target for diverse diseases. *Nat Rev Drug Discov* **11**, 709-30.

46. Kaganovich, D., Kopito, R. & Frydman, J. (2008). Misfolded proteins partition between two distinct quality control compartments. *Nature* **454**, 1088-95.
47. Petroi, D., Popova, B., Taheri-Talesh, N., Irniger, S., Shahpasandzadeh, H., Zweckstetter, M., Outeiro, T. F. & Braus, G. H. (2012). Aggregate clearance of alpha-synuclein in *Saccharomyces cerevisiae* depends more on autophagosome and vacuole function than on the proteasome. *J Biol Chem* **287**, 27567-79.
48. Matsuura, A., Tsukada, M., Wada, Y. & Ohsumi, Y. (1997). Apg1p, a novel protein kinase required for the autophagic process in *Saccharomyces cerevisiae*. *Gene* **192**, 245-50.
49. Woolford, C. A., Daniels, L. B., Park, F. J., Jones, E. W., Van Arsdell, J. N. & Innis, M. A. (1986). The PEP4 gene encodes an aspartyl protease implicated in the posttranslational regulation of *Saccharomyces cerevisiae* vacuolar hydrolases. *Molecular and Cellular Biology* **6**, 2500-10.
50. Emter, R., Heese-Peck, A. & Kralli, A. (2002). ERG6 and PDR5 regulate small lipophilic drug accumulation in yeast cells via distinct mechanisms. *FEBS Lett* **521**, 57-61.
51. Jones, E. W. (2002). Vacuolar proteases and proteolytic artifacts in *Saccharomyces cerevisiae*. *Methods Enzymol* **351**, 127-50.
52. Kamada, Y., Sekito, T. & Ohsumi, Y. (2004). Autophagy in yeast: a TOR-mediated response to nutrient starvation. *Current Topics in Microbiology and Immunology* **279**, 73-84.
53. Lee, D. H. & Goldberg, A. L. (1998). Proteasome inhibitors: valuable new tools for cell biologists. *Trends in Cell Biology* **8**, 397-403.
54. Dobson, C. M. (2002). Protein-misfolding diseases: Getting out of shape. *Nature* **418**, 729 - 730.
55. Taylor, J. P., Tanaka, F., Robitschek, J., Sandoval, C. M., Taye, A., Markovic-Plese, S. & Fischbeck, K. H. (2003). Aggresomes protect cells by enhancing the degradation of toxic polyglutamine-containing protein. *Human Molecular Genetics* **12**, 749-57.
56. Kopito, R. R. (2000). Aggresomes, inclusion bodies and protein aggregation. *Trends Cell Biol* **10**, 524 - 30.
57. Coquel, A. S., Jacob, J. P., Primet, M., Demarez, A., Dimiccoli, M., Julou, T., Moisan, L., Lindner, A. B. & Berry, H. (2013). Localization of Protein Aggregation in *Escherichia coli* Is Governed by Diffusion and Nucleoid Macromolecular Crowding Effect. *PLoS Computational Biology* **9**, e1003038.
58. Chen, B., Retzlaff, M., Roos, T. & Frydman, J. (2011). Cellular strategies of protein quality control. *Cold Spring Harb Perspect Biol* **3**, a004374.

59. Fredrickson, E. K., Rosenbaum, J. C., Locke, M. N., Milac, T. I. & Gardner, R. G. (2011). Exposed hydrophobicity is a key determinant of nuclear quality control degradation. *Molecular Biology of the Cell* **22**, 2384-95.
60. Wimley, W. C., Creamer, T. P. & White, S. H. (1996). Solvation energies of amino acid side chains and backbone in a family of host-guest pentapeptides. *Biochemistry* **35**, 5109-24.
61. Wimley, W. C. & White, S. H. (1996). Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nature Structural Biology* **3**, 842-8.
62. Zhu, X., Zhao, X., Burkholder, W. F., Gragerov, A., Ogata, C. M., Gottesman, M. E. & Hendrickson, W. A. (1996). Structural analysis of substrate binding by the molecular chaperone DnaK. *Science* **272**, 1606-14.
63. Fredrickson, E. K. & Gardner, R. G. (2012). Selective destruction of abnormal proteins by ubiquitin-mediated protein quality control degradation. *Seminars in Cell & Developmental Biology* **23**, 530-7.
64. Rideout, H. J., Lang-Rollin, I. & Stefanis, L. (2004). Involvement of macroautophagy in the dissolution of neuronal inclusions. *International Journal of Biochemistry & Cell Biology* **36**, 2551-62.
65. Neklesa, T. K. & Crews, C. M. (2012). Chemical biology: Greasy tags for protein removal. *Nature* **487**, 308-9.

Fig. 1

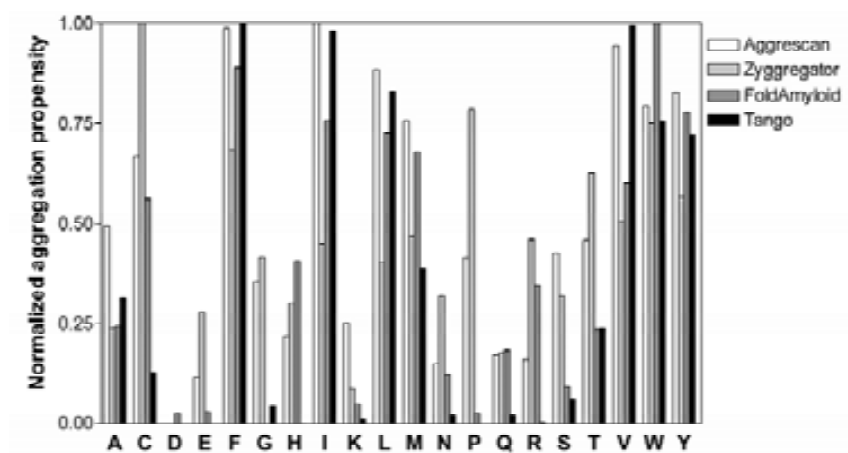


Fig. 2

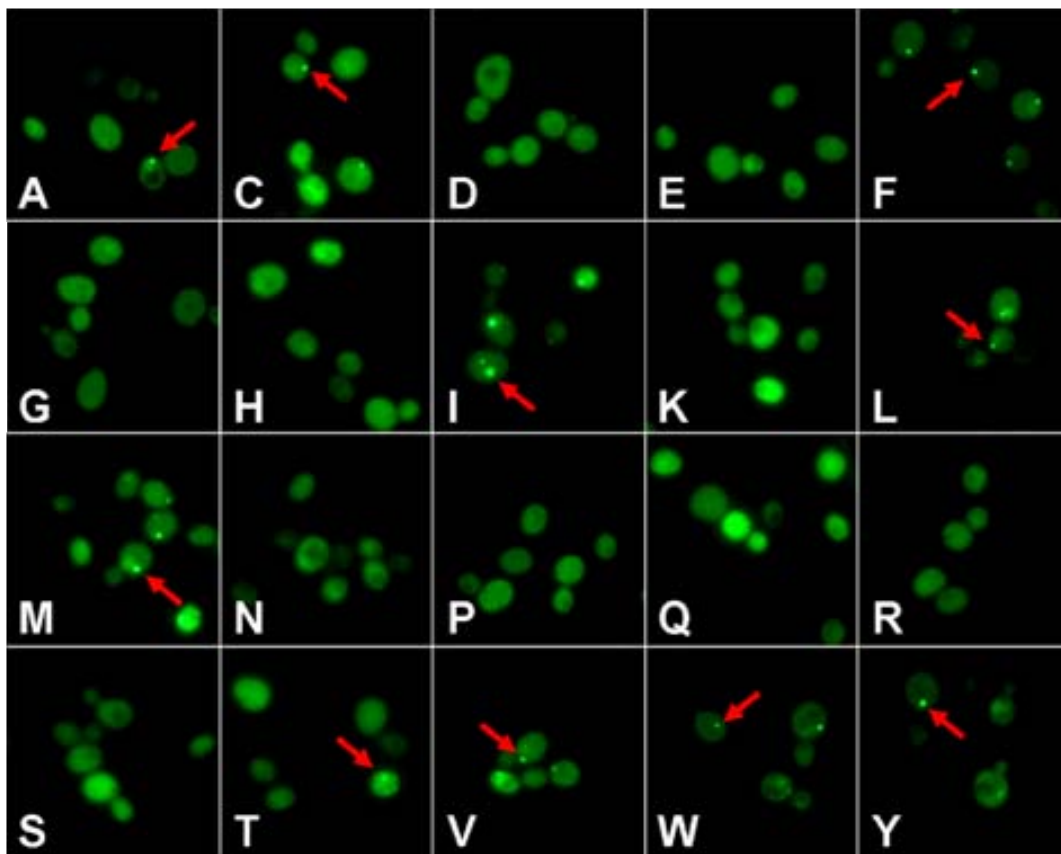


Fig. 3

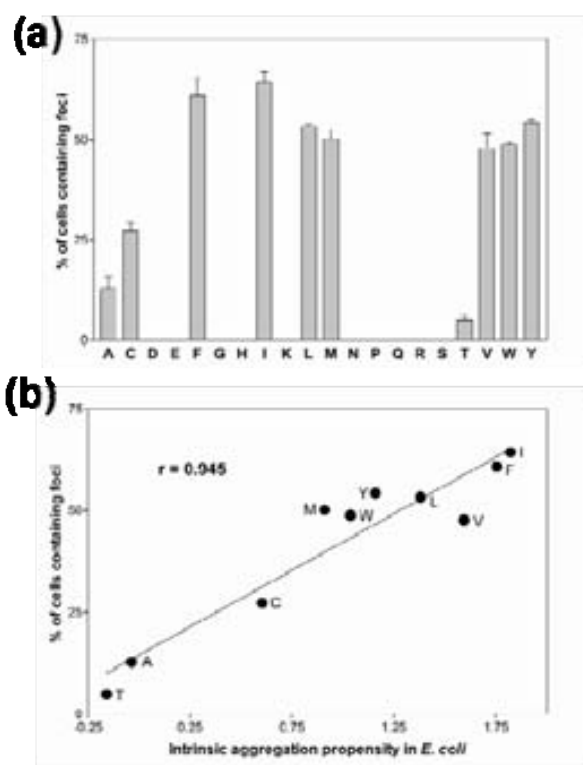


Fig. 4

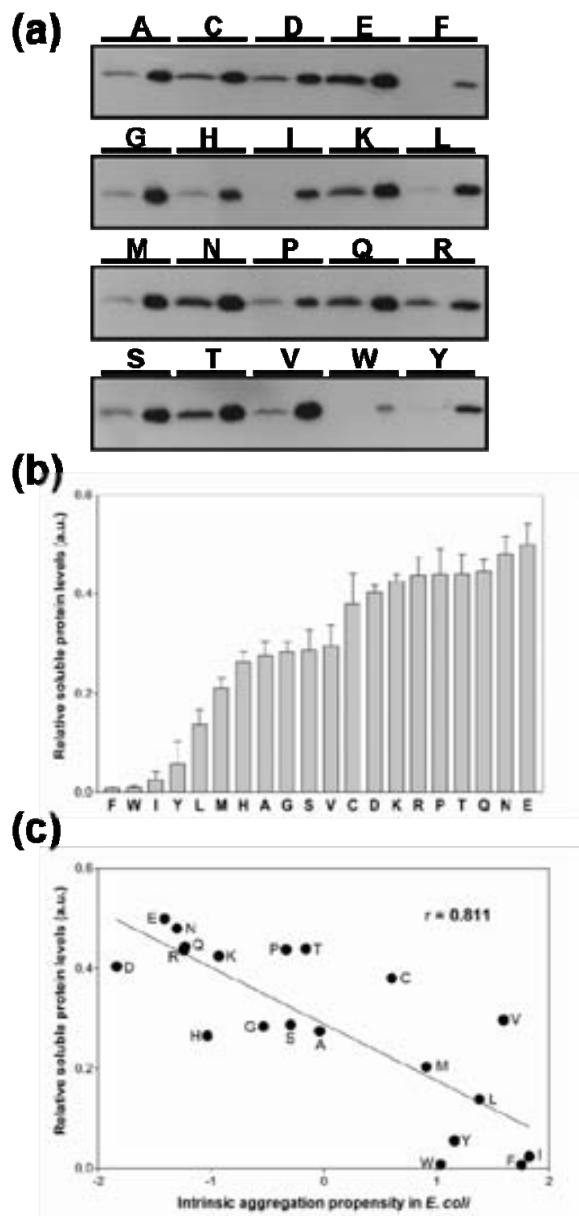


Fig. 5

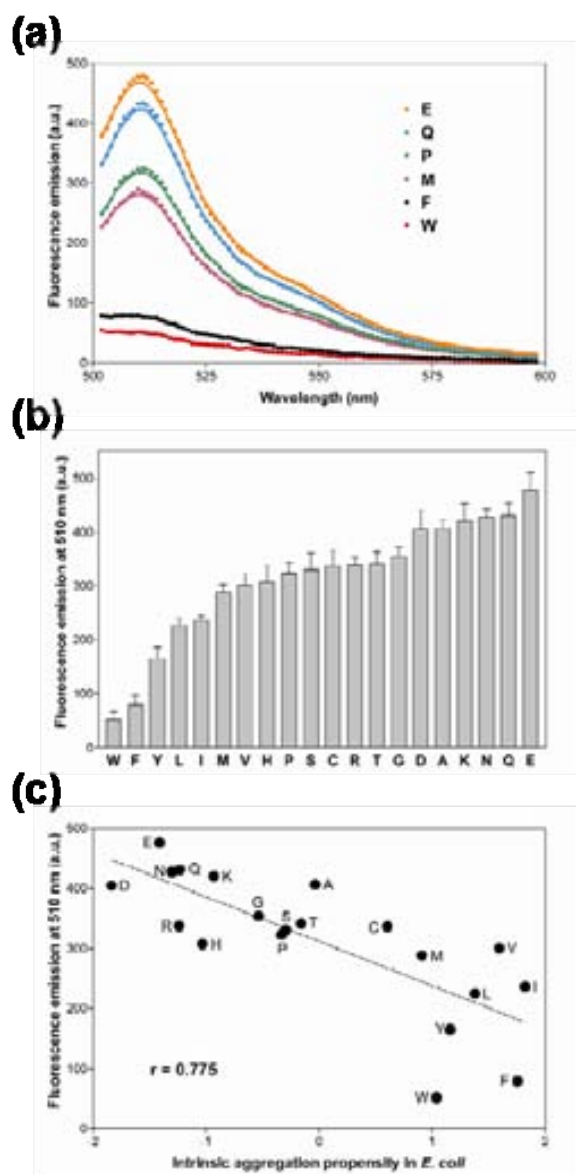


Fig. 6

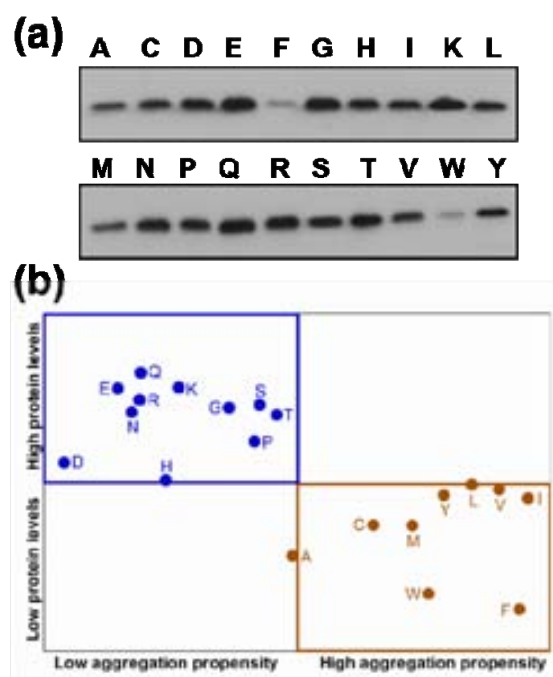


Fig. 7

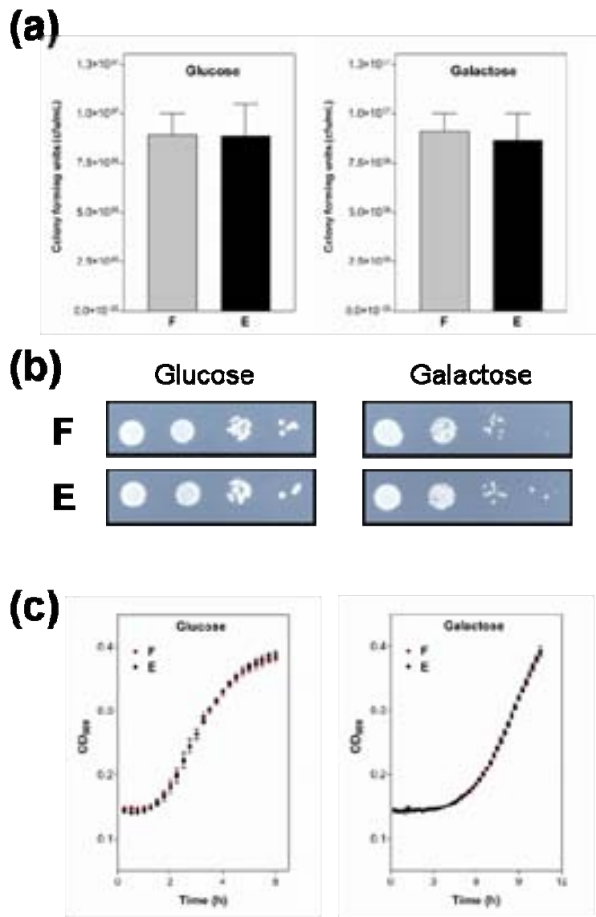


Fig. 8

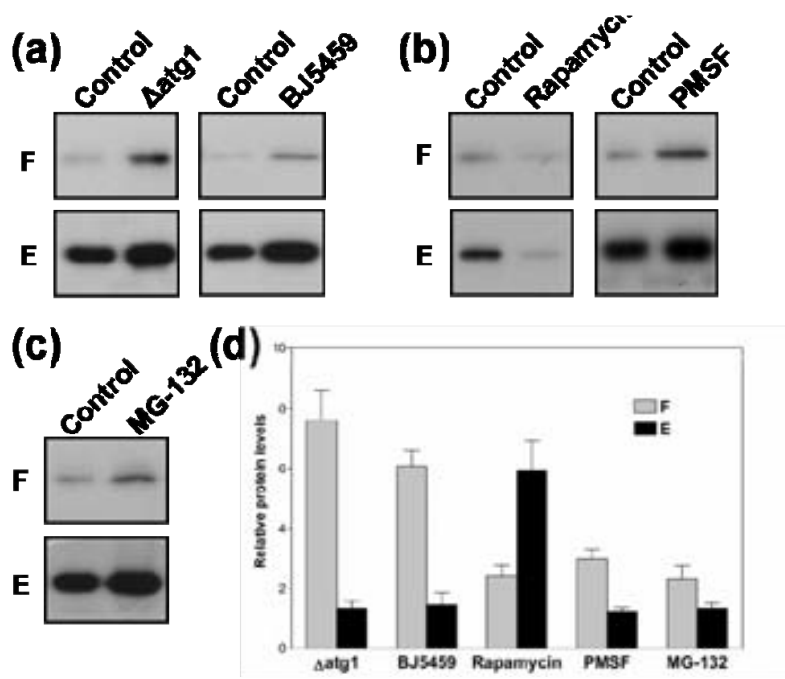


Fig. 9

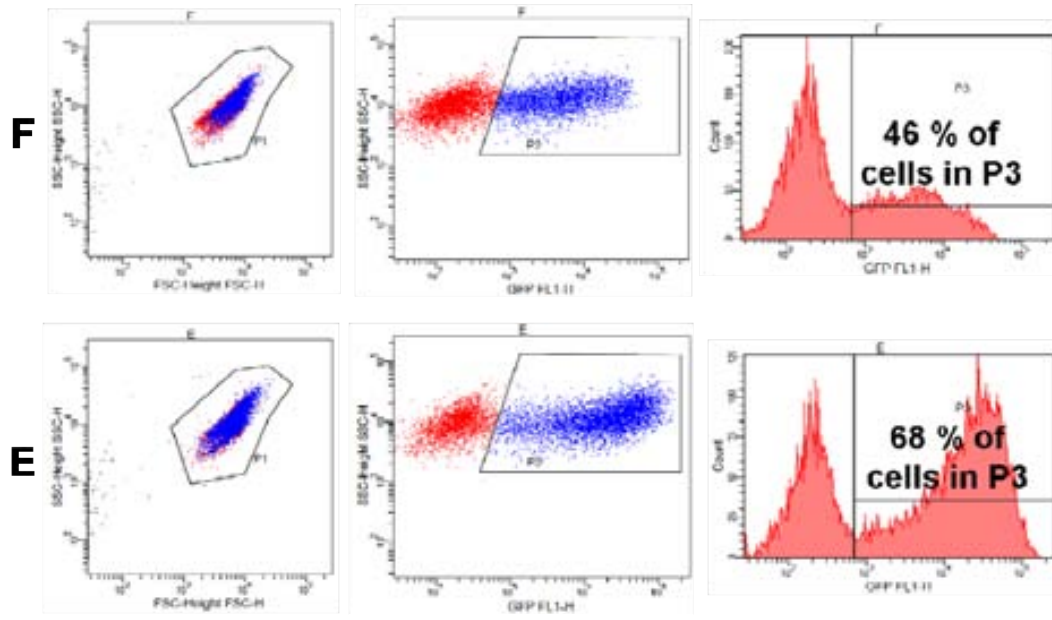


Fig. S1

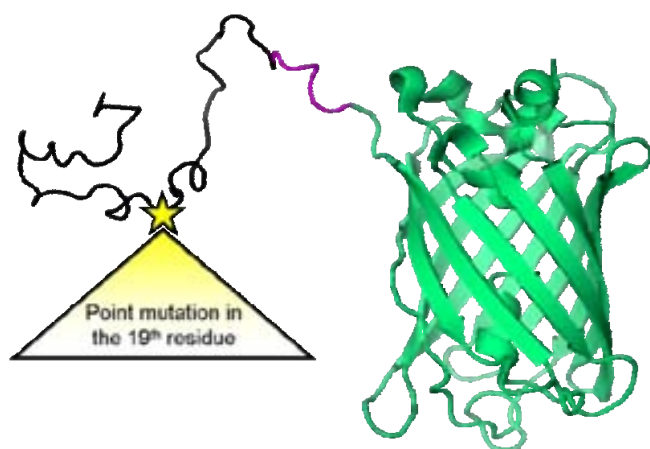


Fig. S2

