Universitat Autònoma de Barcelona

From Sequence to Structure: Determinants of Functional and Non-functional Protein Aggregation

Patrizia Marinelli

September 2015

Universitat Autònoma de Barcelona

Departament de Bioquímica i Biologia Molecular

Institut de Biotecnologia i de Biomedicina

From Sequence to Structure: Determinants of Functional and Non-functional Protein Aggregation

Doctoral thesis presented by Patrizia Marinelli for the degree of PhD 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 Dr. Salvador Ventura Zamora and Dr. Susanna Navarro Cantero.

Patrizia Marinelli

Dr. Salvador Ventura Zamora

Dr. Susanna Navarro Cantero

Bellaterra, September 2015

Summary

Summary

The study of protein aggregation represents a challenging research field which embraces from biomedical to biotechnological areas. The growing number of human diseases associated to the deposition of amyloid aggregates as well as the formation of intracellular protein deposits during the recombinant production of therapeutic proteins in cell factories has pushed the research to understand and develop strategies against protein aggregation. Until the past decade, amyloid aggregation was thought to be only associated with pathological amyloid states or unsuccessful folding during heterologous protein production. This idea was radically changed with the discovery of amyloid fibers being used for functional purposes in bacteria, fungi, insects, invertebrates and humans This suggests that the amyloid propensity is not confined to toxic amyloidosis but it may be considered as an intrinsic property of the polypeptide chain that allows the build up of macromolecular assemblies with exceptional structural properties. In the present thesis we have used a set of biophysical and computational tools combined with in vitro and in vivo models to characterize both functional and non-functional aggregation. On one side, we have analysed the specific role of bacterial peptides in biofilm amyloid formation and virulence in S. aureus. On other side, structurally distinct protein models have been used to study the role of molecular determinants such as disulfide bond crosslinking, harsh environmental conditions and intrinsic amyloidogenic propensities in the formation of non-functional aggregates. Overall, the data reported here underline the tight control exerted by Nature to allow functional amyloid assembly while counterbalancing intrinsic and extrinsic factors that might lead to toxic amyloid aggregation.

Resumen

El estudio de la agregación proteica representa un campo de investigación desafiante que abarca tanto el área biomédica como la biotecnológica. El creciente número de enfermedades humanas asociadas a la acumulación de agregados amiloides, así como la formación de depósitos intracelulares durante la producción recombinante de proteínas terapéuticas en modelos celulares ha impulsado la investigación para comprender y desarrollar estrategias contra la agregación de proteínas. Hasta la última década, se pensaba que la agregación amiloide estabafuese únicamente asociada a estados patológicos o a un plegamiento erróneo durante la producción de proteínas heterólogas. Esta idea fue cambiada radicalmente con el descubrimiento de fibras amiloides utilizadas con fines funcionales en bacterias, hongos, insectos, invertebrados y seres humanos. Esto sugiere que la tendencia a formar estructuras amiloides no se limita exclusivamente a amiloidosis de tipo tóxico si no que puede ser considerada como una propiedad intrínseca de cada cadena polipéptidica que permite la acumulación de agregados macromoleculares con propiedades estructurales excepcionales. En la presente tesis se ha utilizado un conjunto de herramientas biofísicas y computacionales combinadas con modelos in vitro e in vivo para caracterizar la agregación funcional y no funcional. Por un lado, hemos analizado el papel específico de péptidos bacterianos en la formación de biofilms de tipo amiloide así como en la virulencia de S. aureus. Por otro lado, modelos de proteínas estructuralmente distintas se han utilizado para estudiar el papel de los determinantes moleculares tales como la restricción conformacional aportada por puentes disulfuros, condiciones ambientales críticas y propiedades amiloidogénicas intrínsecas en la formación de agregados no funcionales. En general, los datos aquí presentados resaltan el estricto control ejercido por la Naturaleza en permitir que permite el ensamblaje amiloide funcional contrarrestando factores intrínsecos y extrínsecos que pueden conducir a la agregación toxica de tipo amiloide.

Abbreviations

a.u.	Arbitrary units
AcP	Acylphosphatase
AD	Alzheimer's disease
AFM	Atomic force microscopy
ALS	Amyotrophic lateral sclerosis
APR	Aggregation Prone Region
ATR-FTIR	Attenuated Total Reflectance-Fourier Transform Infrared spectroscopy
Aβ42	Amyloid beta peptide 42
Bis-ANS	4'-dianilino-1,1'-binaphthyl-5,5'-disulfonic acid
BPTI	Bovine pancreatic trypsin inhibitor
BSE	Bovine spongiform encephalopathy
CD	Circular dichroism
CI2	Chymotrypsin inhibitor 2
CJD	Creutzfeldt-Jakob Disease
CR	Congo red
DNA	Deoxyribonucleic acid
DSC	Differential scanning calorimetry
DTT	Dithiothreitol
EGCg	Epigallocatechin-3-gallate
ER	Endoplasmic reticulum
EWSR1	EWS RNA-binding protein
FAP	Familial amyloidosis polyneuropathy
FT-IR	Fourier transformed infrared spectroscopy

FTLD	Frontotemporal lobar degeneration
FUS	Fused in sarcoma
GFP	Green fluorescent protein
GITC	Guanidinium thiocyanate
GSH	Glutathione
GST	Glutathione S-Transferase
GuHCl	Guanidinium chloride
HnRNP	Heterogeneous nuclear ribonucleoprotein
HNRPDL	Heterogeneous nuclear ribonucleoprotein D-Like
HPLC	High-performance liquid chromatography
I	Intermediate state
i.e.	id est (that is)
IBs	Inclusion bodies
IDP	Intrinsically disordered protein
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LBs	Lewy bodies
MD	Molecular dynamic
MG	molten globule state
mRNA	Messenger ribonucleic acid
N	Native state
NMR	Nuclear magnetic resonance
PAPA	Prion aggregation prediction algorithm
PBS	Phosphate buffered saline

PCR	Polymerase chain reaction
PD	Parkinson's disease
PDI	Protein disulfide isomerase
PI	Propidium iodide
PK	Proteinase K
PLAAC	Prion-like amino acid composition
PQC	Protein Quality Control
PrLD	Prion-Like domain
PSMs	Phenol soluble modulins
RNA	Ribonucleic acid
RNase A	Ribonuclease A
RONN	Regional order neural network software
ROS	Reactive oxygen species
RRM	RNA Recognition Motif
SCCmec	Staphylococcal cassette chromosome mec
SSA	Senile systemic amyloidosis
ssNMR	Solid-state NMR
TAF15	TATA-inding protein-ssociated factor 2 N
TDP-43	TAR DNA-inding protein 43
TEM	Transmission electron microscopy
TFE	2,2,2-trifluoroethanol
Th-S	Thioflavin-S
Th-T	Thioflavine T

TTR	Transthyretin
U	Unfolded state
UPS	Ubiquitin proteasome system
URN1-FF	FF domain of the URN1 splicing factor

Contents

PART I - General introduction

1 Protein Folding

1.1 Proteins

Proteins are the most abundant cellular macromolecules, which are involved in a wide range of functions (Lodish et al., 2000). Each protein is composed by a different combination of amino acids that defines its primary structure. Specifically, DNA molecules store the genetic information that is translated through RNA molecules (mRNA) into amino acids that, during the protein synthesis, are assembled across the ribosome machinery in a linear combination to form the polypeptide chain (Crick, 1970).

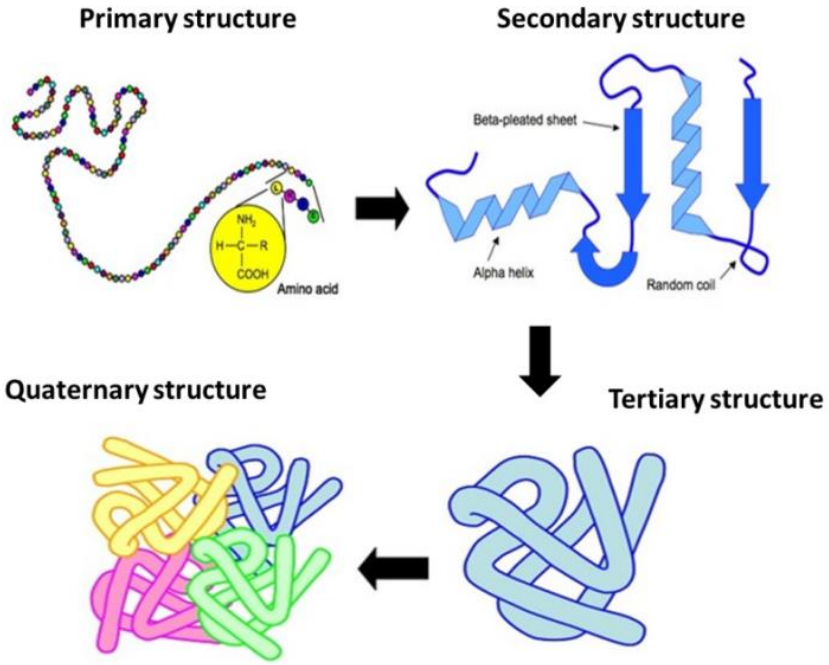

Figure I-1.1| **Different levels of protein structure**. From polypeptide chain to functional conformation. Adapted from (http://www.vce.bioninja.com.au/aos-1-molecules-of life/biomolecules/proteins.html).

In nature there are twenty standard amino acids, which are classified according to their different side chains that confer them distinct chemical and physical properties. The principal forces that drive the formation of the secondary structure include local interactions such as hydrogen bonds and ionic interactions. The secondary structure is constituted by ordered elements, mainly α-helices and β-sheets, but proteins can also exhibit a random coil conformation without a preferred regular structure (Figure I-1.1). The tertiary structure corresponds to the spatial distribution of the secondary structure in the protein. The major force stabilizing the tertiary structure is

the hydrophobic interaction among non-polar side chains in the core of the protein. Additional stabilizing forces include salt bridges, hydrogen bonds, disulfide bonds, and even post translational modifications. Finally, the quaternary structure arises from joining individual subunits with a tertiary structure in a multimeric complex (Creighton, 1993).

1.2 Historical overview

Protein folding is generally defined as the process by which a protein obtains its functional three dimensional structure that defines the native state or functional state in physiological conditions. At the beginning of 1960s Christian B. Anfinsen and co-workers analyzed the *in vitro* refolding of the ribonuclease A (RNase A), showing that the native state depends firstly on its amino acid sequence. They elaborated the thermodynamic hypothesis, or Anfinsen's dogma, affirming that in a given environmental context the native state will be the most stable structure, corresponding to the minimum of its Gibbs free energy (Anfinsen et al., 1961). On the other hand, Cyrus Levinthal demonstrated that the acquisition of the native structure by a random search would take astronomical time periods due to very large number of freedom degrees in unfolded polypeptide chains (Levinthal, 1968; Zwanzig et al., 1992), which crashed with the evidence that proteins usually fold in a time range from microseconds to seconds. The so called "Levinthal Paradox" promoted the study of folding mechanisms to explain how proteins can fold in a biologically relevant timeframe.

1.3 Native state and folding species

Anfinsen suggested the main driving force for folding as the amino acidic composition of proteins, but this statement was early debated by numerous studies demonstrating that proteins with different amino acidic content can share a similar fold, such as the case of horse and yeast globins or the cytochrome c family. Hence the folding code cannot be seen as the simple mRNA triplet code but should be like directed by a complex series of factors. It was thought that the identification of all the species involved in the folding process would allow understanding the folding code (Kim & Baldwin, 1982). Many experiments have been performed to study the folding mechanism of proteins departing from artificial denaturing conditions. In the presence of a strong chemical denaturant (urea or guanidium chloride), high or low pH or high temperature the protein is in its unfolded state and the removal of this conditions allows, in some cases, the spontaneous refolding to the native structure. The use of different techniques such as fluorimetry, circular dichroism, FT-IR and NMR spectroscopy allows to track the refolding process and discriminate between unfolded

and native states (Dill et al., 2008; Baldwin, 2008). In this context, the native or folded state of a protein is defined by the most stable, ordered and functional structure, and the unfolded state is characterized by a lower stability, a lower degree of structure and a high entropy. The native state can be attained through a simple cooperative process where only two species are involved (two-state mechanism) or through a more complex pathway, with the formation of various intermediate states. The structural analysis of intermediates is limited by its transient nature and the few amount of them that can be detected or isolated during the folding process. Nevertheless under mild denaturation conditions (low pH or lower denaturing agent concentration) it is possible to identify equilibrium intermediates, as it has been shown for globular proteins like α-lactalbumin (Kuwajima, 1996) where partially folded species form a compact nucleus with native-like secondary structure and fluctuating tertiary structure called molten globule, described for the first time in the cytochrome c (Ohgushi & Wada, 1983). A main theme of debate is the nature of the intermediates, whether, once identified, they are on-pathway or off-pathway species. The on-pathway intermediate has native-like structure and participates directly to promote the attainment of the native state (U ↔ I ↔ N), in contrast the off- pathway intermediate is trapped in a way that cannot achieve the native state representing a kinetic dead-end (I ↔ U ↔ N) with a strong probability to lead a protein aggregation phenomenon (Baldwin, 1996; Jahn & Radford, 2008). Even though the importance of intermediates in folding pathways has been largely demonstrated (especially for small disulphide rich-proteins, as discussed in section 1.8), numerous works have also described the folding of many small monomeric proteins through a simple two-state kinetic mechanism indicating that intermediates are not essential elements for an efficient folding (Jackson, 1998). However, the use of new sensitive experimental techniques and the integration with novel molecular dynamic simulations has led to a vision where the intermediates can be seen as ubiquitous steps along the route to native state in such a way that the folding energy landscapes of even small proteins exhibit significant ruggedness in which intermediates stabilized by both native and non-native interactions are populated (Brockwell & Radford, 2007).

1.4 Folding models: How a protein can fold?

Several simplified models have been described to explain possible mechanisms by which a protein can properly fold.

First, the *nucleation/growth model* supposed that the folding limiting rate step is the formation of an earlier native nucleus constituted of secondary structure elements able to propagate gradually to the final native structure (Wetlaufer, 1973). Still, the absence of folding intermediates described in this model was clearly in contrast with the increasing number of these species studied in the 1970s and 1980s. Later, Karplus

& Weaver proposed the *diffusion collision model* that involves a hierarchical process where "microdomains" constituted predominantly by secondary structure elements are initially formed and coalesce into the native tertiary structure through a mechanism of diffusion or collision that corresponds to the rate-limiting step of the process (Karplus & Weaver, 1976, 1994). In the *hydrophobic collapse model*, the hydrophobic effect induces a rapid collapse of the polypeptide chain with the instauration of unspecific interactions that leads to the formation of intermediates and the consequent attainment of secondary and tertiary structure (Dill, 1985). Harrison and Durbin presented the *jigsaw puzzle model* where native-like local structures drive an efficient folding for compact proteins without the necessity of unique mechanism likewise there is not a unique manner to solve a jigsaw puzzle (Harrison & Durbin, 1985).

Afterwards, Alan Fersht proposed the *nucleation-condensation mechanism* based on the study of the chymotrypsin inhibitor 2 (CI2) folding pathway (Fersht, 1995; Fersht, 1997). In order to connect features of both nucleation and hydrophobic collapse models he suggested a folding nucleus structurally similar to the native state (molten globule) that collapses and achieves the native state by the simultaneous formation of both secondary and tertiary interactions. Therefore, the nucleation and hydrophobic collapse models are considered as two extremes of the nucleation–condensation mechanism (Daggett & Fersht, 2003). Nowadays, the nucleation-condensation and diffusion-collision models are considered different manifestations of a common unifying mechanism for protein folding, where proteins undertake a specific folding pathway depending on the propensity of its secondary structure or ability to form a folding nucleus (Nickson et al., 2013).

1.5 Folding and Energy Balance

Protein folding is directed by both kinetic and thermodynamic factors. In a simple two state mechanism (Figure I-1.2), the protein in its unfolded state U (energetically unfavourable) passes through the transition state (a maximum free energy state) to achieve its native conformation N that corresponds to the minimum of Gibbs free energy; this process is defined as a reversible reaction (1) described by the van't Hoff equation (2):

$$(1)\quad U \underset{K_U}{\overset{K_F}{\leftrightarrow}} N \qquad (2)\quad \Delta G_{N\text{-}U} = -RT \ln(K_F/K_U) = \Delta H - T\Delta S$$

Where G is the Gibbs free energy, H and S the enthalpy and entropy, respectively; R represents the gas constant, T is the temperature in Kelvins and K_F and K_U are the folding and unfolding rate constants (Privalov, 1979; Creighton, 1992).

Therefore ΔG_{N-U} is the free energy difference between native and unfolded state and represents the protein thermodynamic stability. Assuming that T is constant, protein stability results from the balance between entropy and enthalpy energies, where the first represents the "order" of the system in terms of flexibility and freedom degrees of the molecules, and the enthalpic factor corresponds to the formation of native interactions such as hydrogen bonds, ionic interactions, van der Waals forces and hydrophobic interactions.

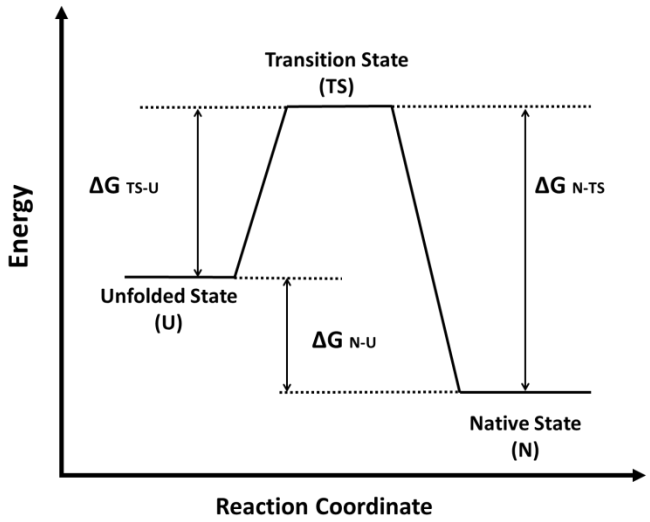

Figure I-1.2| **Energy diagram scheme.**

When the unfolded state converts into the native structure there is a gain in enthalpy due to the formation of native interactions (i.e. H-bonds and hydrophobic interactions) but a loss in entropy due to the acquirement of a highly ordered and compact structure. Therefore, it is possible to assert that a folded protein is in a thermodynamic compromise, where the entropic disadvantage is compensated by the numerous weak native interactions formed. Most of the folded proteins are only marginally stable since the energy barrier between native and unfolded state is relatively low, just 21-42 KJ/mol or 5-10 kcal/mol (Petsko & Ringe , 2004).

1.6 The principle of minimal frustration and the folding funnel theory

In the early 1990s the energy landscape theory was proposed to explore the protein conformational space. Indeed, protein folding is not a simply process that occurs through a defined series of intermediates, but it can be rather imagined as a folding funnel (Figure I-1.3) where a kinetic flow of an ensemble of metastable states, with gradually lower in energy, slopes down through uncountable routes to converge into a global minimum that represents the native state (Bryngelson et al., 1995). The

roughness of the folding funnel symbolizes numerous trapped-states that correspond to transient local free energy minimums along the potential surface, and the transition state ensemble or "bottleneck" is described as the rate-limiting step of folding process, since it represents the major energetic barrier that metastables species must surpass to achieve the native state. Lattice simulations have been employed to characterize deeply the nature of the transition state ensemble revealing the presence of a large number of native contacts (50%) contributing to its stabilization (Onuchic, 1996).

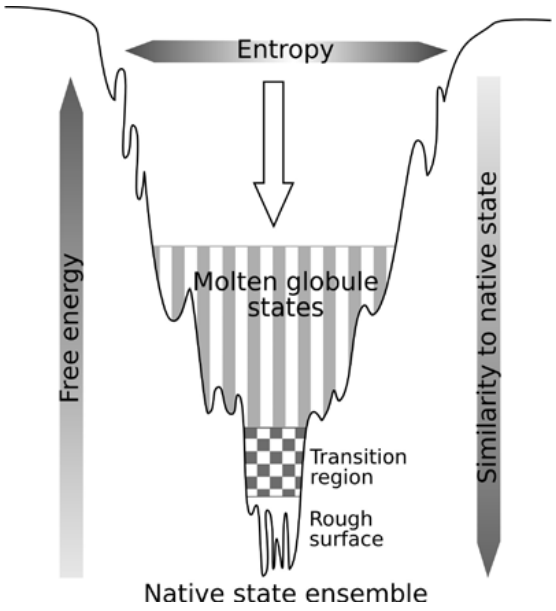

Figure I-1.3| **Folding funnel scheme (Xing-Lai & Shu-Qun Liu, 2011).**

This leads to the concept of minimal frustration. This principle states that the general energy of the protein decreases more than what may be expected by chance as the protein assumes conformations progressively more like the ground (native) state. In other words, there is a strong energetic bias toward the native basin that overcomes both the roughness of the landscape, which stabilizes kinetic traps, and also ultimately the entropy of the chain. Nature chooses only specific interactions that are not in conflict (minimally frustrated) and participate towards the attainment of the low-energy structure (Onuchic et. al, 1997; Onuchic & Wolynes, 2004). Thus the interplay between key residue interactions may decide for the rapid exploration of the native state through the formation of a folding nucleus that determines the native topology in the transition state ensemble (Lindorff-Larsen et al., 2005).

In the last decades, many approaches have been used to decipher the folding code through which the sequence of a protein can drive the correct search and attainment of the functional state. On one side, protein-folding experiments have been employed, such as ϕ-value analysis that assesses mutational effects on folding rates or temperature-jump methods to fast kinetic studies. On the other side, novel theoretical

and computational approaches have been developed including databases with an ever increasing number of protein structures and models, and new and powerful bioinformatic tools supported by the evidence that solving the folding problem could be commercially relevant (Dill et al., 2008; Dill & MacCallum, 2012).

In conclusion, since 50 years ago a lot of information about the protein folding problem has been compiled and used to understand the principles that govern this process. However, despite this effort there is still controversy on how proteins achieve their functionality. For instance, the starting point of folding problem is dictated by the Anfinsen's dogma and the evidence that the native state is in a global minimum of the Gibbs energy; nevertheless, nowadays many studies report the native state not being a real energy minimum but rather an intrinsically "metastable state", when confronted with highly thermodynamically highly stable states constituted by protein aggregates (Baldwin et al., 2011; Eichner & Radford, 2011).

1.7 Protein "non folding" problem: the IDP case

Since the first protein structure was solved by X-ray crystallography, more than 100000 structures have been deposited into the Protein Data Bank (Bernstein et al., 1977). During a long time it was assumed that proteins need to adopt a defined three-dimensional structure to perform their functions. Interestingly, in the last decade a great number of solved structures have shown missing electron densities related to portions of sequence with flexible or disordered regions. Genome analysis from different organisms has shown as 33% of eukaryotic proteins contain disordered regions based of more than 30 residues in length and involved in a series of cellular functions (Ward et al, 2004). These data suggest a new point of view about the structure-function paradigm and the beginning of an intense field of research focused on the study of intrinsically disordered proteins (IDPs). Hence, IDPs are proteins that do not fold into a tridimensional structure under physiological conditions; they can be divided into partially or totally unfolded proteins depending on whether they present only local disordered regions, such as DNA/RNA binding proteins, or they are unstructured along their entire lengths, such is the case of polypeptides like α-synuclein or the Tau protein (Wright & Dyson, 1999). Theoretical and experimental tools have allowed IDPs characterization. Structural techniques as X-ray crystallography, NMR spectroscopy and circular dichroism joined to bioinformatic methods have revealed the chemical properties of disordered regions to be different from that of globular proteins. First, Uversky and co-workers analyzed 275 protein sequences from the SWISS-PROT protein sequence data bank displaying totally or partially unfolded segments showing that low mean hydrophobicity combined with relatively high net charge in disordered regions would be an important prerequisite to sustain unfolded conformations under physiological conditions (Uversky et al., 2000). The analysis of the amino acidic content of disordered regions and globular proteins

indicates the presence of low-complexity sequences for IDPs, with a high hydrophilic amino acid content (Arg, Gln, Glu, Lys, Pro, and Ser) and a lower hydrophobic residue component (Cys, Ile, Leu, Phe, Trp, Tyr, and Val) (Romero et al., 2001). Thus, IDPs amino acidic sequence would encode for the "non folding event" since the absence of the hydrophobic core and the presence of the high net charge would preclude a spontaneous folding and leads to the formation of an ensemble of dynamic and fluctuating species with diverse conformational possibilities. It is reported that IDPs quickly interconvert as unfolded states of globular proteins at least until four different conformations: native (ordered), molten globule, pre-molten globule, and coil-like (Uversky & Keith Dunker, 2010). The lack of a rigid structure and the ability to adopt diverse conformations confers advantages to these proteins; in fact the presence of a great surface area of interaction and a dynamic conformation, due to the high IDPs flexibility, play a key role in the regulation of a lot of cellular functions, from splicing and post-translational modifications to signaling in cellular pathways. This makes the study of IDPs regulation an important area of study since the altered expression of IDPs is associated with many diseases (Babu et al., 2011; Wright et al., 2015).

1.8 Disulfide bridges and Folding

The introduction of disulfide bonds is considered as the most common and probably most ancient co-/post-translational modification. Disulfide bonds are covalent links between pairs of cysteine residues playing an important role in the folding and stability of proteins. It has been largely described how the presence of disulfide bridges results in a series of advantages, among them the most important is undoubtedly the increase of protein stability (Trivedi et al., 2009). In this way, several works have described how the introduction of new disulfide bridges can improve protein stability, protect from damage by proteolysis and increase protein half-life (Zavodszky et al., 2001; Takagi et al., 1990). Interestingly, although it was initially believed that the presence of disulfide bonds in mature proteins was only an evolutionary strategy to gaining stability, it has been demonstrated how disulfide bonds do not represent inert structural elements since they may possess also a functional role. Therefore, for instance, the reversible cleavage of disulfide bonds by catalysts in the extracellular environment may control the secretion of soluble proteins and play a role in cell-surface receptor regulation (Hogg, 2003).

The process involving the formation of native disulfide bonds, *disulfide bond regeneration*, and the final native conformation acquirement or *conformational folding* is defined as oxidative folding (Wedemeyer et al., 2000). Disulfide bond regeneration proceeds by a complex mechanism called thiol/disulfide exchange including a coupled redox reaction (Figure I-1.4). When a thiolate anion is located at an adequate distance from a redox agent or other disulfide in the same protein (*disulfide reshuffling*), it can

delocalize its negative charge between the three sulphur atoms with the concomitant formation of a mixed-disulfide, which leads to sulfhydryl oxidation and reduction of the starting disulfide (Sevier & Kaiser, 2002).

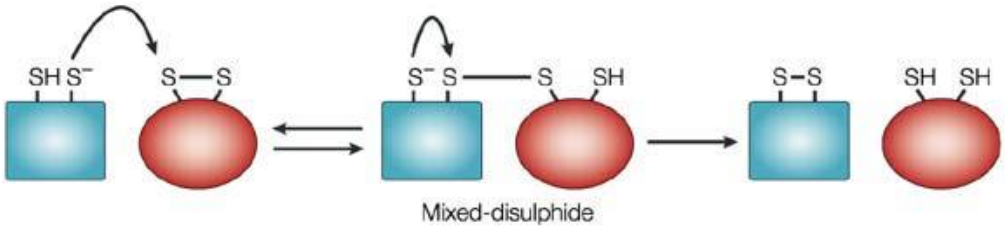

Figure I-1.4| **Thiol disulfide between a protein (blue) and a redox agent (red) (Sevier & Kaiser, 2002).**

In eukaryotic cells, the formation of disulfide bridges occurs in the endoplasmic reticulum (ER) by specialized enzymes: protein disulfide isomerases (PDIs), while in prokaryotic cells oxidoreductases belonging to the Dsb protein family are responsible for this process in the periplasmic space. Protein disulfide isomerases catalyse thiol/disulfide exchange reactions and also help to break (reduction) or rearrange (isomerization) incorrect disulfide bonds. After a thiol/disulphide exchange cycle the enzyme active site cysteines are regenerated by other protein or with a redox agent such as glutathione (Freedman, 1989; Ellgaard & Ruddock, 2005). The thiol/disulphide exchange rate is strictly conditioned by the cellular environment, accordingly, disulfide formation is not promoted in the exceedingly reducing cytosolic ambient and disulfide containing proteins are modified only in the ER (an oxidizing environment) and generally secreted as extracellular proteins. Effectively, cysteine reactivity plays an important role in the formation of disulfides since an insufficient proximity, the cellular pH or a steric impediment by the tertiary structure would prevent the reaction. The environment conditioned reactivity of cysteine residues has been used to characterize the different species (i.e. trapping intermediates) occurring during the folding process (as discussed below).

The bovine pancreatic ribonuclease A (RNase A) has been largely used as protein model to study the stability conferred by disulfide bridges. The protein stability increase is initially attributed to a destabilization of the unfolded state due to a loss of entropy caused by the molecular constraint generated with the formation of the disulphide bond. Furthermore, the entropic factor is not the only implicated in the stability gain but also the instauration of obligatory favourable interactions in the hydrophobic core may contribute enthalpically to the stabilization of the native state (Betz, 1993; Wedemeyer et al., 2000).

In order to characterize the mechanism of folding, more than 30 small disulfide-rich proteins have been analysed in the last 20 years using the technique of the oxidative folding. This approach is based on the use of disulfide bond and/or reactive cysteines

to follow the formation of transient folding intermediates which are difficult to characterize with classical studies of conformational folding. Initially a reducing agent (i.e. DTT) and a chemical denaturant (i.e. urea) are used to unfold the protein and the incubation of the reduced and denatured protein in oxidized conditions in the absence of denaturant leads to spontaneous protein refolding into the native state. The use of this technique allows the characterization of trapped intermediate species by quenching the reaction with acid or a chemical agent (alkylation) and a further analysis by chromatography (reverse phase HPLC). The acid-trapped method has been widely used to study the role of kinetic intermediates in the folding mechanism, since the rates of disulfide formation can be easily adjusted by modification of the oxidizing environment (i.e. with the addition of a redox agent) (Darby & Creighton, 1997). The study of all disulfide protein models revealed an important variety of oxidative folding pathways defined by the heterogeneity of intermediates accumulated during the process. Understanding the intermediate structural nature is fundamental, since it may select for an specific oxidative folding pathway. The intermediates can be defined depending on the accessibility of their disulfide bonds and/or thiol groups. *Disulfide-insecure* intermediates are dead-end metastable species which disulfide bonds are exposed as their thiol groups promoting reshuffling rather than oxidation. *Disulfide-secure* intermediates are defined as productive intermediates with buried disulfide bonds and exposed thiol groups favouring the oxidation and native state acquisition (Narayan et al., 2000). The oxidative folding pathways can be classified according to (i) the amount of heterogeneous folding intermediates, (ii) the prevalence of native-like intermediates with native disulfide bonds, and (iii) the presence of intermediates containing at least two non-native disulfide bonds. According to these criteria there exist two extreme folding models. The first model is based on the bovine pancreatic trypsin inhibitor (BPTI) study where a few number of productive intermediates with exclusively native disulfide bonds are formed. In contrast, the second model corresponds to that of the hirudin folding pathway, characterized by a primary non-specific packing of a large number of heterogeneous intermediates followed by the formation of scrambled isomers, which ultimately reshuffle into the native protein. In the case of BPTI and BPTI-like proteins, the key event is represented by the acquisition of a stable tertiary structure that buries the native disulfide bonds (locking in) leading to the native state by the oxidation of free thiol groups in these productive intermediates. On the opposite, in the case of hirudin and hirudin-like proteins the limiting step is represented by native disulfide formation that drives the high heterogeneous ensemble of scrambled intermediates toward the native state. However, mechanisms between these two extreme models have been reported, where both pathways are coupled together showing similitudes and divergences to BPTI and hirudin mechanism (Arolas et al., 2006; Chang, 2011).

2 Protein Oxidation by Reactive Oxygen Species (ROS)

In the last few years, oxidative modifications are receiving increasing interest, especially the Reactive Oxygen Species (ROS) effect upon proteins, nucleic acids and lipids. ROS are defined as reactive chemical compounds containing oxygen. The ROS term includes superoxide anion, hydrogen peroxide, hydroxyl radical and a non-radical ROS such as singlet oxygen, an excited form of oxygen, where one of the electron jumps to a superior orbital (Turrens, 2003). During their lifetime all organisms are continually exposed to ROS that are produced by a large number of metabolic processes, such as aerobic mitochondrial metabolism, reduced electron carriers autoxidation, inflammatory reactions, oxidase catalysed reactions, glycation/glycoxidation reactions, and metal catalysed reactions. Furthermore, it has been demonstrated that environmental factors such as irradiation or contaminants in the atmosphere also contribute to ROS generation (Stadtman & Levine, 2000). Cellular excess of pro-oxidant species, a phenomenon known as oxidative stress, leads to dangerous modification of cellular biomolecules underlining the importance of controlling oxidative homeostasis in the cell (Rahal et al., 2014). Cells have developed a diversified set of mechanisms to prevent, control and repair oxidative damages. They dispose of a complex battery of antioxidants that interconverts ROS to unreactive derivatives. Antioxidants include enzymes such as superoxide dismutases, catalases and glutathione (GSH) peroxidases, and non-enzymatic molecules as phenolic compounds (tocopherol), vitamins or free fatty acids (Sies, 1996). Cells use a multi-complex degradation system to remove irreversible oxidized proteins. This cellular degradation system includes lysosomal systems, mitochondrial and calcium-dependent proteases, and the 20S proteasome, an evolutionary ancient system considered to be the main oxidized proteins removing system (Jung et al., 2014). Hence, the correct balance (Figure I-2.1) between pro-oxidant, antioxidant activities and degradation systems controls the intracellular accumulation of oxidized species. Protein oxidation is the most recurrent modification mediated by ROS since proteins are the major group of macromolecules in the cell. Proteins can be modified covalently either on backbone or on amino acid side chains (Berlett & Stadtman, 1997). All amino acids are potential targets for ROS oxidation, but methionine and cysteine residues, both containing susceptible sulfur atoms, show particularly sensitivity to almost all ROS forms. Methionine can be oxidized to a sulfoxide and cysteine to sulphide or sulfenic, sulfinic, and sulfonic acids, and these oxidized products are reduced enzymatically through specialized reductases with the exception of sulfonic acid.

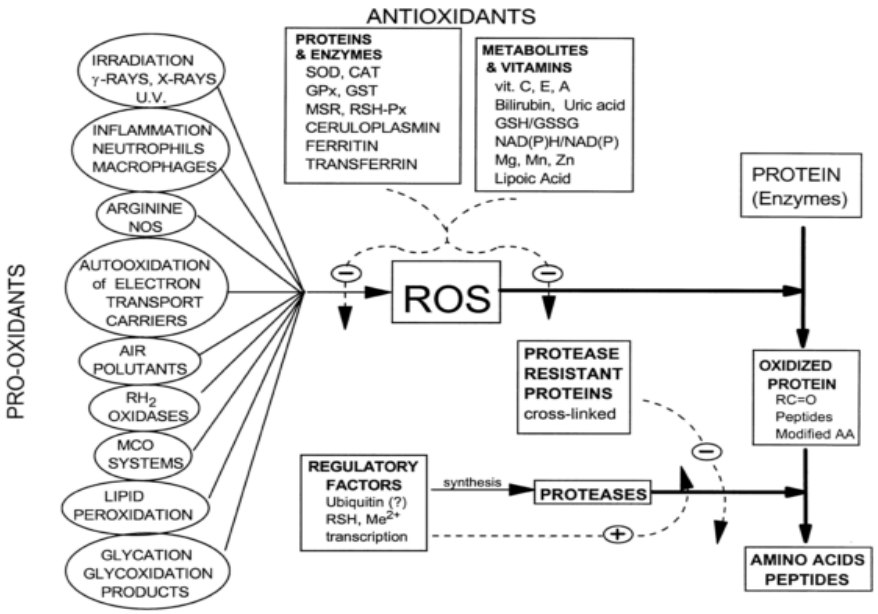

Figure I-2.1| **Factors affecting reactive oxygen species (ROS) and the level of oxidatively modified proteins (Berlett & Stadtman , 1997**).

In biological systems these are the only oxidative modifications that can be repaired, hence it has been proposed that sulfur-containing aminoacids may function as cellular antioxidants working as “built-in” ROS scavenger systems to protect proteins from more extensive irreversible oxidation (Levine et al., 1996).

On the other hand, oxidation of specific amino acids such as, lysine, arginine, proline and histidine, and protein backbone cleavage at proline, glutamate and aspartate residues lead to irreversible formation of protein carbonyl groups (aldehydes and ketones). Due to their chemical stability, protein carbonyl groups have been used as a ROS-mediated protein oxidation marker related with aging, oxidative stress and several human diseases such as Alzheimer’s disease (AD) (Dalle-Donne et al., 2003; Smith et al., 1991).

Oxidized proteins degradation plays a key role in maintaining cellular homeostasis, and many studies have been carried out dealing with protein oxidation and its effect upon degradation systems. It has been described that under mild oxidation (stage I, Figure I-1.6) proteins lose partially their function but maintain their structures intact. When oxidative pressure increases proteins change their structure (exposing hydrophobic regions) and are recognized by the degradation machinery (stage II, Figure I-2.2) which works efficiently to remove damaged proteins. The proteolytic system efficacy, however, is affected when oxidative stress overcomes its degradation rate and unfolded oxidized species initiate aggregational processes with further accumulation of toxic aggregates (stage III, Figure I-1.6) (Nowotny et al.,2014).

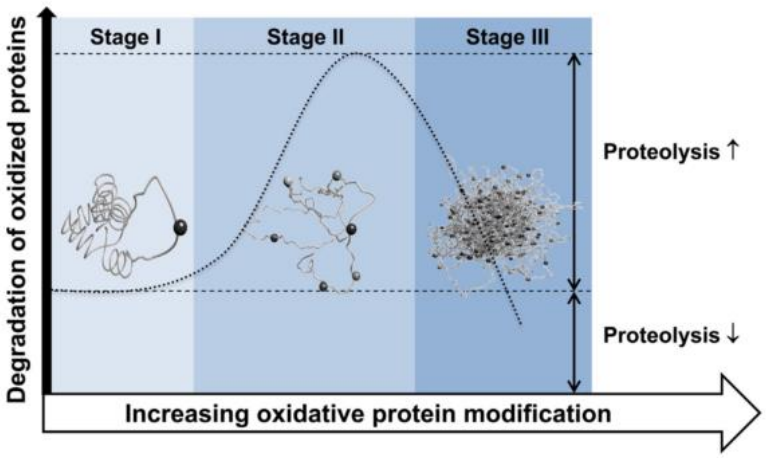

Figure I-2.2| **Protein oxidation versus protein degradation systems.** Under mild oxidative conditions structure of proteins is minimally modified (stage I) and it is not required the degradation machinery intervention. When oxidative level increases proteins lose their structures and degradation system removes modified proteins efficiently (stage II). In stage III, under high oxidative stress proteolysis does not degrade all damaged species accumulated in the cell (Nowotny et al., 2014).

During the past decade many works have been addressed the controversial role of ROS in the cell. On one side, it has been described that the damaging effect of these species results in multiple diseases including cancer, cardiovascular diseases, metabolic disorders, neurodegenerative pathologies and premature aging. On the other side, recent evidences suggest ROS involvements in the functional regulation of cellular pathways. ROS would participate as signalling molecules to control different physiological mechanisms as the regulation of neuronal ion channels, kinases, and transcription factors (Thomas et al.,2007; Brieger et al., 2012) or the possible implication in the apoptosis cascade (Zorov et al., 2005). Clearly, understanding the ROS control mechanisms requires the analysis of multiple variables, where undoubtedly the amount of cellular ROS seems to have a crucial role. Hence, lower ROS concentrations are necessary for physiological cell regulation whereas higher levels spring irreversible oxidative modifications (Sena & Chandel, 2012).

2.1 Cysteine Oxidation

Cysteine residues, as discussed above, are one of the major ROS targets. Even if these residues are one of the less abundant amino acids (1-2%) in organisms, they take part in numerous cellular functions: from protein stability to catalysis, regulation and cofactor-binding, among other. The high thiol group reactivity and chemical plasticity results fundamental to direct different cysteine functions as well as ROS oxidation. Specifically, the reactivity of thiol groups depends on several factors such as the solvent exposure, pH or protein microenvironment that would change the pKa values and cysteine nucleophilicity (Marino & Gladyshev, 2012; Roos, Foloppe & Messens, 2013).

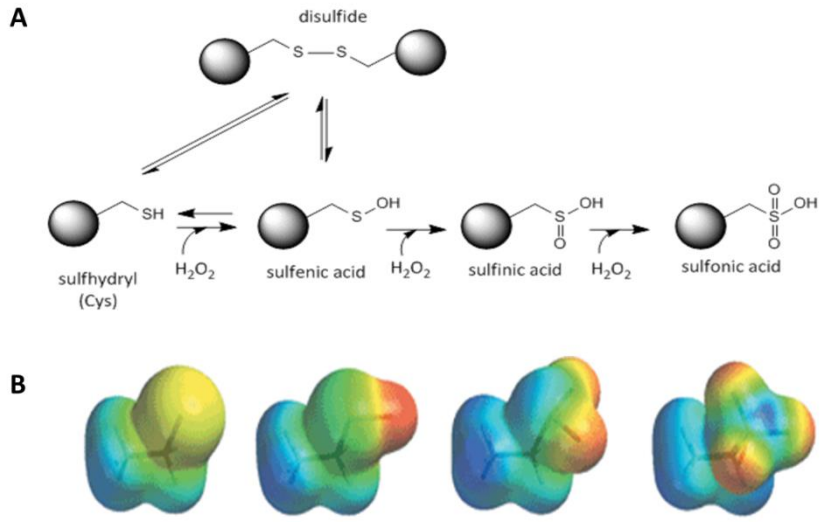

Figure I-2.3|**Cysteine oxidation.** (A) Cysteine redox and oxidation mechanism in presence of ROS (hydrogen peroxide) (B) Electrostatic potential surfaces of cysteine thiolate (from left), sulfenate, sulfonate and sulfonate (Red is most negative; blue is most positive) (Yarnell, 2009).

During cysteine oxidation (Figure I-2.3A), a thiol group is attacked by a ROS such as hydrogen peroxide forming the first short life-time intermediate, cysteine sulfenic acid (–SOH), which quickly converts to disulfide bond (Cys-S-S-Cys) or sulfinic (–SO2H) and sulfonic (–SO3H) acids (Yarnell, 2009). Sulfinic acid is a relatively stable species reversibly reduced to sulfhydryl by sulfiredoxin (Rhee et al., 2007) or oxidized by ROS to form sulfonic acid, the final irreversible species of cys-oxidation. Accordingly, as shown in Figure I-2.3B, the sulphur negative charge increases proportionally to the oxidation state indicating the relative reactivity of these species.

Direct and indirect approaches have been used to identify oxidative cysteine modifications. The majority of indirect methods analyse changes in the cysteines redox state monitoring, for example, the binding to thiol-specific alkylating agents such as N-ethylmaleimide (NEM) or iodoacetamide (IAM) which can be detected by the decrease in the signal when the probe is oxidized. Other indirect technique is based on the detection of oxidation by restoration of thiol-labelling by reducing agents. In this case, a previous step generates the irreversible sulfhidril groups alkylation, followed by reduction of oxidized forms and derivatization with detectable specific alkylating agents. However, this type of approach is limited to cell lysates or purified proteins analysis. Thus, new methodologies have been recently introduced where specific cysteine oxoforms are directly detected. In this case, chemo-selective alkylating agents have been used for specific cysteine oxoforms, such as 5,5-dimethyl-1,3-cyclohexanedione (dimedone) that recognizes sulfenic acid forms (Leonard & Carroll, 2011). The sulfenic acid-dimedone complex is detected directly by mass spectroscopy or by conjugation with biotin or fluorophores allowing a faster identification of oxidized forms (Kim et al., 2014; Tyther et al., 2010).

Thanks to the increasing number of detecting techniques, protein sulfenylation has been identified in a large set of proteins, from enzymes as kinases, phosphates, peroxidases, oxidoreductases, cysteine proteases, transferases to proteins as transcription factors, chaperones, apoptotic regulators, oxygen carriers and serum proteins. The biological meaning of these post translational modifications has been largely studied in the last years showing that reversible cysteine oxidation to sulfenic and sulfinic acid plays a regulatory role in the cell influencing protein activity and taking part in the signalling pathways (Lo Conte & Carroll, 2013). The presence of transient sulfenic acid in the enzymes active site may temporally regulate activation/inhibition mechanisms or lead to changes in the protein conformation and modulate protein activity as in the case of the oxidation-dependent enhancement of tyrosine kinase activity in the receptor of epidermal growth factor (Paulsen et al., 2011). Protein sulfinylation has been considered for a large time as an artefact of protein purification, however recent evidences have been reported about its biological role in the cell. The absence of robust methods for sulfinic form detection has allowed its identification only in relatively few proteins, mostly peroxidases and sulfiredoxins, indicating the implication of protein sulfinylation mainly in protective mechanisms against oxidative stress, such is the case of the Parkinson's disease protein DJ-1, where the formation of sulfinic acid at cys106 directs the re-localization of the protein, avoiding the cellular death (Canet-Avilés et al., 2004).

Irreversible cysteine oxidation under oxidative stress conditions leads to sulfonic acid formation. An increasing number of works are reporting the toxic effect of this hyper-oxidized species in the cell. For example, parkin protein sulfonation alters its E3 ligase activity, promoting the formation of Lewy bodies (LBs), amyloid aggregates found in Parkinson's disease (PD) (Meng et al., 2011).

3 Protein Aggregation

3.1 Protein Misfolding and Cell Response

The crowded cellular environment requires a tight control of folding to avoid the population of kinetic traps, misfolded or partially unfolded species that can culminate in the accumulation of potentially toxic aggregates (Golding & Cox, 2006; Ellis, R. J., & Pinheiro, 2002). A critical equilibrium exists between favorable and unfavorable conformational arrangements during the folding process since, in the cell, continuous molecular collisions promote a strong competition among native intra and non-native inter-molecular interactions; these latter leading to the irreversible assembly of

thermodynamically highly stable aggregated structures located in a deep free energy minimum (Figure I-3.1).

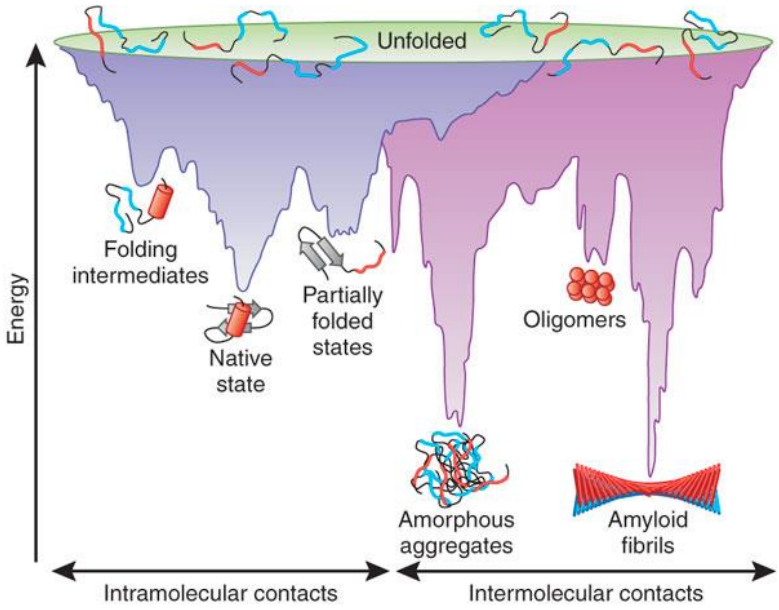

Figure I-3.1|**Energy state of protein folding under physiological and misfolding conditions.** The shape of the graph shows energy state of the protein conformations moving toward its native or misfolded condition through multiple inter- and intramolecular contact arrangements (Hartl & Hayer-Hartl, 2009)

To prevent aberrant aggregation, cells have evolved a complex machinery of molecular chaperones that help and assist protein synthesis across ribosomes and support efficient folding. Molecular chaperones control folding kinetics and inhibit events that may lead to unproductive end points, resulting in aggregational reactions. Chaperones are ubiquitous proteins that cooperate with nascent polypeptide chains during synthesis and act on translocation to different cellular compartments. The "Protein Quality Control" (PQC) system is composed by molecular chaperones and the ubiquitin proteasome system (UPS), which recognizes and binds, via non-covalent interaction, proteins with abnormally solvent exposed hydrophobic residues and targets them to the degradation pathway (Hartl at al., 2011; Berke & Paulson, 2003). Cells express constitutively a chaperone complex network and their expression is increased significantly under environmental stress conditions, as for example in hyperthermia, hypoxia, oxidative stress or exposure to toxins (Gidalevitz, T. et al., 2011). In eukaryotic cells, when the excess of misfolded proteins overcomes the refolding capacity of molecular chaperones, these aberrant species are transported to the perinuclear zone, the centrosome, and assembled with different components of the UPS to form dynamic structures called aggresomes, which are considered as a cytoprotective strategy to recruit and favour the elimination of these otherwise toxic species (Olzmann, J. et al., 2008; Johnston, J et al., 1998). Under normal conditions molecular

chaperones are implied in protein homeostasis. A strict control of proteostasis networks by the PQC machinery results fundamental for the organism health, while its deregulation during aging becomes one of the main causes of cellular dysfunction and disease onset (Kikis et al, 2010).

3.2 Conformational Diseases

Despite the PQC machinery usually works efficiently, failures in protein folding may occur and culminate in pathologic conditions known as *Protein Misfolding or Protein Conformational diseases*. In the last decades a wide range of disorders have been associated to misfolding events in which a polypeptide chain is not able to acquire or preserve its functional native conformational state. The inability to maintain the correct conformation often leads to toxic gain of function, such in the case of amyloid diseases, characterized by the deposition of particular toxic aggregates in tissues and organs. Misfolding processes may also result in a loss of function, for example in cystic fibrosis, where pathologic states are caused by a direct decrease in the active protein quantity.

Conformational diseases are distinguished, according to the localization where protein aggregates are deposited, (Table I-1) into (a) neurodegenerative, with aggregation occurring in the brain (Alzheimer's (AD) and Parkinson's (PD) diseases), (b) non neuropathic localized amyloidosis, where aggregates are found in a single type of tissue except the brain (types II diabetes), and (c) non neuropathic systemic amyloidosis with the presence of aggregates in multiple tissues (Senile systemic amyloidosis). Conformational diseases can be classified as sporadic (85%) or hereditary (10%) forms, whereas 5% are transmissible in humans as well as in other mammals such is the case of spongiform encephalopathies, also defined as *prion* diseases (Chiti & Dobson, 2006). These latter diseases are associated with the prion protein (PrP) which existence was suggested for the first time by Prusiner in 1982 to describe the unprecedented infectious pathogen implied in important neurodegenerative diseases, such as the progressive dementia caused by Creutzfeldt-Jakob Disease (CJD) and ataxic illnesses of scrapie of sheep and bovine spongiform encephalopathy (BSE). Prp, a constituent of normal mammalian cells, can convert in its isoform PrP^{sc} through a conformational change involving the loss of its α-helical and coil structure and the acquisition of a β-sheet enriched structure. The isoform PrP^{sc} becomes the transmissible agent since it self-propagates in cells and transmits its conformation to other native proteins (Prusiner, 1998). Interestingly, recent reports and clinical observations suggest that cell-to-cell propagation of amyloid aggregates in Parkinson's, Alzheimer's and Huntington's brain diseases is mediated by a prion-like transmission mechanism. In this way, simultaneous studies published in 2008 have shown the typical amyloid aggregates associated to Parkinson's disease (PD) or Lewy bodies (LBs) in originally healthy nerve cells transplanted into brains affected by PD, hence

demonstrating the transmission of the disease from host to grafted cells (Brundin et al., 2010 ; Herva & Spillantini, 2014).

Table I-1|**Human diseases related with formation of extracellular or intracellular amyloid deposits. Adapted from (Chiti & Dobson, 2006).**

Disease	Precursor protein or peptide	Native structure of protein or peptide[1]
a - Neurodegenerative diseases		
Alzheimer's disease	Amyloid β peptide	Natively unfolded
Spongiform encephalopathies	Prion protein or fragments thereof	Natively unfolded and α-helical
Parkinson's disease	α-Synuclein	Natively unfolded
Dementia with Lewy bodies	α-Synuclein	Natively unfolded
Frontotemporal dementia with Parkinsonism	Tau	Natively unfolded
Amyotrophic lateral sclerosis	Superoxide dismutase 1	All-β, Ig like
Huntington's disease	Huntingtin with polyQ expansion	Largely natively unfolded
Spinocerebellar ataxias	Ataxins with polyQ expansion	All-β
Spinocerebellar ataxia 17	TATA box-binding protein with polyQ expansion	α+β, TBP like
Spinal and bulbar muscular atrophy	Androgen receptor with polyQ expansion	All-α
Hereditary dentatorubral-pallidoluysian atrophy	Atrophin-1 with polyQ expansion	Unknown
Familial British dementia	ABri	Natively unfolded
Familial Danish dementia	ADan	Natively unfolded
b - Nonneuropathic systemic amyloidosis		
AL amyloidosis	Immunoglobulin light chains or fragments	All-β, Ig like
AA amyloidosis	Fragments of serum amyloid A protein	All-α, unknown fold
Familial Mediterranean fever	Fragments of serum amyloid A protein	All-α, unknown fold
Senile systemic amyloidosis	Wild-type transthyretin	All-β, prealbumin like
Familial amyloidotic polyneuropathy	Mutants of transthyretin	All-β, prealbumin like
Hemodialysis-related amyloidosis	β2-microglobulin	All-β, Ig like
ApoAI amyloidosis	N-terminal fragments of apolipoprotein AI	Natively unfolded
ApoAII amyloidosis	N-terminal fragment of apolipoprotein AII	Unknown
ApoAIV amyloidosis	N-terminal fragment of apolipoprotein AIV	Unknown
Finnish hereditary amyloidosis	Fragments of gelsolin mutants	Natively unfolded
Lysozyme amyloidosis	Mutants of lysozyme	α+β, lysozyme fold
Fibrinogen amyloidosis	Variants of fibrinogen α-chain	Unknown
Icelandic hereditary cerebral amyloid angiopathy	Mutant of cystatin C	α+β, cystatin like

Disease	Precursor protein or peptide	Native structure of protein or peptide[1]
c - Nonneuropathic localized diseases		
Type II diabetes	Amylin, also called islet amyloid polypeptide (IAPP)	Natively unfolded
Medullary carcinoma of the thyroid	Calcitonin	Natively unfolded
Atrial amyloidosis	Atrial natriuretic factor	Natively unfolded
Hereditary cerebral haemorrhage with amyloidosis	Mutants of amyloid β peptide	Natively unfolded
Pituitary prolactinoma	Prolactin	All-α, 4-helical cytokines
Injection-localized amyloidosis	Insulin	All-α, insulin like
Aortic medial amyloidosis	Medin	Unknown
Hereditary lattice corneal dystrophy	Mainly C-terminal fragments of kerato-epithelin	Unknown
Corneal amylodosis associated with trichiasis	Lactoferrin	α+β, periplasmic-binding protein like II
Cataract	γ-Crystallins	All-β, γ-crystallin like
Calcifying epithelial odontogenic tumors	Unknown	Unknown
Pulmonary alveolar proteinosis	Lung surfactant protein C	Unknown
Inclusion-body myositis	Amyloid β peptide	Natively unfolded
Cutaneous lichen amyloidosis	Keratins	Unknown

[1] According to Structural Classification Of Proteins (SCOP)

3.3 Amyloid fibrils

The previously mentioned Amyloid diseases are characterized by the presence of a particular type of protein aggregate: the amyloid fibril. The word “amyloid” was introduced it in the medical literature for the first time in 1854 by the physician and scientist Rudolph Virchow. He described the same color reaction with iodine and sulfuric acid (a change from brown to blue) for nerve deposits than the typical on starch, concluding that they were starch-like deposits and gave them the name *amyloid* from the Latin *amylum* (starch). In 1859, amyloid true nature was figured out when Friedreich and Kekule verified both the presence of protein in a “mass” of amyloid and the apparent absence of carbohydrate based on the high nitrogen content (Sipe & Cohen, 2000). All amyloid fibrils present a great similarity and share common general features although proceeding from diverse and unrelated precursor proteins implicated in different amyloid diseases. Accordingly, amyloid fibrils have been described as insoluble, resistant and highly ordered aggregates with a typical cross β-structure. The term cross β is used due to the characteristic "cross” pattern showed in the X-ray diffraction analysis (Figure I-3.2A), consisting of two scattering diffraction signals at 4.7 and 10-11 Angstroms arranged longitudinally and transversally to the fibril axis corresponding to the hydrogen bonding distances between intermolecular β-

sheet and side chain packing between the sheets (Figure I-3.2B) (Eanes & Glenner, 1968; Makin & Serpell , 2005). Electronic (EM) and atomic force microscopy (AFM) analysis show long fibrillar structures aggregated side-by-side ranging between 6-10 nm in diameter and composed by 2-6 subunits called "protofilaments" which arrange themselves parallel to each other, longitudinal or slightly oblique to the long fibril axis (Figure I-3.2C). The different disposition of these substructures leads to distinct fibril morphologies like as striated ribbons or twisted paired fibrils as observed in Aβ1-40 fibrils (Shirahama & Cohen, 1967; Malinchik et al., 1998).

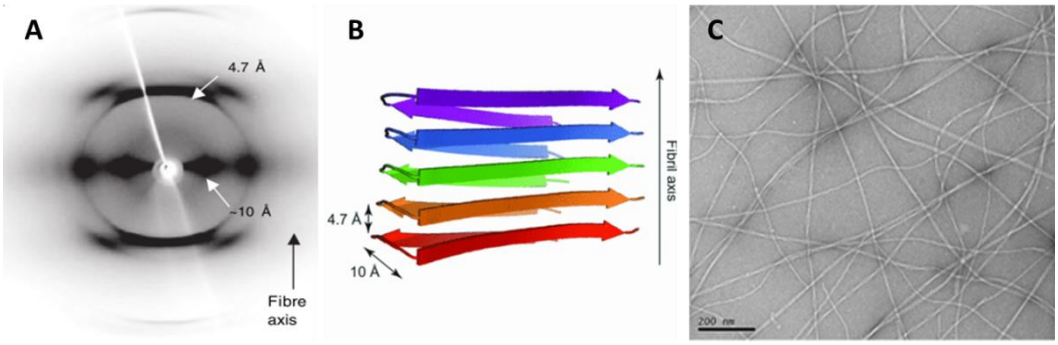

Figure I-3.2|**Structural and morphological properties of amyloid fibrils**. (A) X-ray diffraction pattern from aligned IAPP amyloid fibrils (Makin & Serpell, 2005). (B) Structure of the cross-βcore of amyloid fibrils, derived from NMR data for the Alzheimer's peptide, Aβ (Sawyer et al., 2012). (C) TEM micrograph of Aβ peptide amyloid fibrils (Morris & Serpell, 2010).

Recently, new methodologies such as solid-state NMR (ssNMR) and X-ray micro-crystallography have allowed the β-cross core characterization at atomic level. β–sheets have been reported to be tightly associated forming compact dehydrated interfaces defined as steric-zippers. They has been described up to eight models of steric zippers depending on the the face and strand orientation within the β–sheets (Tycko, 2011; Sawaya et al., 2007; Colletier et al., 2011). Different arrangements impact directly the fibril morphology and depend on several factors such as the amino acidic sequence or assembly conditions as it will be discussed in section 3.6. The characteristic amyloid fibrils structure allows the specific binding of chemical compounds, such as Thioflavin-T, a fluorescent dye that exhibits enhanced fluorescence emission at 482 nm when interacts with amyloid aggregates (LeVine, 1993), and Congo Red, that shows the specific red shift in the absorption maximum from 490 to 540 nm in presence of amyloid fibrils (Holde et al., 1961; Howie & Brewer, 2009). The conformational change from native structure into β-amyloid structure implies a strong shift in secondary structure that can be followed by far-UV circular dichroism and fourier transform infrared (FTIR) spectroscopy, displaying distinctive signals at ≈ 218 nm and at ≈ 1625 cm^{-1}, respectively (Byler et al., 1986; Calero & Gasset, 2004; Sabate & Ventura, 2013; Natalello & Doglia, 2015).

3.4 Functional Amyloid

For a long time, amyloid deposits have been only described as "folding mistakes" associated to a large list of diseases. This view has been revolutionized in the last decade thanks to an increasing number of studies reporting amyloid-like structures with important functional roles in organisms ranging from bacteria to mammals (Fowler et al., 2007). Hence, it has emerged the idea that the propensity to form amyloid structures can be considered as an intrinsic property of each polypeptide chain, although it is controlled by several factors such as the sequence composition (as will be discussed in section 3.6). Living systems largely exploit this generic ability for different purposes: from structural and protective functions to regulative ones (Table I-2). Nature uses the cross- β self-organizing code in many structural settings where a strong but malleable material is needed; such in the case of bacterial biofilms. *Escherichia coli*, *Salmonella*, *Citrobacter* and *Shewanella* microbes produce curli fimbriae, amyloid-like structures that in association with exopolysaccharides lead to the formation of the biofilm extracellular matrix. The presence of amyloid fibrils generates a robust fibrillary scaffold resistant to degradation that is used by the bacteria to promote surface adhesion and host colonization (Barnhart & Chapman, 2006). Another, example of functional amyloid used due its excellent mechanical properties has been reported in insects and spiders during silk production, where soluble proteins are accumulated in the spider gland at high concentration (≈50% w/v) before starting an amyloid-like assembly (Kenney et al., 2002; Hijirida et al., 1996). An additional interesting strategy is based on the switch between amyloid and non-amyloid forms to control specifc functions, such in the case of the bactericidal peptide Microcin E492 of *Klebsiella pneumonia* or the phenol soluble modulins of *S. aureus* which show cytotoxic and anti-microbial activity when released in monomeric forms but also assemble into amyloid fibrils to form bacterial biofilm (Bieler et al., 2005; Schwartz et al., 2012). This ability has been also described in *yeast* where "infectious prions" Sup35 and Ure2p shift between soluble and amyloid forms to enable the transfer of their cytoplasmic amyloid content to the progeny during cell division, engendering an epigenetic non-Mendelian type of inheritance that may promote survival in fluctuating environments, facilitating the evolution of new adapted phenotypes (Halfmann et al., 2012). In mammals, the transmembrane glycoprotein PMEL (Pmel17) is processed via proteolytic cleavage in a post-Golgi compartment and participates in melanin formation producing an amyloid-like fibrillary matrix that acts as a scaffold to accelerate the synthesis of the pigment (Watt et al., 2013). The use of amyloids confers an important tool to execute many functions but a strict regulation of the assembly process is fundamental, since amyloids could become extremely toxic to cells. In this way, different regulatory mechanisms have been described, from a highly operon regulated-pathway in curli proteins to a rigorous control by chaperones of

yeast prions or a control of proteolytic cleavage in Pmel17 (Barnhart & Chapman, 2006; Dueholm et al., 2012; Otzen, 2010a).
Natural functional amyloid fibrils have inspired the use of amyloids in nanotechnology since their physicochemical properties convert them in an incredibly strong and robust material that can be used in a wide range of technological applications (Sawyer et al., 2012).

Table I-2|**Proteins forming naturally non pathological amyloid-like fibrils with specific functional roles, adapted from (Pham, Kwan & Sunde, 2014).**

Species	Protein	Function
Bacteria		
E. coli	Curli	Biofilm formation and host invasion
Salmonella spp	Curli	Biofilm formation and host invasion
S. coelicolor	Chaplin	Modulation of water surface tension
Pseudomonas spp.	FapC	Biofilm formation
Xanthomonas spp. and other plant pathogens	HpaG	Plant pathogen virulence factor
K. pneumonia	Microcin E492 (Mcc)	Storage and regulation of toxicity
Bacillus subtilis	TasA	Biofilm formation
Streptococcus mutans	P1	Adhesion and host interaction
S. aureus	Modulin	Toxin and biofi lm stabilization
M. tuberculosis	Mtp	Pili formation
Unicellular eukaryotes		
Plasmodium falciparum	MSP2	Erythrocyte invasion*
Fungi		
Podospora anserina	HET-s	Regulation of heterokaryon formation
S. cerevisiae	Ure2p	Regulation of nitrogen catabolism
S. cerevisiae	Sup35p	Regulation of stop-codon read through
S. cerevisiae	Nsp1	Control of nucleocytoplasmic mixing at nuclear pore
Candida albicans	Fungal adhesin	Cell adhesion
S. cerevisiae	Fungal adhesin	Cell adhesion
Most filamentous fungi	Hydrophobin	Coat formation and modulation of adhesion and surface tension
Plants		
Hevea brasiliensis	REF	Biosynthesis of natural rubber
Animalia (excluding mammals)		
Insects and fish	Chorion protein	Structural and protection

Species	Protein	Function
Nephila clavipes	Spidroin	Structural
A. diadematus	Spidroin	Structural
C. flava	Silk protein	Structural
M. signata	Silk protein (MalXB)	Structural
A. californica	CPEB	Memory storage
P. dacnicolor	aDrs	Host defence*
Mammals		
Homo sapiens	Pmel17	Melanin synthesis
H. sapiens	Peptide hormone	Storage and controlled release
H. sapiens	RIP1/RIP3	Regulators of necroptosis
Mus musculus and probably other mammals	CRES and cst8	Sperm maturation and maintenance of the luminal milieu

*Function is putative or it is unclear if the amyloid has a function in vivo

3.5 Mechanisms of Aggregation

As described above, proteins adopt their functional state driven by kinetically favourable intramolecular interactions and assisted by the strict control of molecular chaperones. However, during its native state search, the polypeptide chain may transiently populate partially folded forms with a higher propensity to aggregate.

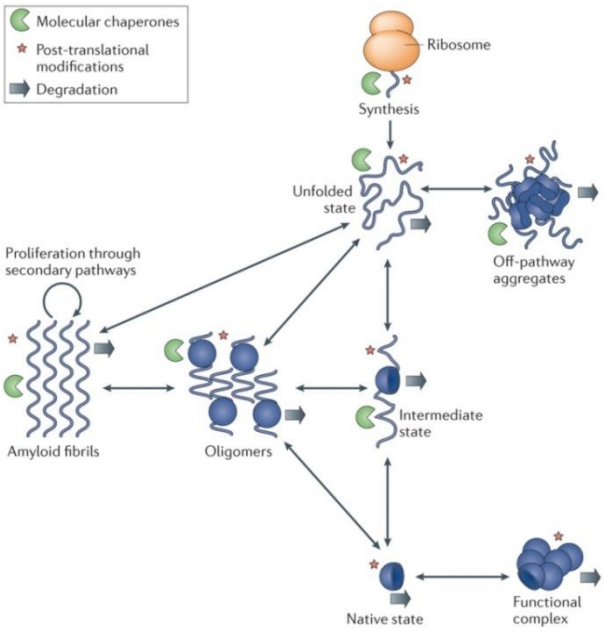

Figure I-3.3|**Different conformational states of a polypeptide chain.** Different states and interconversion rates depend by their thermodynamic stabilities, free energy barriers, the

propensity to interact with chaperones and effect of post-translational and other chemical modifications. When the polypeptide is synthetized across the ribosome is in its unfolded state, later it can associate to form off-pathway aggregates or initiate the aggregation process through partially folded intermediate that lead to oligomeric forms and amyloid fibrils (Knowles, Vendruscolo & Dobson, 2014).

The conversion rates between different soluble and aggregated states are dictated by thermodynamic and kinetic factors as well as by synthesis and degradation processes, interactions with chaperone systems and post-translational or chemical modifications (Figure I-3.3). The heterogeneous ensemble of conformational species at risk of aggregation includes totally or partially unfolded species naturally arising from the native state or promoted by stress conditions and genetic mutations, natively unstructured species, small peptides and fragments of larger proteins generated via proteolysis that cannot fold without the full protein. All these species are considered as aggregation-prone since they lead to amorphous aggregates or self-associate to form oligomers that latter evolve into amyloid fibrils or in some cases functional complexes such as actin, myosin and microtubules (Chiti & Dobson, 2006). Although for a long time it has been thought that amyloid fibrils localized in neuronal deposits might be related directly to disease, recent evidences have demonstrated that the neurotoxicity resides in small soluble oligomers and not in the fibrils themselves (Sakono & Zako, 2010; Winner et al., 2011).

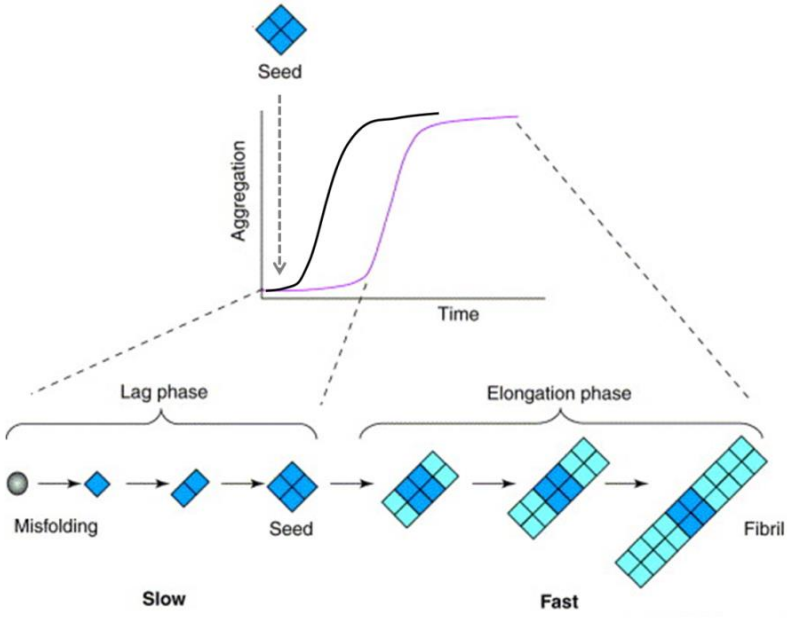

Figure I-3.4|**Nucleation growth model.** Oligomeric nuclei are formed during the lag phase in a slow process directed by unfavorable intermolecular interactions and misfolding of the protein precursor. Once nuclei are formed, a further rapid phase of elongation promotes fibril formation (pink line). When preformed seeds (aggregates or fragments of fibrils) are added to solution the lag phase is shorted resulting in acceleration of fibrils growth (black line). Adapted from (Soto et al., 2002).

This new scenario has promoted the interest in elucidating protein aggregation mechanisms since understanding oligomers assembly results pivotal to develop strategies against conformational diseases. The nucleated growth mechanism, a crystallization-similar model, is classically employed to describe amyloid formation. The conversion of precursor proteins into fibril species monitored by Th-T fluorescence, light scattering or CD signal can be plotted as a function of time as a sigmoidal curve with a first step or lag phase that comprises the formation of "nuclei" or seeds followed by a fast exponential growth or elongation phase and ended with a plateu or stationary state (Figure I-3.4). The lag phase is considered to be the rate-limiting step where partially unfolded species interconvert slowly into oligomeric nuclei; as soon as nuclei are constituted they can quickly associate with monomers or oligomers to trigger the elongation step leading to fibrils formation (Lomakin et al., 1996; Chiti & Dobson, 2006; Knowles, Vendruscolo & Dobson, 2014). The primary nucleation can be perturbed by different factors such as the increase in the initial concentration of monomeric precursor and solution stirring, shaking or ultrasonication that reduce or totally abolish the lag phase, accelerating fibril formation. Besides the addition of preformed aggregates (seeding or templating), which act as templates, promotes rapid fibril growth in a similar manner to a crystallization process. Secondary nucleation events as fibril fragmentation or fibril branching increase the number of fibril ends and soluble precursors attachment also leading to an exponential growth of fibrils (Yoshimura et al., 2012; Knowles et al., 2009; Andersen et al., 2009).

A large set of native protein precursors has been studied in order to describe the amyloid aggregation mechanism. Senile systemic amyloidosis (SSA) and Familial amyloidosis polyneuropathy (FAP) are two important human diseases associated to wild-type or variant forms of the homo-tetrameric all-β protein transthyretin (TTR), respectively. The study of FAP linked amyloid formation has described how the tetramer irreversible dissociation leads to a compact non-native monomer. This event is considered the aggregation triggering step since non-native monomer interconverts into unfolded species with hydrophobic regions exposed to solvent, which increases their tendency to aggregate and to form amyloid fibrils (Figure I-3.5A) (Quintas, Saraiva & Britot, 1999). The small monomeric human lysozyme is characterized by an α+β fold and has been largely utilized for protein stability and folding studies (Swaminathan et al., 2011). Lysozyme amyloidogenic variants are associated to familial non-neuropathic amyloidosis forming amyloid deposits in the liver, spleen, kidneys and other regions of gastrointestinal tract (Pepys et al., 1993; Yazaki & Benson, 2003). Experimental data have shown as these variants present a decrease in stability compared to the wild type resulting in an increased propensity to aggregate for the variants.

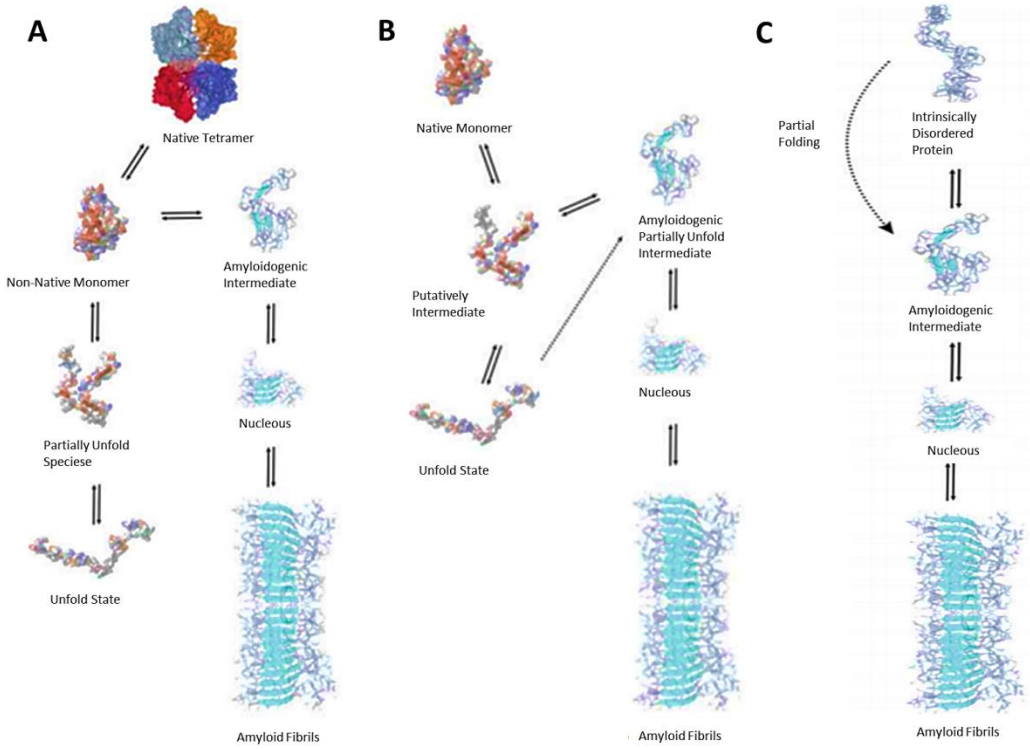

Figure I-3.5| **Schematic representation of different pathways for amyloid fibril formation.** (A) Homo-tetrameric protein precursor is dissociated in non-native monomer that convert through conformational change in unfolded species exposing hydrophobic regions that promote aggregation event. (B) Native monomer precursor is in equilibrium with unfolded states that can convert in amyloidogenic prone-aggregation intermediates. (C) IDP precursor partial folds into a β-sheet intermediate with high aggregation propensity. Adapted from (Quintas, 2013).

The aggregation mechanism proposed involves a dynamic equilibrium between native species and partially unfolded intermediates followed by the self-assembly of unstable intermediates with the subsequent formation of β-sheet ordered aggregates and, finally, amyloid fibrils (Figure I-3.5B) (Canet et al., 1999). It has been suggested that lysozyme I56T and D67H mutations may destabilize the interface between α- and β-domains with a distortion of the hydrophobic core that may encourage intermolecular interactions in the transiently exposed mutant intermediates (Dumoulin, Kumita & Dobson, 2006; Liu et al., 2007).

Proteins as α-synuclein, Aβ-peptide and amylin are intrinsically disordered proteins (IDPs) involved in Parkinson's disease (PD), Alzheimer's disease (AD) or Type II diabetes, respectively (Table I-1). To initiate the aggregation process, IDPs convert their natively unfolded structure into semi-structured species with high aggregation propensity (Figure I-3.5C) (Uversky et al.,2009; Zhang et al., 2013). Several factors such as environmental conditions, covalent modifications, mutations and elevated levels of protein expression induce IDPs partial folding and collapse into pre molten globule-like

intermediates, which can trigger ordered self-assembly, as described for example in the case of α-synuclein (Jain et al., 2011; Uversky et al., 2001; Uversky, 2008).

3.6 Determinants of Protein Aggregation

Numerous *in vitro* studies have revealed how different proteins and peptides not related to diseases can also aggregate into amyloid-like deposits under specific denaturing conditions. In addition, as discussed above, amyloid structures are largely exploited by the Nature for diverse functional roles. These evidences suggest that protein aggregation can be considered as a generic and intrinsic property of polypeptide chains since all these proteins, generally, diverge both in sequence and structure (Stefani, 2004; Chiti & Dobson, 2006). The different parameters that affect the aggregation process can be distinguished in *extrinsic* or *environmental* factors such as protein concentration, temperature, ionic strength and pH, and *intrinsic* factors regarding specific properties of the polypeptide chain related to the amino acidic composition and sequence. The protein concentration is an extremely important factor since it can favor intermolecular interactions that reduce significantly the lag phase of the aggregation reaction. Temperature, ionic strength and pH influence unfolding and aggregation rates as well as the morphology of the aggregates (Li et al., 2010). For instance, it has been described that in the case of the essential luminal polypeptide region of Pmel17 (RPT) the aggregation is regulated by the pH, leading to different fibrils morphology depending of the stage of maturation of the melanosomes (Pfefferkorn et al., 2010). Another interesting example is the polymorphism of α-synuclein aggregates, which is controlled by the salt concentration of the solution, leading to the generation of two strains: α-syn fibrils (at physiological salt concentration) and α-syn ribbons (at low salt concentration) showing different structures, toxicity and propagation properties (Bousset et al., 2013).

As mentioned before, intrinsic factors also influence the propensity of a given protein to aggregate and, accordingly, physicochemical properties of polypeptide chains such as hydrophobicity, charge and the propensity to form secondary structure have been analysed. Point mutations in aggregation-prone regions have been used to characterize the contribution of the amino acidic composition upon global aggregation (Chiti et al., 2003). An increase in the hydrophobicity of the polypeptide usually correlates with an increase in the aggregation propensity whereas increasing the net charge decreases it due to electrostatic repulsions that preclude the formation of stable intermolecular contacts (Otzen et al., 2000; Kim & Hecht, 2008; Chiti et al., 2002a). Because the aggregates are enriched in β-sheet structure, mutations promoting β-sheet are pro-aggregational whereas and increase in α-helical structure protects from aggregation. (Villegas et al., 2000; Chiti et al., 2002b). Remarkably, the large number of works that have used mutational analysis in different protein models and sequential regions have shown that not all the amino acidic residues in a protein

play the same role in the promotion of aggregation, rather short specific amino acid stretches can act as facilitators favoring the intermolecular interactions and nucleating fibril formation, according to the so-called amyloid stretch hypothesis (Ventura et al., 2004; Ivanova et al.,2004; Pastor, Esteras-Chopo & Serrano, 2007; Teng & Eisenberg, 2009). These Aggregation Prone Regions (APRs) also known as "hot spots" are generally composed of 5-15 amino acids, enriched in hydrophobic residues and with a low net charge, that under native conditions are usually buried in the hydrophobic core or implicated in intermolecular contacts sustaining the secondary and tertiary structure of the protein or involved in protein-protein interactions. Nevertheless when mutations, oxidative stress or aging cause the APRs solvent exposure, they rapidly assemble forming β-sheet enriched aggregates (Sánchez de Groot, et al 2005; Rousseau et al., 2006).

A strong selective pressure is necessary to the avoid aggregation of essential globular and especially intrinsically disordered proteins since, in this latter case the APRs are in contact with the solvent (Sanchez de Groot et al., 2012). Accordingly, Nature has evolved a protective sequence code constituted by "gatekeepers" or "amyloid breakers", which are short regions that flank APRs and disrupt their aggregative capability. On one side, gatekeepers are constituted of charged amino acids (Arg, Lys, Asp, Glu) that inhibit interactions via the repulsion effect of the charge, on the other side the presence of large and flexible side-chains such as those of Arg or Lys penalize aggregation acting upon the entropic factor (Otzen et al., 2000; Beerten et al., 2012). In addition, the presence of amino acids as proline or glycine lowers the β-structure forming capability. Finally, the presence of disulfide bonds and salt bridges in aggregation-prone interfaces protect the native protein ensemble from conformational rearrangements that lead to misfolding and aggregation (Monsellier & Chiti, 2007; Pechmann et al., 2009).

3.7 Inclusion Bodies (IBs)

The use of bacterial cells as machineries for the production of recombinant proteins has been extensively exploited in different fields such as industry and in research due principally to their low-cost and easy handling, as in the paradigmatic case of *E. Coli.* However, protein over-expression in bacteria leads, in many cases, to the formation of large insoluble aggregates called inclusion bodies (IBs) being a challenging problem for the biotechnological production of soluble active proteins (García-Fruitós, 2010). Many strategies have been developed in the last years to optimize heterologous protein expression and to avoid the loss of solubility by the control of culture media composition, decreasing the growth temperature, modulating the transcription rate or enhancing the accessibility to chaperone systems (Sørensen & Mortensen, 2005). In addition, numerous refolding procedures have been also designed with the aim to recover insoluble proteins from IBs (Middelberg, 2002; Fahnert et al., 2004).

Using transmission electron microscopy, IBs appear as amorphous and refractile aggregates displaying sphere-like or rod-like shapes with diameters ranging from 0.2 to 1.2 µm, located at both poles (Figure I-3.6A) or in the periplasmic space of the cell (Figure I-3.6B).

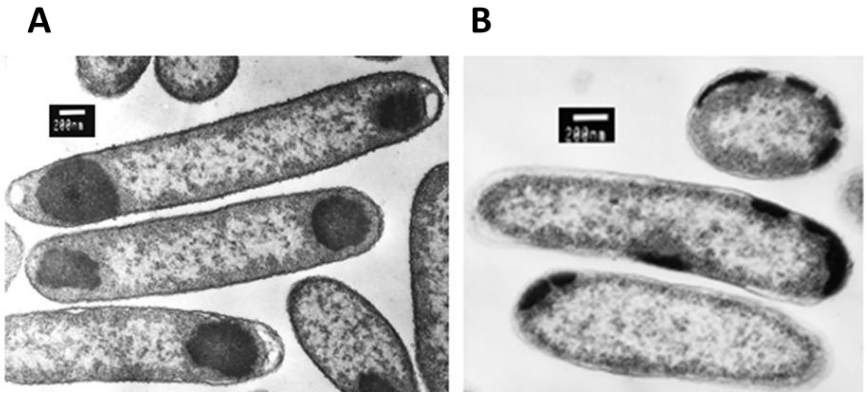

Figure I-3.6| **Microscopic view of *E.Coli* Inclusion bodies.** (A) Electron micrograph of aggregated polypeptides of tailspike protein accumulated as electron-dense inclusion bodies localized at one or both poles of the cells. (B) TEM image of pectate lyase C aggregates IBs at the cell periphery suggests that they are formed in the periplasmic space. Adapted from (Betts, 1998).

IBs result to be composed usually by 90% of the recombinantly over-expressed proteins and by low quantities of molecular chaperones (Carrió et al., 1998; Rattenholl et al., 2001). During many years, IBs have been considered as "molecular dust balls", ordinary misfolding and inactive products which formation was driven by non-specific interactions between exposed hydrophobic surfaces of folded or partially denatured polypeptides (Fink, 1998; Ventura & Villaverde, 2006).

The IBs definition has undoubtedly changed with recent and numerous evidences confirming that aggregation reactions are usually specific events mainly controlled by the sequence of the polypeptide chain, both in eukaryotic and prokaryotic organisms (Rajan et al., 2001; Ventura, 2005). Accordingly, it has been demonstrated that IBs formation corresponds in many cases to a specific self-associative process, as illustrated by the fact that when two different aggregation-prone proteins are co-expressed in the same cell they segregate into different aggregates and do not co-aggregate (Ventura & Villaverde, 2006; Morell et al., 2008). The use of techniques with high resolution, such as X-ray and NMR spectroscopy, has allowed a better molecular characterization of these apparently amorphous deposits, revealing the presence of amyloid-like structures., IBs have been analyzed with the battery of methods typically used to detect amyloid fibrils related to human diseases, from binding to specific dyes such as Th-T and Congo Red to determination of secondary structure by CD and FT-IR, confirming the presence of classical amyloid cross-β features. Besides, typical mature fibrils have been observed by TEM after IBs partial digestion with proteinase K (Wang et al., 2008; Morell et al., 2008). These properties have made of IBs an interesting and

useful tool that can be exploited to model protein folding, misfolding and amyloid-like aggregation *in vivo*. (De Groot et al., 2006). Furthermore, IBs not only contain β-sheet structures but also intermediate and native conformations that can maintain their biological activity, offering new possibilities for their potential use as biocatalysts or as nanoparticles (García-Fruitós et al., 2005; Villaverde et al., 2012).

PART II - Objectives

The general aim of this thesis is the further understanding of the molecular determinants of functional and non-functional protein aggregation. On one side, it was analyzed the regulative mechanism used by Nature to direct the assembly of functional amyloids and, on the other side, the role of the different intrinsic and extrinsic factors that can lead to misfolding and harmful amyloid aggregation.

Hence, this thesis was subdivided in two main sections and the following objectives were proposed for each presented work:

1. Functional aggregation

- To dissect the role of phenol soluble modulins (PSMs) in *S. aureus* biofilm structuration and virulence.

2. Non-functional aggregation

- To characterize the amyloidogenic properties of the intracellular aggregates formed by the human prion-like RNA-processing protein HNRPDL.
- To investigate the impact of cosolvent-induced structural rearrangements on the process of amyloid formation by the all α-helical FF domain.
- To study the contribution of configurational entropy to the folding, stability and amyloid fibrillation of the FF domain.
- To analyze the effect of a specific oxidative modification upon the structure, conformation and aggregation propensity of the FF domain.

PART III - Material & Methods

1 Protein Models

1.1 α-PSMs peptides of *S. aureus*

An increasing interest is being shown on the amphipathic peptides family produced by *S. aureus* and other staphylococci generally named as "phenol soluble modulins" (PSMs), since they have been recognized as crucial determinants of staphylococcal pathology (Wang et al., 2007; Peschel & Otto, 2013). The term of phenol soluble modulins was introduced for the first time in 1999 by Mehlin and colleagues when, during hot phenol extraction, they identified a virulent polypeptide complex of *S. epidermidis* into the phenol phase (Mehlin, Headley & Klebanoff, 1999). Thereafter, a group of peptides related in part to PSMs of *S. epidermidis* were described in *S. aureus.* The use of different techniques such as analytical and preparative reversed-phase HPLC/electrospray mass spectrometry (RP-HPLC/ESI-MS) and N-terminal peptide sequencing allowed a better characterization of PSMs, which can be classified in α- and β-type depending on their length. In this way, it is possible to distinguish between shorter α-PSMs constituted by 20-25 amino acids including the previously described δ-toxin, and two β-PSMs peptides with 43-45 residues (Figure III-1.1A).

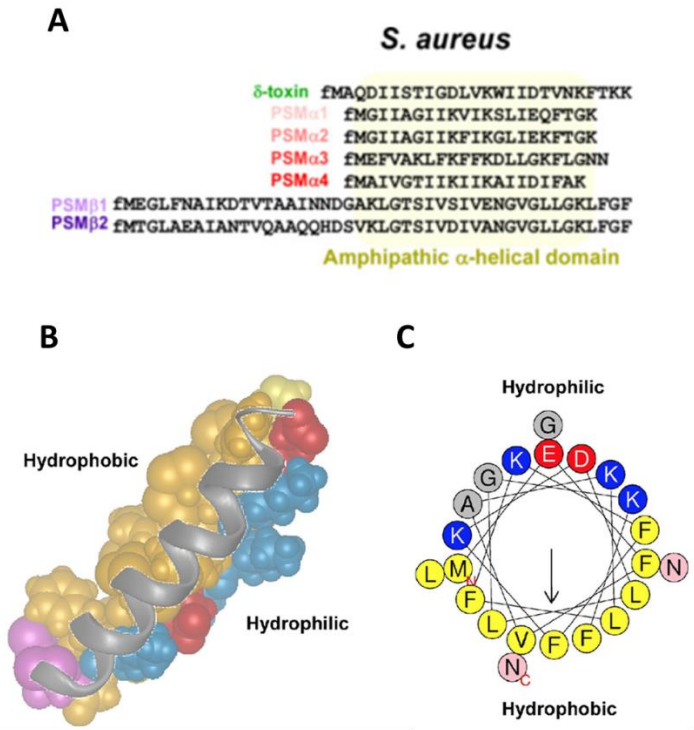

Figure III-1.1| **Sequence and structure of PSMs.** (A) Amino acidic sequence of PSMs. (B) Structural representation of αPSM3 modeled from NMR data available for δ-toxin and (C) an α-helical wheel projection of the same peptide. Hydrophobic amino acids are in yellow,

hydrophilic amino acids in blue (positive charge) or red (negative charge). The two C-terminal asparagine residues are in pink. Adapted from (Cheung et al., 2014a).

All PSMs are characterized by an amphipathic α-helix as shown for the case of α-PSM3 in Figure III-1.1B. The α-helical wheel projection of this peptide presents the disposition of hydrophilic and hydrophobic residues in opposite sides showing the amphipathic nature of the α-helix (Figure III-1.1C) (Wang et al., 2007; Laabei et al., 2014; Cheung et al., 2014a). The genes encoding for α-PSMs and β-PSMs are arranged in two clusters, in the alpha and beta operon respectively, while δ-toxin is encoded within the regulatory RNA, RNAIII, which also controls the positive regulation of the other staphylococcal exotoxins through the global regulatory quorum-sensing system. The psm-mec gene located in SCCmec, a mobile genetic element that confers methicillin resistance, regulates *S.aureus* virulence controlling indirectly PSM production (Kaito et al., 2011; Kaito et al., 2013). PSMs are involved in a series of biological functions that are critical for staphylococcus pathogenesis. PSMs trigger the inflammatory response while specifically α-PSMs of *S .aureus* and *S. epidermidis* can efficiently lyse white and red blood cells. Furthermore, it was also demonstrated that PSM peptides are responsible for the development of biofilm structures contributing to chronic *S. aureus* infections due to their resistance to host immune response and chemotherapies (Periasamy et al., 2012a; Periasamy et al., 2012b).

1.2 Heterogeneous Nuclear Ribonucleoprotein D-Like (HNRPDL)

The human Heterogeneous Nuclear Ribonucleoprotein D-Like (HNRPDL) belongs to the subfamily of ubiquitously expressed heterogeneous nuclear ribonucleoproteins (hnRNPs), associated with pre-mRNAs in the nucleus and involved in pre-mRNA processing and other aspects of mRNA metabolism and transport (Dreyfuss et al., 1993 ; Krecic & Swanson, 1999). The hnRNPs are characterized by the presence of RNA Recognition Motifs (RRMs), exceptionally versatile domains that result indispensable in the mRNA binding to perform their numerous biological functions in the cell (Maris et al., 2005). Interestingly in the last years, the development of algorithms designed to detect prion domains has allowed the identification of a selected group of RNA-binding proteins containing Prion-Like Domains (PrLDs) that have been recently linked to different neurodegenerative disorders in humans, suggesting that they are potentially pathogenic (King et al., 2012). The algorithms have been build up taking as models yeast prion domains, which are characterized by sequences of low complexity, highly enriched in asparagine (N) and/or glutamine (Q) residues and predicted to be intrinsically unstructured (Alberti et al.,2009). Yeast Prion Domains can switch between an unfolded conformation and a transmissible cross-β conformation, being both necessary and sufficient for amyloid formation and propagation (Li & Lindquist, 2000).

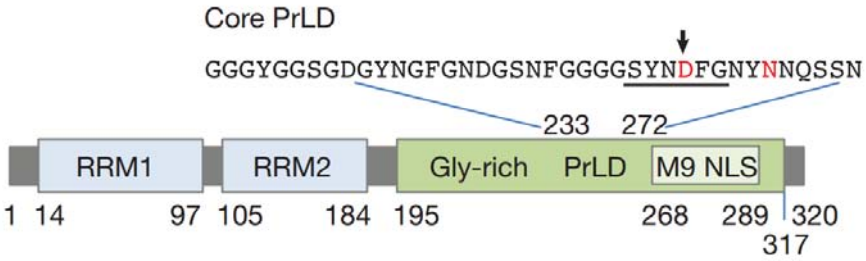

Figure III-1.2| **Domain architecture of hnRNPA1.** In the figure are shown the RNA recognition motifs 1 and 2 (RRM1 and RRM2), the C-terminal glycine-rich domain and an M9 nuclear-localization signal. The PrLDs are centered in the C-terminal glycine-rich domain. Adapted from (Kim et al., 2013).

HNRPDL, such as hnRNPA1 shown in Figure III-1.2, presents two contiguous RRMs and is predicted to contain a PrLD at its C-terminus. Recently it has been shown to be linked to limb-girdle muscular dystrophy 1G, a genetically determined muscle disorder with a primary or predominant involvement of the pelvic or shoulder girdle musculature (Vieira et al., 2014).

1.3 Yeast URN1-FF Domain

FF domains are small protein-protein interaction modules consisting of 50–70 residues often organized in tandem arrays, which display two conserved Phe residues at the N- and C-termini (Bedford & Leder, 1999). These domains consist of three α-helices arranged as an orthogonal bundle with a 3_{10} helix in the loop connecting the second and the third helix (Figure III-1.3). FF domains are involved in RNA splicing, signal transduction and transcription pathways and they are present in different nuclear transcription factors such as the human CA150 and splicing factors like as human HYPA/FBP11, yeast Prp40 and yeast URN1 (Bonet et al., 2008; Allen et al., 2002; Gasch et al., 2006).

The conformational properties of this domain are receiving increasing attention, since its folding reaction involves the transient population of an on-pathway intermediate whose structure has been solved at atomic resolution by NMR, providing thus high-quality data on the structural constraints driving the early stages of folding (Korzhnev et al.,2010; Korzhnev et al., 2012). The URN1 protein from *Saccharomyces cerevisiae* is a pre-mRNA splicing factor containing only a FF domain. For this simplicity it is used as the most representative model of FF domain-containing splicing factors.

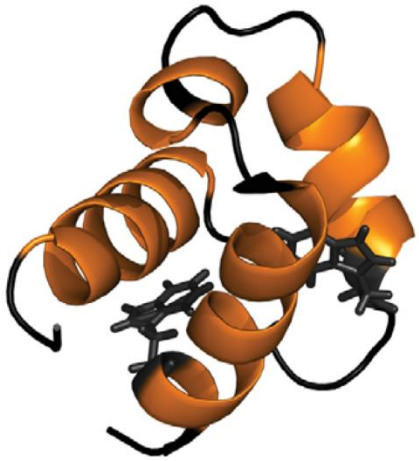

Figure III-1.3| **Ribbon representation of the URN1-FF domain structure.** Tryptophan residues are shown in black. The Protein Data Bank accession code for the structure is 2JUC.This figure was prepared with PyMOL.

From the study of its structure in solution, it has been reported that URN1-FF domain presents a classical FF fold but a distinctive charge distribution on the surface compared to other FF domains (Bonet et al., 2008). Interestingly, we have shown in a recent work that the URN1-FF domain populates a molten globule (MG) state under mild denaturing conditions at low pH, which, despite being devoid of any detectable β-sheet element, self-assembles into highly ordered amyloid fibrils (Castillo et al., 2013).

2 Analysis of Protein Aggregation

2.1 Aggregation Prediction Methods

In the last years, the increasing knowledge about amyloid fibril formation and the study of the determinants involved in this process has pushed up the development of several mathematical tools for aggregation prediction and the identification of aggregation prone regions. The rationalization of sequential changes effects on protein aggregation resulted fundamental to develop efficient predictors. From the first mathematical tool published in 2003 by Chiti and co-workers, more than 20 computational methods have been developed to predict the aggregation behavior of polypeptide chains (Belli et al., 2011; Ahmed & Kajava, 2013). These predictive methods can be classified in empirical or phenomenological models and in structured-based approaches. Empirical methods use the physicochemical properties of the amino acid sequence such as hydrophobicity, β-propensity and solubility to explain

experimental data and, on the other side, the structured-based models analyze the few existing three-dimensional (3D) structures of peptides that adopt a known fibrillar conformation to perform atomistic simulations of potential amyloidogenic fragments (Caflisch, 2006).

As mentioned above, Chiti and colleagues were the first to propose a mathematical tool to obtain predictions consistent with *in vitro* experimental data. The study of single point mutants of the human muscle acylphosphatase (AcP) showed a correlation between the change in aggregation rates of mutants and changes in their hydrophobicity, propensity to form α-helical and β-sheet structures and overall charge (Chiti et al., 2003). Therefore, these intrinsic properties were used to develop an empirical and simple formula based on three parameters that allows an accurate prediction only for all unstructured or intrinsically disordered protein states. In order to obtain a prediction of absolute aggregation rates, Dubay and co-workers proposed a seven-parameter formula where intrinsic properties of the polypeptide chain were combined with extrinsic environmental factors such as peptide concentration, pH value and ionic strength of the solution (DuBay et al., 2004). This parameters were further integrated to provide a scale of the individual intrinsic aggregation propensities of the 20 natural amino acids and to identify specific "sensitive regions" inside the sequence of unfolded peptide or protein (Pawar et al.,2005). A further refinement of Pawar´s function was presented with the Zyggregator method which not only takes into account the contribution of intrinsic properties of amino acidic sequence in fully or partial unstructured proteins but also the tendency of a protein region to adopt a folded conformation, considering its local stability as derived from the protein flexibility and solvent accessibility provided from the CamP analysis which uses hydrogen exchange measurements to predict them directly by the amino acid sequence without any knowledge of the native structure of the protein.(Tartaglia et al., 2007; Tartaglia, & Vendruscolo, 2008; Tartaglia, & Vendruscolo, 2007).

TANGO was the first empirical method that analysed the sequences propensity to form β-aggregating regions calculating the different capability of each residue to attain several competing conformations as β-turn, α-helix, β-sheet and β-sheet aggregating. According to TANGO, a polypeptide chain presents propensity to aggregate when it contains a fragment of at least five consecutive residues with β-aggregation tendency higher than 5%. This statistical mechanics algorithm explores both protein stability and intrinsic protein properties but also takes into account extrinsic factors such as pH, protein concentration and ionic strength (Fernandez-Escamilla et al., 2004).

The SALSA algorithm was designed to calculate the location of regions with high propensity to form β-strands, since these fragments are considered to be "fibrillogenic hot spots" embedded in the protein sequence (Zibaee et al, 2007).

Remarkably, AGGRESCAN was the first empirical predictor server developed with experimental data obtained from *in vivo* systems. The intrinsic aggregation–propensity scale for the 20 natural amino acids was calculated monitoring the GFP solubility of

GFP-Aβ42 variants expressed in *E.coli* (Sánchez de Groot et al., 2005). These data were used to evaluate the aggregation propensity of each amino acidic residue according to its relative position in the sequence generating an aggregation propensity profile from the primary structure of the analysed protein (Conchillo-Solé et al., 2007). Recently, an evolution of AGGRESCAN method was presented where a structure-based approach is used to resolve the limitations of sequence-based algorithms. Hence, AGGRESCAN3D (A3D) detects specifically hot spots in globular proteins taking into account protein dynamic fluctuations and spatial clustering of residues distant in the protein sequence, providing the possibility to virtually mutate residues to increase the solubility or test pathogenic mutations (Zambrano et al., 2015a).

The FoldAmyloid algorithm, predicts short aggregation prone regions in a similar way to AGGRESCAN, but the calculation of individual amino acid aggregation propensities proceeds from the statistical analysis of known tridimensional structures of globular proteins using two characteristics, the expected packing density and the hydrogen bonds formation probability, which correlate with amyloidogenicity (Garbuzynskiy et al., 2009).

From the statistical analysis of a non-redundant dataset of native globular proteins, it was developed the PASTA algorithm that analyzes the role of specific sequence to form self-propagating cross-β structures. For this purpose, it was calculated the propensity of two residues to be located into a β-sheet facing one another in parallel or antiparallel neighboring strands (Trovato et al., 2006; Trovato et al., 2007). In a similar manner, BETASCAN algorithm calculates the most preferred β-pairing considering the likelihood scores for potential β-strands based on correlations derived from known amphipathic parallel β-sheets described in the PDB database (Bryan et al., 2009).

The use of the 3D profile method, a structure-based algorithm, allowed the analysis of more than 20000 protein sequences and the collection, in the ZIPPER database, of all those segments with strong fibrillation propensity able to form a steric zipper. The 3D profile algorithm was developed on the basis of the crystal structure of the fibril-forming peptide NNQQNY that was used as structural backbone to analyse a large set of six-residue peptides not containing proline residues, which energetically fitted into the crystal structure through the ROSETTADESIGN software (Thompson et al., 2006; Kuhlman & Baker, 2000).

The analysis of amyloid-forming hexapeptides (AmylHex dataset) and the consideration of the physicochemical properties of their amino acidic residues were used in the WALTZ algorithm to generate a position specific scoring matrix (PSSM) of hexapeptides useful to discriminate between amyloid-forming sequences and amorphous β-sheet aggregates (Maurer-Stroh et al., 2010). Interestingly, a WALTZ scoring function update (pWALTZ) was recently used for the identification and evaluation of amyloid cores in putative prion domains through the PrionW server that allows the recognition of prion-like domains (PrLDs) scanning for the presence of

amyloid cores in glutamine/asparagine (Q/N) rich sequences (Sabate et al., 2015 ; Zambrano et al., 2015b).

Finally, in order to enhance aggregation prediction accuracy it was developed the consensus algorithm AMYLPRED which includes the employment of 5 of the described methods. This consensus method generates an integrated prediction result at the residue level that allows a simultaneous comparison between the various outputs, which in principle might minimize overestimation by a particular individual predictor. The early version of AMYLPRED has been recently updated to AMYLPRED2 that now combines 11 different methods (Frousios et al., 2009; Tsolis et al., 2013).

PART IV - Results & Discussion

1 Functional aggregation

Functional amyloid-like aggregates are exploited by Nature for different functions. Most of them are used by several species of bacteria to build up strong and resistant fibrillar matrices, essential elements of biofilms that result fundamental for the attachment and colonization of bacterial colonies (Barnhart & Chapman, 2006; Dueholm et al., 2012; Oli et al., 2012). In the first part of this thesis the molecular determinants of the functional amyloid matrix of *S. aureus* biofilm were analyzed in order to understand their involvement in staphylococcal pathogenesis.

S. aureus biofilm

The opportunistic pathogen *Staphylococcus aureus* is recognized as one of the most frequent causes of biofilm-associated infections. Methicillin-resistant *S. aureus* (MRSA) strains are involved in several hospital acquired "nosocomial" infections such as endocarditis, necrotizing pneumonia and septic shock from clinical site contamination (Rubin et al., 1999). Furthermore, in the last few decades, new community-associated (CA-) MRSA strains have emerged that are able to infect healthy people showing both antibiotic resistance and enhanced virulence and fitness (Chambers, 2001; M. Otto, 2010). Thus, *S. aureus* infection ability results to be tightly associated to the extraordinary bacterial biofilms resistance to host immune response and chemotherapies (Otto, 2008). Staphylococcal biofilm development can be generally divided in four steps: (a) cell attachment to a surface, (b) assembly of attached initial cells to constitute a micro-colony, (c) growth to form a mature structure and finally (d) biofilm disassembly, characterized by the detachment and dispersion of bacterial clusters. The biofilm extracellular matrix comprises chemically different macromolecules as polymeric proteins, teichoic acids, extracellular DNA (eDNA) and polysaccharides (Otto, 2008; Boles & Horswill, 2011). Recently, small amphipathic peptides defined as phenol soluble modulins (PSMs) have been recognized as the main molecular effector of staphylococcal biofilm maturation and dispersal disrupting non-covalent forces in the biofilm matrix to facilitate the detachment of agglomerate cells. (Otto, 2013; Le et al., 2014). Interestingly, it has also been reported that PSMs contribute to the biofilm integrity promoting the formation of an extracellular fibril structure displaying amyloid-like properties suggesting a potential regulative role of aggregation on PSM biological activity (Schwartz et al., 2012).

1.1 Dissecting the contribution of *S. aureus* αPSMs to biofilm amyloid structure and toxicity

αPSMs sequence analysis and aggregation propensity

S. aureus phenol-soluble modulins (PSMs) are small peptides, which show amphipathic, α-helical secondary structures and surfactant-like properties (Mehlin et al., 1999; Peschel & Otto, 2013; Cheung et al., 2014a). PSMs play multiple functions in *S. aureus* pathogenesis, from cytolytic activity toward leukocytes to biofilm assembly (Otto, 2014). In order to understand the PSMs role in staphylococcal virulence, we analyzed the shorter αPSMs contribution. We also include in the analysis δ toxin, a short PSM intimately linked to the αPSMs in the biofilm matrix. PSMs share barely conserved sequence with similar physicochemical properties, as revealed from the αPSMs sequence alignment (Figure IV-1.1) by T-Coffee method (Notredame et al., 2000). The conserved residues are shown in grey and marked by asterisks, while amino acidic clusters are distinguished by different colors according to Taylor's classification (Hossain et al., 2011). In all PSMs sequences**,** the predominant hydrophobic character is coupled to the presence of tiny and small residues. Positively charged amino acids are in general well conserved, whereas negative residues could only be aligned for αPSM1, αPSM2 and αPSM4 sequences.

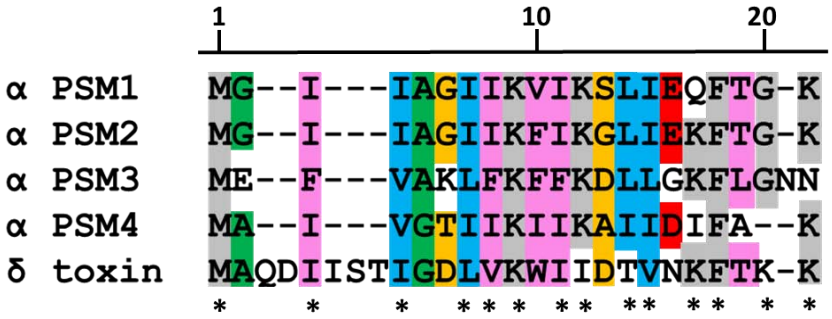

Figure IV-1.1| **α-Phenol Soluble Modulins Alignment.** Sequences of α-PSMs are aligned with T-Coffee server (Notredame et al., 2000). Conserved residues (at least 3) are shown in grey with an asterisk. Tiny (green), small (orange), hydrophobic (pink), hydrophobic and aliphatic (blue) and negatively charged (red) amino acids are indicated.

In an attempt to find PSMs sequence/aggregation correlation, Amylpred2 consensus method was used to predict the aggregation propensity of each individual peptide. This software allows sequence analysis using simultaneously up to 11 validated prediction methods. The consensus prediction output results by the hit overlap of at least n/2 (rounded down) out of n selected methods (Tsolis et al., 2013). The PSM sequences were processed with Amylpred2 program and both the individual and consensus outputs are shown in Table IV-1.1. The PSMs generally show a high

predicted aggregation tendency when individual algorithms are used such in the cases of WALTZ or AGGRESCAN methods. However, the Amylpred2 consensus analysis indicates that there exist certain differences in the aggregation propensity of PSMs showing longer and shorter aggregation-prone regions for αPSM4 (residues 1-19) and αPSM3 peptides (residues 5-8), respectively; suggesting the possibility that these PSMs peptides might contribute in a distinct manner to biofilm formation and stability.

Table IV-1.1| **Predicted aggregation propensities of PSM peptides using the consensus method Amylpred2.**

	αPSM1	**αPSM2**	**αPSM3**	**αPSM4**	**δ toxin**
AGGRESCAN	1-17	1-17	1-10, 16-20	1-20	4-8, 13-19
AmyloidMutants	-	-	1-8, 14-22	1-8, 11-19	1-10, 15-23
Amyloidogenic Pattern	-	-	-	-	5-10
Average Packing Density	-	-	5-13	8-13	13-18
Beta-strand contiguity	1-19	2-16	-	1-19	4-24
Hexapeptide Conf. Energy	1-12	1-12	3-8	1-20	5-10, 13-20
NetCSSP	4-14	5-19	6-12	1-19	3-21
Pafig	1-6, 13-19	1-6, 8-19	2-8	1-7, 15-20	5-10, 15-24
SecStr	4-14	4-12	-	3-4, 6-13	12-15
TANGO	3-8	3-8	-	2-8, 10-19	12-20
WALTZ	4-21	4-21	6-11, 14-22	3-20	1-26
Amylpred2	**3-14**	**3-16**	**5-8**	**1-19**	**4-10, 12-21**

Structural and morphological characterization of PSM aggregation *in vitro*

As, discussed above, it has been recently reported that the conjunct of *S. aureus* PSMs participate in biofilm assembly forming functional amyloid structures (Schwartz et al., 2012). However, the individual contribution of each particular PSM to amyloid formation was unknown. Since computational predictions suggested that they might differ in their amyloidogenic properties, we decided to characterize PSMs aggregation *in vitro*. Each PSM peptide was incubated at 200 μM final concentration at room temperature and the aggregation process was monitored by the binding to the amyloid-specific dye Thioflavin-T (Th-T) (Figure IV-1.2A). PSMs aggregation was monitored during twenty-five days. Only αPSM4 and αPSM1 promote a significant increase in fluorescence emission of the Th-T dye. αPSM4 binds Th-T dye at six days while αPSM1 shows Th-T binding only at twenty-five days indicating a faster aggregation rate for αPSM4. This trend was observed also by far-UV CD (Figure IV-1.2B) where the typical amyloid β-sheet signal at ≈218 nm is detected only for αPSM1

and αPSM4. The amyloid properties of PSM aggregates were corroborated by Congo red (CR) binding (Figure IV-1.2C) displaying for αPSM1 (25 days) and αPSM4 (6 and 25 days) the characteristic band at ≈540 nm, while no significant red shift change is observed for all other peptides.

Accordingly, the morphological analysis of αPSM aggregates by Transmission Electronic Microscopy (TEM) reveals the presence of well ordered amyloid-like fibrils for αPSM1 after twenty-five days of incubation and αPSM4 mature fibrils are already observable at six days (Figure IV-1.3).

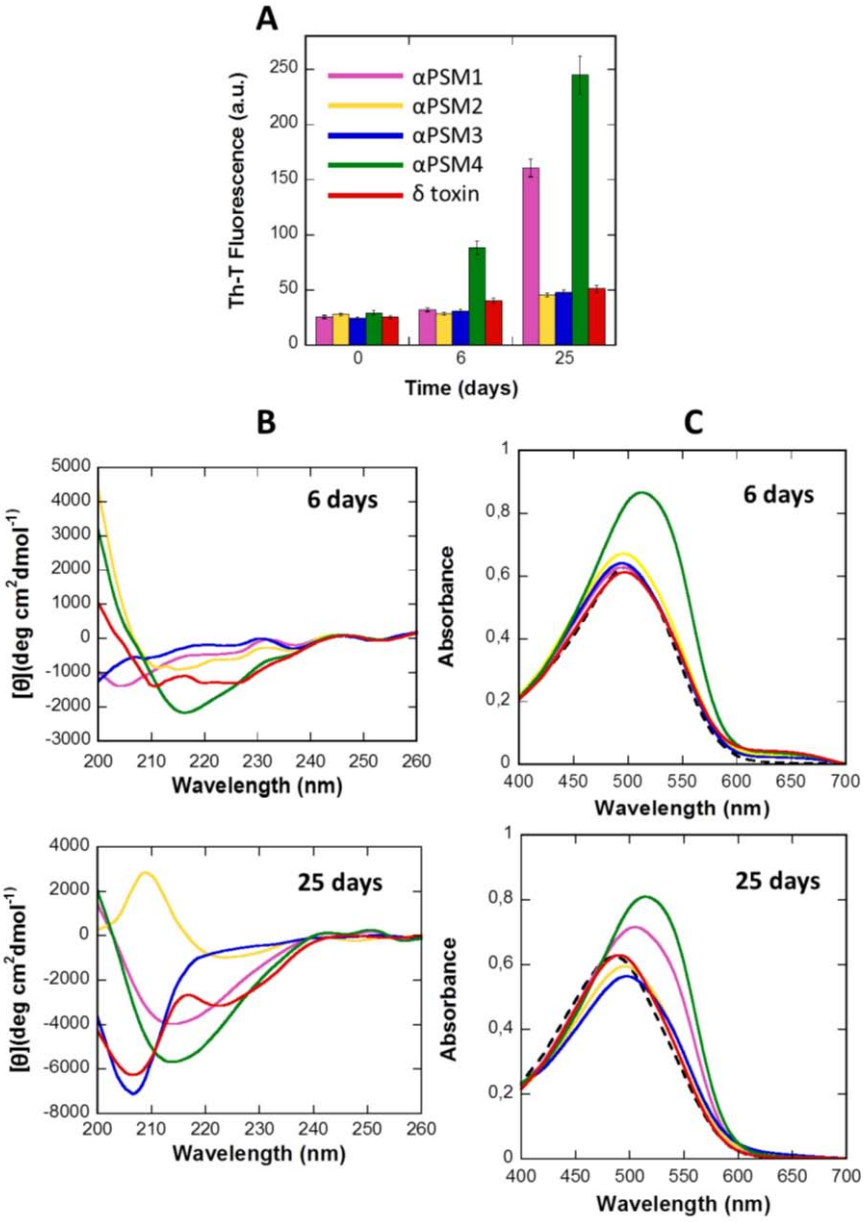

Figure IV-1.2| **Aggregation kinetics and characterization of PSM aggregates.** (A) Kinetics were monitored by following the change in relative Th-T fluorescence during 25 days. Error bars indicate ± SE (n= 3). (B) Far-UV CD spectra of α-PSM aggregates and (C) Congo Red (CR) binding was registered at 6 days and 25 days. Free Congo Red is represented by a discontinuous black line.

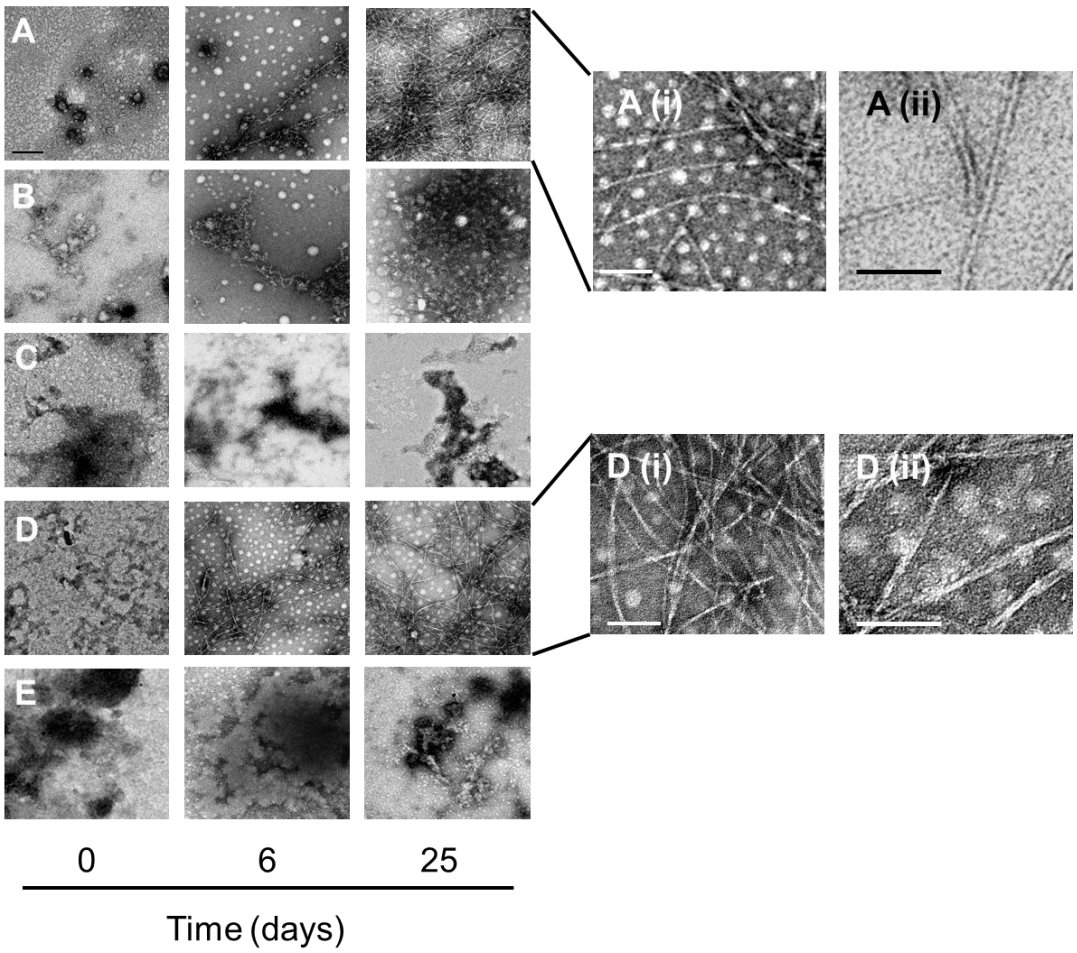

Figure IV-1.3| **Morphological properties of PSM aggregates.** TEM micrographs of the aggregates formed by α-PSM1 (A), α-PSM2 (B), α-PSM3 (C), α-PSM4 (D) and δ-toxin (E) at different time points. The scale bar represents 0.5 µm. α-PSM1 (Ai, Aii) and α-PSM4 (Di, Dii) fibrils (25 days) are shown at different magnification, the scale bar represents 100 nm.

Unbranched, long and untwisted fibrils are observed for αPSM1 and αPSM4. Conversely, αPSM2, αPSM3 and δ-toxin show amorphous aggregates during all incubation time. The similar morphology shared for αPSM1 and αPSM4 fibrils (incubated 25 days) was confirmed by Attenuated Total Reflectance-Fourier Transform Infrared spectroscopy (ATR-FTIR) evaluating their secondary structure content (Figure IV-1.4). Deconvolution of the absorbance spectra in the amide I region allows to analyze the individual secondary structure components and their relative contribution to the main signal. As shown in Table IV-1.2, an intense signal at ≈1624 cm^{-1} and ≈1626 cm^{-1} for αPSM1 and αPSM4, respectively, indicates the presence of intermolecular β-sheets packing into the characteristic amyloid structure (Natalello & Doglia , 2015). αPSM1 and αPSM4 also display common bands at ≈1636 cm^{-1} associated to intramolecular β-sheet structure and a band at ≈1667 cm^{-1}, assigned to β-turns. A band at ≈1652 cm^{-1} typically assigned to helical/random conformations is also detected. Overall, these data confirm the PSMs ability to form amyloid structure suggesting that this property may be specifically attributed to αPSM1 and αPSM4

peptides, which contribute to amyloid formation with distinct aggregation propensities.

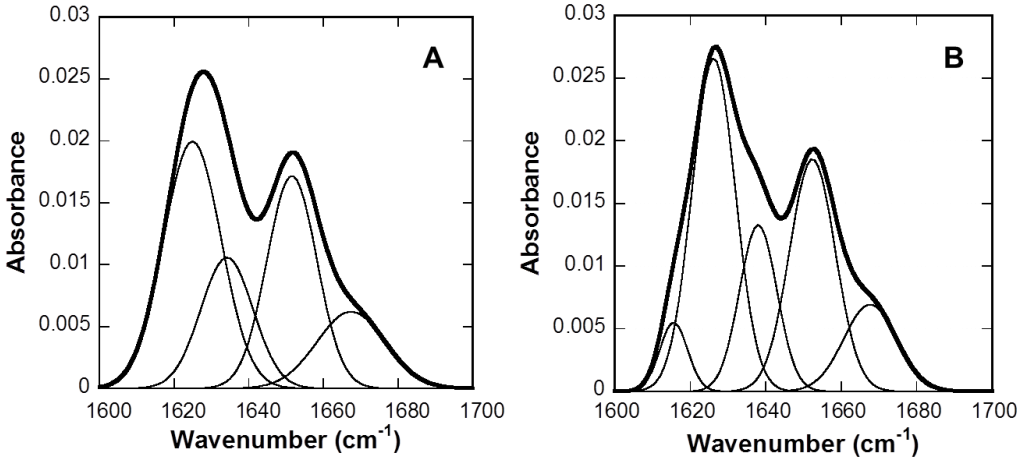

Figure IV-1.4| **α-PSM1 (A) and α-PSM4 (B) aggregates characterization by ATR-FTIR.** Absorbance spectra of the amide I region (continuos thick line) showing the component bands (continuos thin lines). The sum of individual spectral components after Fourier self-deconvolution closely matches the experimental data.

Table IV-1.2| **Secondary structure bands in the deconvoluted absorbance FTIR spectra of α-PSM1 and α-PSM4 aggregated peptide solutions.**

	Bands (cm^{-1})	**% Area**	**Structure**
αPSM1	1624	38.5	β - sheet (inter)
	1634	18.5	β - sheet (intra)
	1651	28.8	α-helix/random
	1667	14.2	turn
αPSM4	1615	5	β – sheet (inter)
	1626	38.2	β - sheet (inter)
	1638	16.5	β – sheet (intra)
	1652	28.2	α-helix/random
	1667	12.1	turn

Seeding of amyloid formation in PSMs peptides

The addition of preformed fibrils enhances the rate of fibril formation in a phenomenon defined as seeding. The sequence specificity in the seeding of protein amyloid fibrils was demonstrated (Krebs et al., 2004) indicating that protein aggregation can be nucleated by fibrils that share a similar sequence.

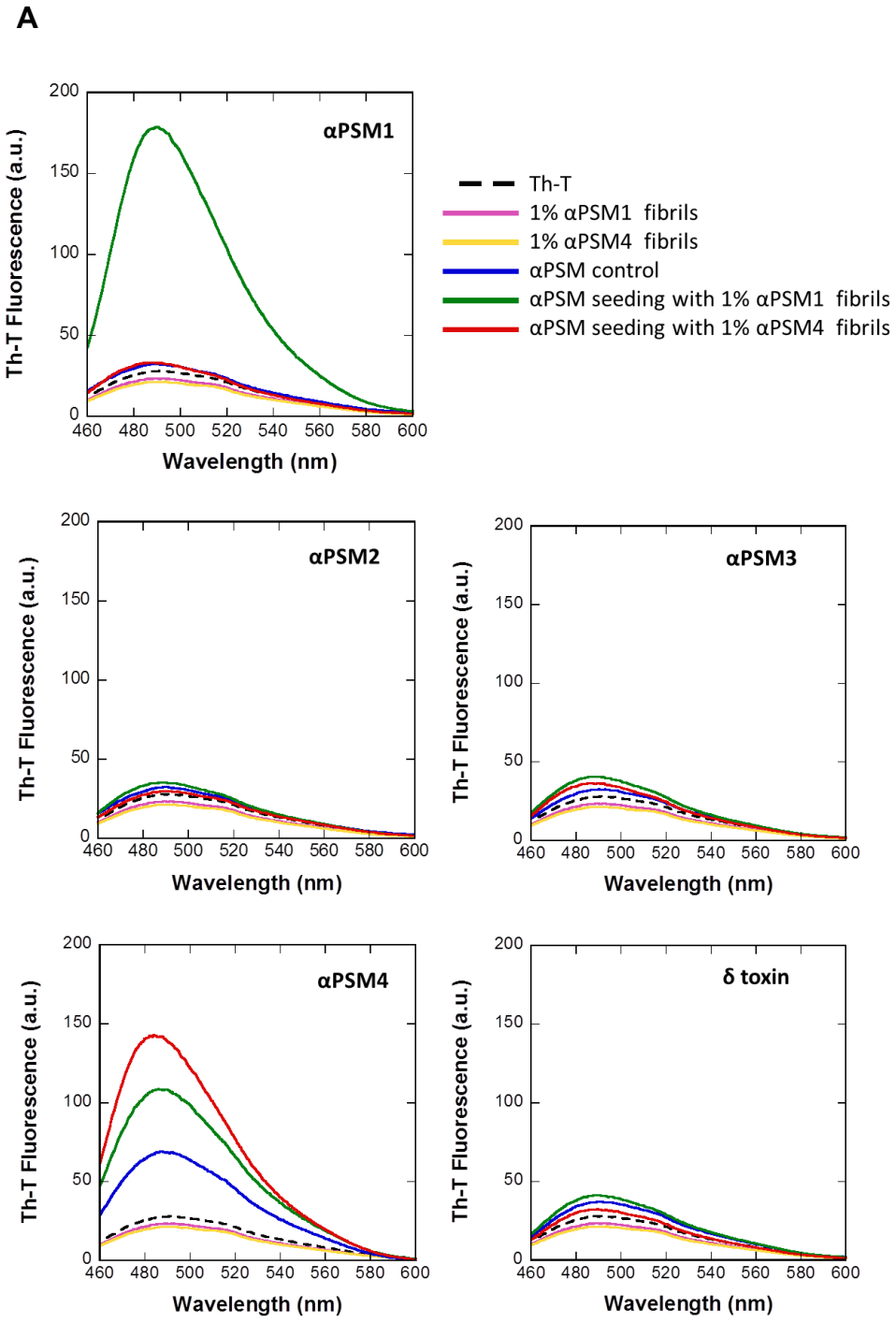

Figure IV-1.5 continued

Figure IV-1.5 continued

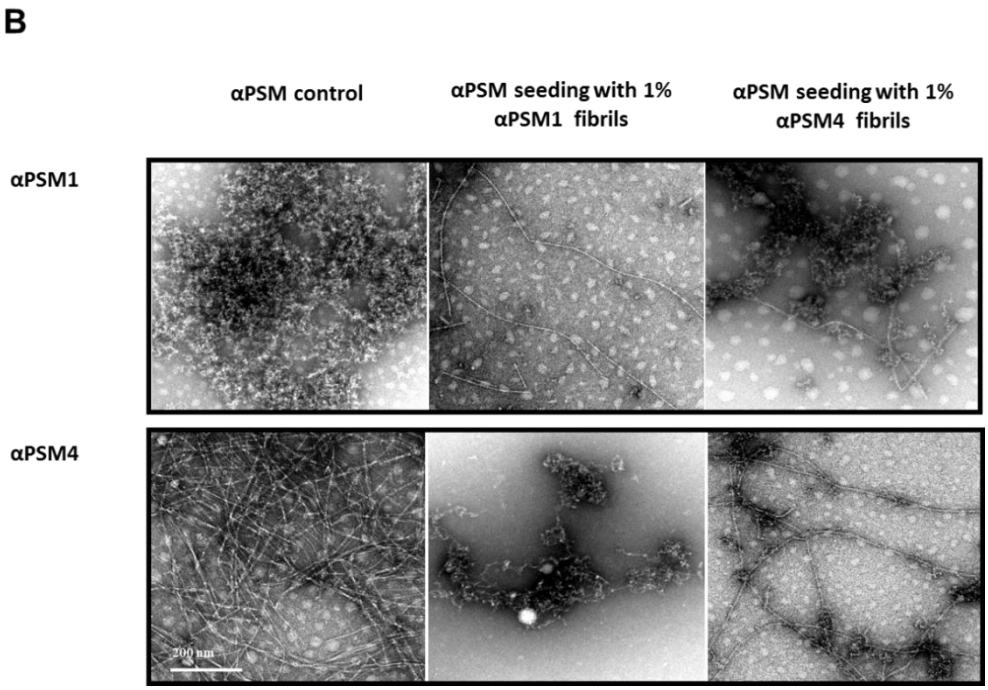

Figure IV-1.5| **Seeding and cross seeding test.** (A) PSMs seeding and cross-seeding Th-T emission spectra recorded after 6 days of incubation time are reported. In all the cases, Th-T free is represented by a black dotted line. (B) TEM micrographs of α-PSM1 and α-PSM4 in absence (α-PSM control) and in presence of α-PSM preformed fibrils are shown. The scale bar represents 200 nm.

In this way, seeding and cross-seeding experiments were carried out to test seeding selectivity for αPSM peptides. The αPSMs were incubated at 200 µM in their initial soluble states in the absence and in the presence of 1% (v/v) preformed PSM fibrils and analyzed by Th-T fluorescence after six days of incubation (Figure IV-1.5A). The presence of preformed homologous amyloid fibrils induces a large increase in Th-T maximum fluorescence emission compared to control peptide solutions, for both αPSM1 and αPSM4. The cross-seeding effect is observed only for αPSM1 which, when added to the αPSM4 solution, promotes its aggregation increasing Th-T fluorescence. However, analysis of αPSM1 and αPSM4 seeded and cross-seeded solutions by TEM (Figure IV-1.5B) shows that mature well-structured amyloid-like fibrils are detected only in homologous seeding reactions suggesting that the presence of heterologous fibrils can increase the formation of initial β-sheet enriched aggregates able to bind Th-T, but only an strict sequential identity allows an ordered incorporation of the heterologous monomers on top of seeds.

The ZipperDB algorithm (Thompson et al., 2006) was used to predict αPSM1 and αPSM4 hexapeptide segments corresponding to the putative amyloid core nuclei

(seed), which are supposed to initiate the self- assembly and might be responsible for cross-seeding effects (Figure IV-1.6A). αPSM4 shows more predicted aggregation-prone regions than αPSM1 according to its experimentally faster aggregation reaction. The αPSM1 and αPSM4 six-amino-acid stretches having lowest free energies, indicative of a high amyloid-forming potential, are shown in Figure IV-1.6B as PDB steric zipper structures. Their high similarity in sequence might explain why αPSM1 is able to accelerate the aggregation of αPSM4. The lack of cross-seeding when using αPSM4 preformed fibrils argues that likely amyloid formation in this peptide is leaded by alternative aggregation-prone low energy sequence stretches with lower homology to αPSM1 or that the homologous region is deeply buried in its amyloid conformation.

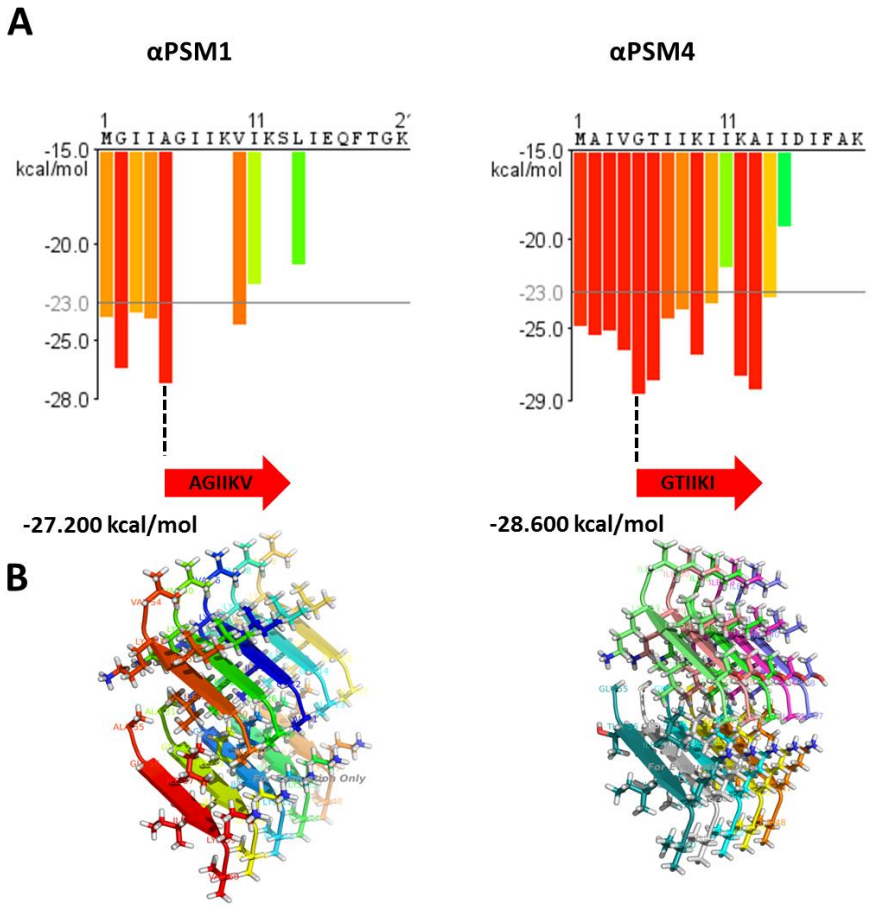

Figure IV-1.6| **αPSM1 and αPSM4 fibril-forming segments prediction by Zipper DB algorithm (Thompson et al., 2006).** (A) The propensity profile graph predicts hexapeptide segments (beginning at the indicated residue) that are highly likely to assemble into amyloid fibrils. Orange-red segments with energy values below the indicated threshold of −23 kcal/mol (gray line) are expected to form steric-zipper spines of fibrils. (B) The energetically favourable steric-zipper structures for both peptides is shown.

αPSM disaggregation and aggregation kinetics in the presence of EGCg

The natural compound epigallocatechin-3-gallate (EGCg) (Figure IV-1.7A) is the most abundant polyphenolic catechin of the green tea composing more than 50% of its leaves (Nagle et al., 2006). In the past two decades EGCg has been widely studied for its multiple activities as anti-proliferative (Milligan et al., 2009), anti-mutagenic (Šmerák et al., 2006) and anti-cardiovascular (Widlansky et al., 2007) effects, among others. Several works reported that this natural compound shows also antimicrobial properties against a wide variety of microorganisms (Das et al., 2014). It has been demonstrated that EGCg inhibits *S. epidermis* biofilm formation interfering with the extracellular polymeric material (glycocalyx) (Blanco et al., 2005). Interestingly, EGCg has also emerged as a potent inhibitor of Aβ peptide and α-synuclein amyloidogenesis *in vitro* (Ehrnhoefer et al., 2008).

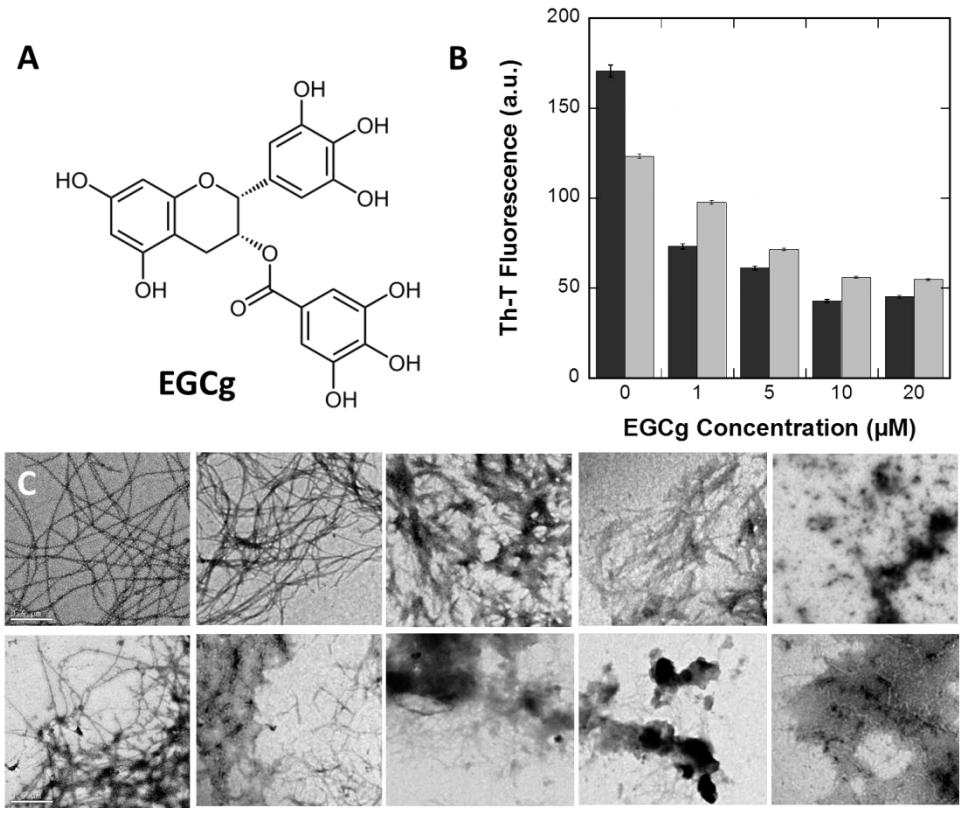

Figure IV-1.7| **Disaggregation kinetics of α-PSM1 and α-PSM4 fibrils by EGCg.** (A) Chemical structure of EGCg (B) Disaggregation of α-PSM1 (grey) and α-PSM4 (black) mature fibrils were followed by Th-T emission fluorescence at different EGCg concentration; error bars indicate ± SE (n= 2). (C) TEM micrographs of α-PSM1 and α-PSM4 preformed fibrils are displayed in the upper and bottom panels respectively, in presence of 0, 1, 5, 10 and 20 μM of EGCg (from left to right). The scale bar represents 0,5 μm.

To test if the previously reported activity of EGCg on biofilm formation might be mediated by an anti-amyloidogenic effect affecting PSM assemblies, αPSM disaggregation and aggregation kinetic assays were performed. αPSM1 and αPSM4 preformed fibrils (5% v/v) were dissolved in a 50 µM Th-T solution and incubated

during for 10 min with different EGCg concentrations. A concentration-dependent EGCg disaggregation effect is evident as revealed by the decrease in dye fluorescence emission signal for both αPSM fibrils (Figure IV-1.7B). This behavior was corroborated by TEM observation of the disaggregation reactions where long amyloid fibrils in the control solution (without EGCg) are progressively replaced by amorphous aggregates as the EGCg concentration increases (Figure IV-1.7C).

To study EGCg amyloid inhibitory effect, αPSM1 and αPSM4 peptide solutions were prepared at 200 µM in absence and in presence of 20, 100 and 200 µM of EGCg and incubated for one and two weeks (Figure IV-1.8). The inhibition of αPSM1 and αPSM4 aggregation kinetics was followed by monitoring the Th-T emission signal at different time points.

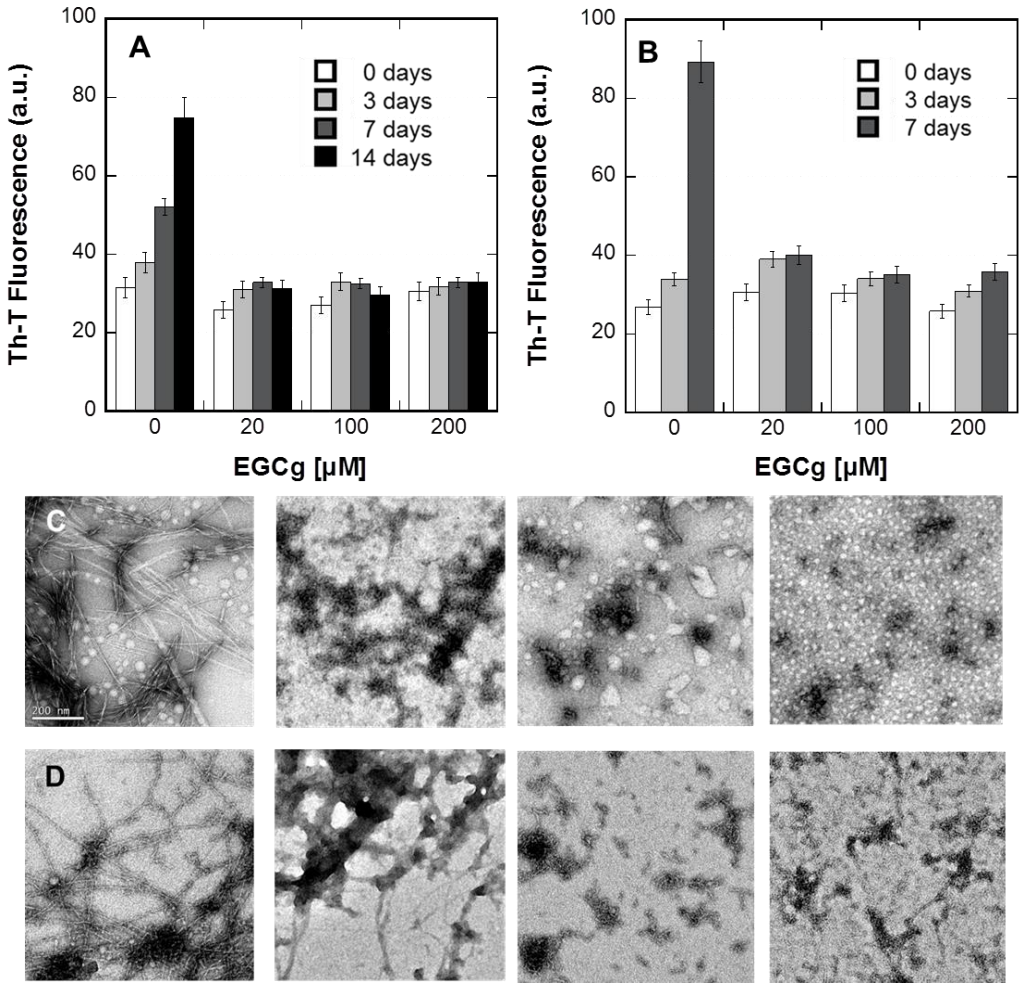

Figure IV- 1.8| **EGCg inhibition effect upon α-PSM1 and α-PSM4 aggregation kinetics.** Th-T kinetics of α-PSM1 (A) and α-PSM4 (B) peptide solutions in presence of EGCg have been monitored at different time points. Error bars indicate ± SE (n= 2). Inhibitory effect of EGCg upon α-PSM1 (C) and α-PSM4 (D) fibril formation at final point was also tested by TEM at 0, 20, 100 and 200 μM EGCg (from left to right). The scale bar represents 200 nm.

For both peptides EGCg inhibited amyloid formation at all tested concentrations. TEM morphological analysis of αPSM1 and αPSM4 solutions confirms the fibrillation inhibition in the presence of EGCg. Showing the absence of long ordered fibrils and the formation of smaller size aggregates (Figure IV-1.8C,D). These results confirm the EGCg effect on the amyloidogenic process. The disaggregational and inhibitory activity against αPSM amyloid formation suggests that the previously observed inhibitory activity on biofilm formation on *S. epidermis* could be mediated by this mechanism and suggests its applicability to *S.aureus* biofilm inhibition.

αPSMs expression in *Escherichia coli* model

The multiple roles executed by PSMs in *S.aureus* virulence have been demonstrated by several studies in the last years. PSM toxins, in particular αPSM3, may cause lysis of human neutrophils, stimulate inflammatory response or kill infected osteoblasts as recently reported (Wang et al., 2007; Rasigade et al., 2013; Laabei et al., 2014; Cheung et al., 2014a). PSMs are secreted through a specific ABC transporter, blockage of this transporter results in the accumulation of PSMs in the cytosol (Chatterjee et al., 2013). Cytosolic accumulation of PSMs is accompanied by abnormal cell division and severe damage to the cytoplasmic membrane. Therefore, the PSM has emerged as a drug target of exceptional value. However, it is not know which is the mechanism and the effectors of this cytotoxic effect. With the aim to understand the correlation between their aggregation potential and their toxicity properties, intracellular αPSM expression was analyzed *in vivo* using the *E.coli* bacterial system.

αPSM (1-4) peptides were fused to the green fluorescent protein (GFP) and expressed in *E. coli* cells according to a previous described approach (Wurth et al., 2002) which demonstrated how the presence of active GFP in bacterial intracellular aggregates (IBs) depends on the aggregation propensity of its fusion protein partner (de Groot et al., 2006). In this way, IBs fluorescence can be used as *in vivo* protein aggregation reporter since higher aggregation rate corresponds to lower GFP fluorescence and vice versa (Villar-Pique et al., 2012).

GFP fluorescence emission of intact cells expressing αPSM-GFP fusion proteins was measured by spectrofluorimetry (Figure IV-1.9A). A strong GFP emission signal for cells expressing αPSM3-GFP fusion protein is observed while lower fluorescence emissions are recorded for cells containing the other αPSM-GFP fusion proteins. Fluorescence microscopy was used to identify the cellular location of the detected GFP emission (Figure IV-1.9B). In all cases, the fluorescence is confined mainly in the IBs at the poles of the cell. However, αPSM3-GFP IBs are clearly more fluorescent than those formed by the other αPSM-GFP IBs, consistent with a lower αPSM3 aggregation propensity *in vivo*. After the αPSM-GFP cell extract lysis, anti-GFP Western Blotting was performed to analyse soluble and insoluble fractions (Figure IV-1.9C). All αPSM-GFP fusion proteins are located in the insoluble fraction and, as expected, only the αPSM3-GFP construct is also observed in the soluble fraction.

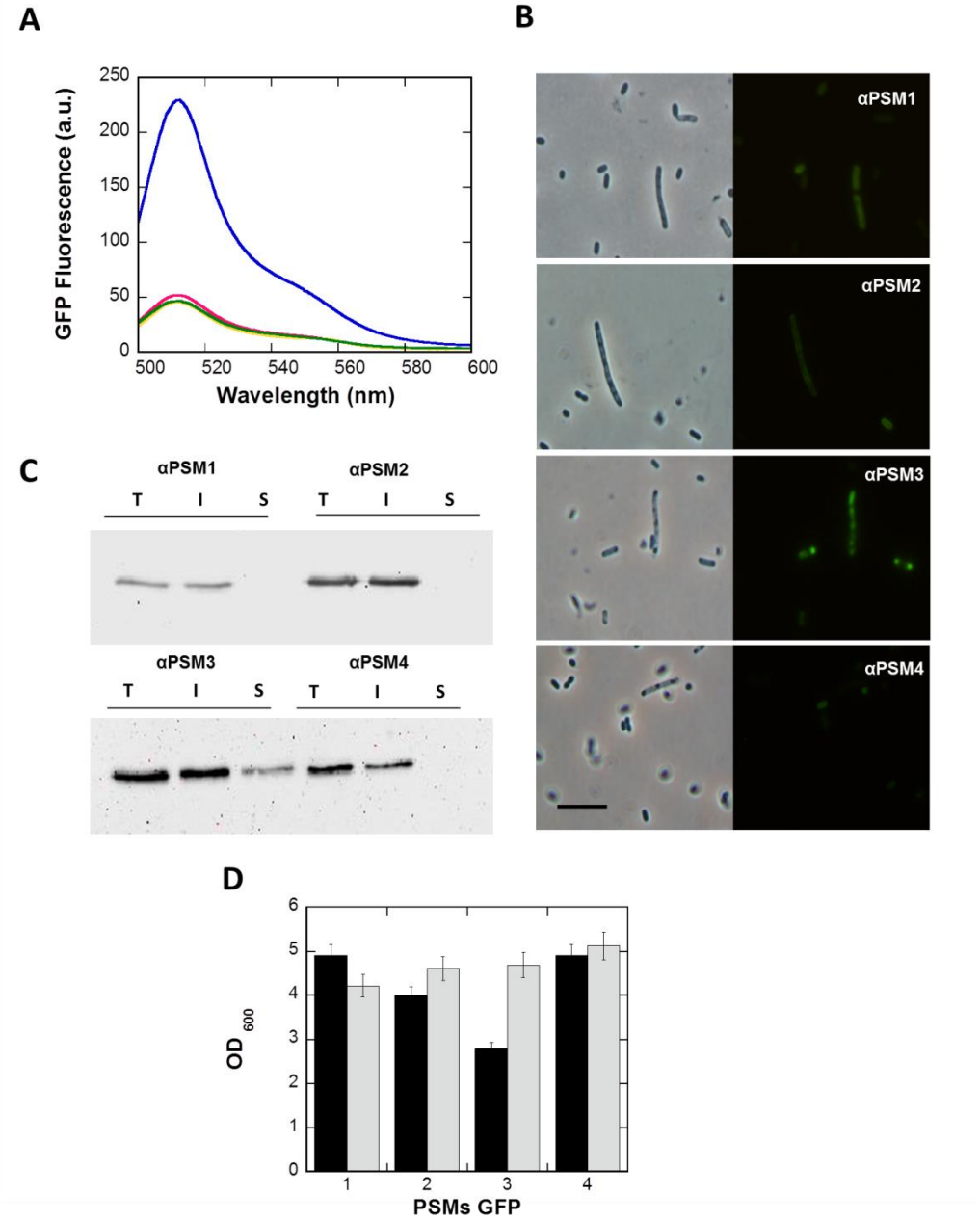

Figure IV-1.9| **Characterization of α-PSM GFP fusion protein expression in *E.Coli*.** (A) GFP fluorescence emission spectra of entire cells expressing α-PSM1 (red), α-PSM2 (yellow), α-PSM3 (blue) and α-PSM4 (green) fusions were collected. (B) α-PSMs GFP containing cells were observed using fluorescence microscopy (right panels) and phase-contrast microscopy (left panels); the scale bar represents 5 μm. (C) Total (T), insoluble (I) and soluble (S) cell fractions of α-PSMs fusions were detected by western blotting. (D) α-PSM fusion cells growth in presence (black) and in absence (grey) of IPTG was analyzed measuring OD_{600} after 20 h of expression. Error bars indicate ± SE (n= 3).

In order to study the toxicity *in vivo*, αPSM-GFP fusion proteins were expressed at 37˚C for 20h in presence and absence of IPTG and their growth was monitored spectrophotometrically measuring the optical density at 600 nm (OD_{600}). As shown in Figure IV-1.9D, OD_{600} measurements of cells containing αPSM1-GFP, αPSM2-GFP and

αPSM4-GFP fusion proteins do not show significant changes respect to non-induced controls, while of αPSM3-GFP induced sample shows a clear turbidimetric decrease if compared with the respective control. These data may suggest that intracellular αPSM-GFP aggregation does not exert a direct toxic effect on cell growth in this *in vivo* system whereas the expression of soluble αPSM3-GFP fusion protein may affect the *E.coli* growth rate. To analyze the αPSM solubility/toxicity correlation, αPSM-GFP fusion proteins were expressed at lower temperatures for 20h and intact cells were monitored by GFP fluorescence (Figure IV-1.10A) and by turbidimetric analysis (Figure IV-1.10B). The most and less aggregation prone αPSM4-GFP and αPSM-GFP3 fusion proteins, respectively, have been used in this assay. GFP fluorescence of αPSM-GFP cells expressed at 37, 25 and 18 ˚C was monitored showing in all cases an increase of emission signal for samples expressed at lower temperature indicating a major solubility respect to non-induced controls. However, cell growth decrease monitored by OD_{600} measurements is observed again only for αPSM-GFP3 fusion proteins indicating that the αPSM effect upon cell growth is not associated to the general αPSM solubility but rather to the specific αPSM3 toxicity.

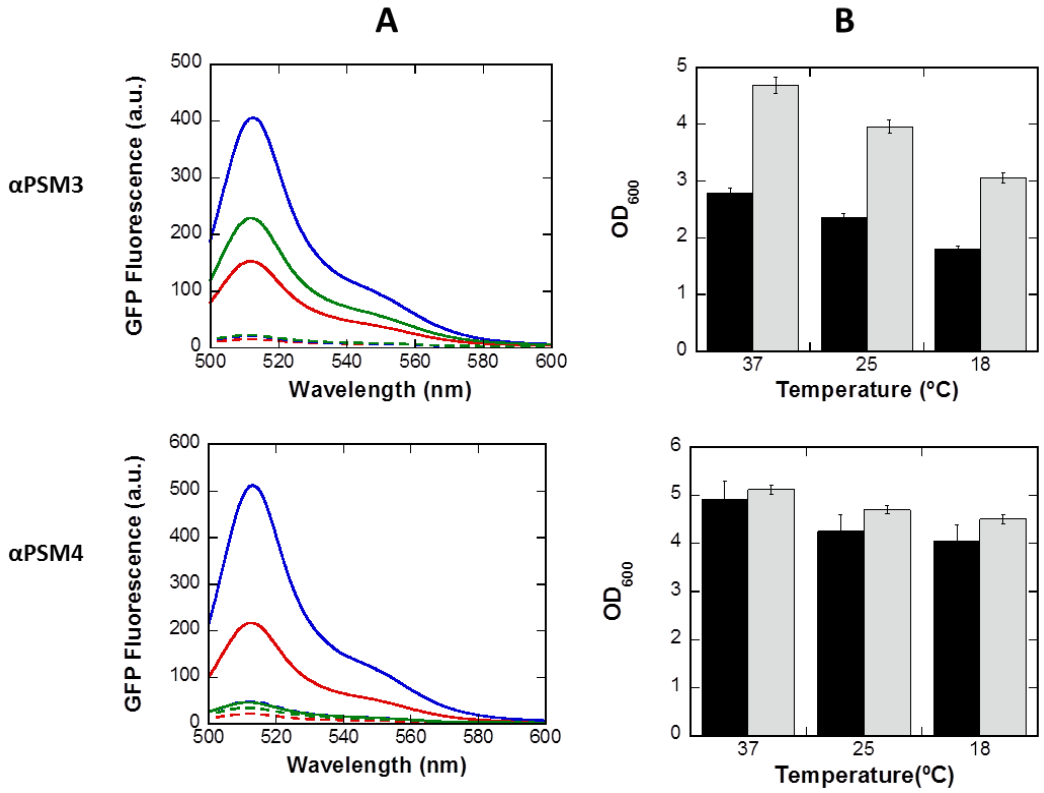

Figure IV-1.10| **Effect of temperature upon α-PSMs solubility and cell growth.**(A) GFP fluorescence was measured for intact cells expressing α-PSM3 and α-PSM4 fusion proteins at 37 (green), 25 (blue) and 18 (red) degrees in presence (solid line) and in absence (dotted line) of IPTG. (B) OD_{600} (20 h) of α-PSM3 and α-PSM4 fusion cells grown at different temperature, with (black) and without (grey) IPTG, was analyzed too. Error bars indicate ± SE (n= 3).

Discussion

S. aureus αPSMs are the smaller staphylococcal toxins belonging to PSM family. They have been well characterized because their pronounced cytolytic activity toward leukocytes (Wang et al., 2007; Cheung et al., 2012). Recently, many studies have reported their involvement in biofilm formation and detachment triggering the search of new insights about PSM function mechanisms in staphylococcal pathogenesis (Otto, 2013; Le et al., 2014). Interestingly, the PSMs act forming fibrillary amyloid structures that promote the integrity of *S. aureus* biofilms *in vitro* (Schwartz et al., 2012). In the present work, we have analyzed the *in vitro S. aureus* αPSMs-mediated amyloid formation mechanism in order to dissect their specific contribution to the biofilm structure and their intracellular toxicity *in vivo*.

Despite the high aggregation propensity predicted for all PSMs due to their similar structure and physicochemical properties, the use of the Amylpred2 consensus algorithm has allowed us to predict the αPSM4 peptide as the most aggregation prone and the αPSM3 peptide as the less aggregation prone (Table IV-1.1). These theoretical data have been confirmed *in vitro* demonstrating that contrary to what it was previously thought, not all PSMs form amyloid structures. Effectively, our results show that only αPSM1 and αPSM4 peptides are responsible for fibrillogenesis (Figure IV-1.2A) since the αPSM1 and αPSM4 aggregates bind to amyloid dyes, are enriched in intermolecular β-sheet conformation and seed freshly homologous peptide solutions. We have also demonstrated that the amyloid nature of *S. aureus* biofilm represents a potential target for the natural compound EGCg, which is active both inhibiting the aggregation of αPSMs and disrupting their preformed fibrils *in vitro*, as reported for other amyloidogenic proteins (Rezai-Zadeh et al., 2005; Ehrnhoefer et al., 2008 ; Bieschke et al., 2010). Besides, the *in vivo* αPSMs expression in *E. coli* reveals that αPSMs aggregation does not perturb cell fitness consistent with a functional role of αPSM amyloid aggregation in the *S. aureus* biofilm formation, a process in which they are expected to be innocuous. Conversely, the less aggregation prone αPSM3 is expressed also in the soluble form and shows a specific capability to affect cell growth in our bacterial system in agreement with the pronounced toxic extracellular activity described for this peptide (Wang et al., 2007). As discussed above, all PSMs share an amphipathic α-helical structure. Remarkably, the degree of helicity of these peptides has been described as a critical property linked to their lytic activity (Cheung et al., 2014b; Laabei et al., 2014). This effect seems to be driven mainly by the hydrophobic character of α-helical motif since hydrophobic residues promote the PSM aggregation, concomitantly decreasing the membrane lytic capability by lowering the interaction with negatively charged membranes, as described in previous studies on α-helical antimicrobial peptides (AMPs) (Kim & Hecht, 2006 ; Giangaspero et al.,2001).

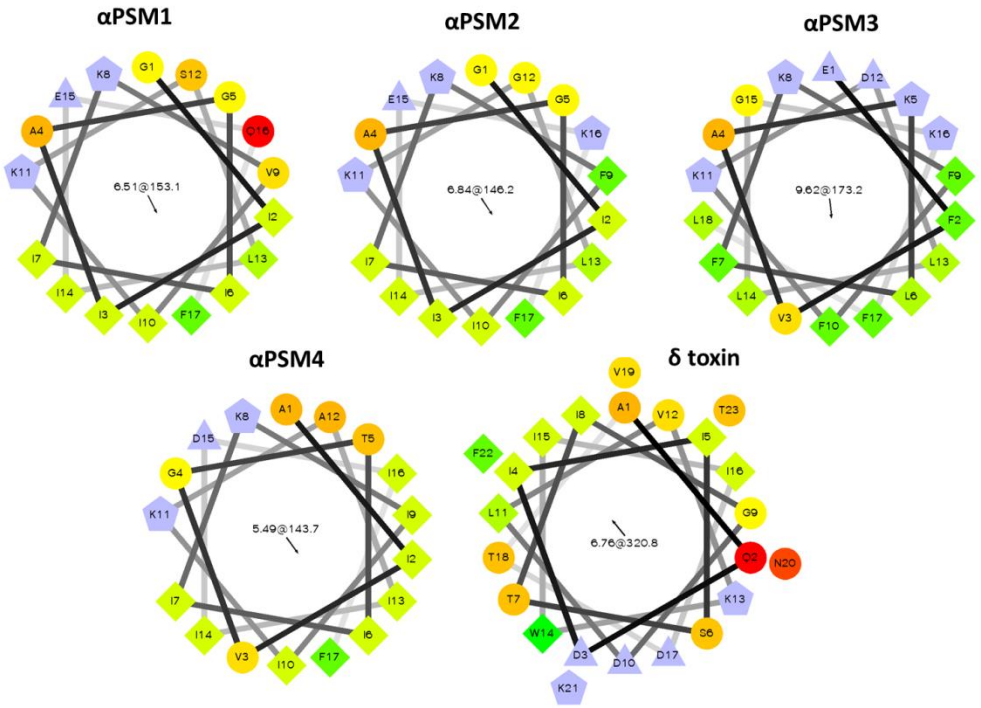

Figure IV-1.11| **Helical wheel projections of αPSMs.** αPSMs sequence forming α helical structures predicted through JPRED4 server () are represented. Hydrophobic residues are color coded from green (most hydrophobic) to yellow (zero hydrophobicity). Hydrophilic residues are coded with pure red (the most hydrophilic), and the amount of red decreasing proportionally to the hydrophilicity. The potentially charged residues are light blue. Helical wheel projections were performed using the Helical Wheel Plotting program by (http://rzlab.ucr.edu/).

In order to understand the "weight" of hydrophobicity in the α-helical structure of each PSMs we have analyzed the five sequences with the JPRED4 program (Drozdetskiy et al., 2015) to predict the stretch of amino acidic residues forming the α-helix and then represented them using helical wheel projections (Figure IV-1.11). We have calculated the hydropathy index of the predicted sequences using the GRAVY value (Kyte & Doolittle, 1982) as shown in Table IV-1.3. As expected, αPSM4 shows the most hydrophobic score (2.01) while αPSM3 (0.67) and δ toxin (0.43) result to be the less hydrophobic peptides in good agreement with their distinct functions. αPSM1 and αPSM2 peptides show a very similar hydropathy index which can explain why both aggregate intracellularly in bacteria despite they do not share the same capability to participate in amyloid formation *in vitro,* a property more linked to the sequential and structural properties than to the hydrophobic character. Interestingly, the hydropathy calculation of the residues belonging to hydrophobic and hydrophilic halves of the wheel projections suggests that the hydrophilic residues contribution could modulate the final solubility of the peptides, αPSM4 and αPSM1 displaying the less negative GRAVY values at the hydrophilic face. In this context the presence of the residue in position 16 may play a critical role since in PSM2 and PSM3 is a Lys that interrupts the

hydrophobic face, whereas in PSM1 this residue has a neutral (Gln) and in PMS4 a hydrophobic (Ile) nature, respectively, allowing a larger continuous hydrophobic interface. Overall this indicates that the PSM function might be driven by the fine balance of hydrophobic/ hydrophilic forces into the α-helical structure, which relies indeed in a few aminoacidic diferences.

Table IV-1.3| **αPSMs hydropathy calculated.** The GRAVY value is calculated by adding the hydropathy values (Kyte & Doolittle, 1982) of each amino acid residues and dividing by the number of residues in the sequence. Increasing positive score indicates a greater hydrophobicity.

	Sequence[a]	T.GRAVY[b] value	P.GRAVY[c] value	P.GRAVY[d] value
αPSM1	GIIAGIIKVIKSLIEQF	1.36	4.2	-1.15
αPSM2	GIIAGIIKFIKGLIEKF	1.28	4.0	-1.82
αPSM3	EFVAKLFKFFKDLLGKFL	0.67	3.3	-2.65
αPSM4	AIVGTIIKIIKAIIDIF	2.01	4.3	-1.25
δtoxin	AQDIISTIGDLVKWIIDTVNKFT	0.42	3.5	-2.40

[a] αPSMs sequence represented as wheel projection in Figure IV-1.11.

[b] GRAVY value calculated for the fully α helical structures predicted by JPRED4 server in a.

[c] GRAVY value calculated for the residues (underlined in the sequence[a]) belonging to the hydrophobic half of wheel projection in Figure IV-1.11.

[d] GRAVY value calculated for the residues belonging to the hydrophilic half of wheel projection in Figure IV-1.11.

Methods

Peptides and preparation

αPSM1 (MGIIAGIIKVIKSLIEQFTGK), αPSM2 (MGIIAGIIKFIKGLIEKFTGK), αPSM3 (MEFVAKLFKFFKDLLGKFLGNN, αPSM4 (MAIVGTIIKIIKAIIDIFAK) and δ-toxin (MAQDIISTIGDLVKWIIDTVNKFTKK) peptides (ChinaPeptides, Shanghai, China) with a purity > 95% were dissolved to a final concentration of 0.5 mg/mL in a 1:1 mixture of trifluoroacetic acid (TFA) and hexafluoroisopropanol (HFIP). Peptides were then sonicated for 10 minutes and incubated for 1 h at room temperature. Stock solutions were divided into aliquots, solvent TFA/HFIP dried with a speed vac at room temperature and stored at -80ºC. All assays were performed with equal stoichiometric ratios of 0.5 mg/mL (200 µM) peptide unless otherwise noted. When required, samples were resuspended in anhydrous dimethyl sulfoxide (5%) and sonicated for 10 minutes. Sonication was crucial for removing any trace of undissolved seeds that may

resist solubilization. This preparation yielded phenol soluble modulins in monomeric form. Peptide aliquots were prepared in MilliQ water yielding a final peptide concentration of 200 µM.

Prediction of amyloid sequence stretches and aggregation propensities

PSM peptides aggregation-prone regions prediction was performed using the consensus method Amylpred2 (Tsolis et al., 2013) that employs 11 different algorithms to analyze and identify the consensus region for each peptide. In this analysis the following algorithms were included: AGGRESCAN, AmyloidMutants, Amyloidogenic Pattern, Average Packing Density, Beta-strand contiguity, Hexapeptide Conf. Energy, NetCSSP, Pafig, SecStr, TANGO and WALTZ, employing default settings.

Thioflavin-T (Th-T) binding

200 µM peptide solutions were incubated at room temperature. Peptide aggregation was monitored by measuring the transition from the non-aggregated to the aggregated state by relative Th-T (50 µM) fluorescence at 490 nm upon excitation at 440 nm. In the seeding and cross-seeding assays, 1% (v/v) preformed aggregates were added to peptide solution and incubated during 6 days at room temperature. For aggregation monitoring, 50µM Th-T dye was added to 10 µM peptide solution. Spectra were recorded as the accumulation of three consecutive scans between 460 and 600 nm with an excitation wavelength of 440 nm on a Jasco FP8200 spectrofluorometer (Jasco, Japan). For all measures, a 5 nm slit width was used for both excitation and emission.

Congo Red binding

Congo Red (CR) binding with aggregated peptides was analysed in the range 400 – 700 nm using a 1 cm optical length quartz cuvette placed in a thermostated cell holder at 25 °C on a Cary-400 UV/Vis spectrophotometer (Varian Inc). 20µM CR was mixed with 12 µM peptide sample solution. The differential CR spectrum in the presence and absence of peptide solution was analyzed in order to detect a typical amyloid band at 540 nm.

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

ATR-FTIR analysis was performed using a Bruker Tensor 27 FTIR Spectrometer (Bruker Optics Inc.) with a Golden Gate MKII ATR accessory. Aggregated peptide solutions were dried out under a N_2(g) atmosphere and each spectrum was measured as the accumulation of 16 scans at a spectral resolution of 2 cm^{-1} within the range 1800–1500 cm^{-1}. All infrared spectra were deconvoluted into overlapping Gaussian curves by employing the non-linear fitting program PeakFit v4.12 (Systat Software Inc.).

Circular Dichroism analysis

Far-UV CD spectra were measured in a Jasco-710 spectropolarimeter (Jasco, Japan) thermostated at 25 ºC. Aggregated peptides were prepared at 20 µM and measured immediately. Spectra were recorded from 260 to 200 nm, at 0.5 nm intervals, 1 nm bandwidth, and a scan speed of 100 nm/min. For each spectrum twenty accumulations were averaged.

Transmission electron microscopy (TEM)

Fibrillar peptide solutions were resuspended in water, placed on carbon-coated copper grids, and left for 5 min. The grids were then washed twice with distilled water, stained with 2% (w/v) uranyl acetate for a minute, and dried out before analysis. TEM JEM-1400 microscope was used operating at an accelerating voltage of 120 kV.

Epigallocatechin gallate (EGCg) disaggregation and inhibition assays

To test the effect of EGCg disaggregation, 5% v/v preformed fibrils were added to 50 µM Th-T solution and incubated with 0, 1, 5, 10 and 20 µM EGCg (99% pure; Sigma). After 10 min of incubation time Th-T fluorescence emission at 490 nm was recorded for each sample. Peptide aggregation kinetics were prepared at 200 µM in presence of 0, 10, 20 and 200 µM of EGCg and monitored as described in the previous section.

Cloning and expression of PSM-GFP fusions

Single stranded oligonucleotides (ssDNA) corresponding to the PSM peptides (αPSM1, αPSM2, αPSM3, αPSM4) were purchased from Invitrogen (Thermo Fisher Scientific Inc.). PSM ssDNAs were phosphorylated and annealed to obtain double stranded oligonucleotides (dsDNAs). A pET28a vector already containing a fusion of Amyloid β-42 (Aβ-42) peptide with the Green Fluorescent Protein (GFP) (de Groot et al., 2006) was engineered by PCR reaction to remove Aβ-42 gene and introduce the PSM insert upstream of the GFP sequence finally encoding an N-terminal PSM-GFP fusion protein. All constructs were verified by DNA sequencing. Plasmids were then transformed into Escherichia coli BL21 (DE3) cells and cultures were grown in Luria Broth (LB) medium, containing appropriate antibiotics, with agitation (250 rpm) at 37 °C (25 or 18 ˚C when needed). Protein expression was induced at an OD600 of 0.5, with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 20 h.

Cell optical density (OD) and Fluorescence measurements

Cell cultures induced for 20 hours were harvested at 3000 rpm and washed in PBS for three times. Optical density at 600 nm (OD_{600}) was recorded for each sample on a Cary 100 UV/Vis spectrophotometer (Varian Inc., Palo Alto). GFP fluorescence of intact cells at an OD_{600}= 0.1 was recorded in the range of 500–600 nm, using an excitation wavelength of 470 nm on a Jasco FP8200 spectrofluorometer (Jasco, Japan). For microscopic analysis, 10 µL of washed intact cells were deposited on top of glass slides

and imaged with a phase-contrast and fluorescence microscopy under UV light using a Leica Q500 MC fluorescence DMBR microscope (Leica Microsystems, Mannheim, Germany).

Western Blotting

Cell pellets were resuspended in PBS at an OD_{600} =10. Samples were sonicated and centrifuged at maximum speed for 20 min at 4ºC. The soluble, insoluble and total fractions were resolved on 18% w/v SDS-PAGE gel and transferred onto a PVDF membrane. Immunodetection was performed using anti-GFP antibody (BD Biosciences) and membranes were developed with ECL Chemiluminescent HRP Substrate (Millipore) according to the manufacturer's protocols. Densitometry bands analysis was performed using the Image J software.

2 Non-functional aggregation

Non-functional aggregation is associated to the onset of an increasing number of human disorders, from neurodegenerative diseases such as Alzheimer's and Parkinson's disease, through transmissible prionic encephalopathies, to non-neurodegenerative amyloidoses such as type II diabetes (Fernandez-Busquets et al., 2008; Invernizzi et al., 2012; Dobson, 2003). Despite the proteins being involved in these pathologies are not related in sequential or structural terms, in many cases, their aggregation involves the acquisition of a common cross-β conformation leading to the formation of amyloid fibrils. Moreover, the adoption of such macromolecular structures is not restricted to disease-linked polypeptides and might constitute a generic feature of proteins chains (Dobson, 2004; Jahn & Radford, 2005). It was reported that in globular proteins there are almost three times much more aggregation prone regions than in intrinsically disordered proteins (IDPs) and that the formation of highly structured globular proteins competes with β-aggregation propensity because both structure and aggregation obey very similar physicochemical constraints (Linding et al., 2004). In this context, the origin of protein aggregation and amyloid formation is poorly understood for IDPs that do not have a fixed three-dimensional structure in physiological conditions. Some IDPs are resistant to protein aggregation while others are directly involved in amyloid formation. What makes some IDPs foldable or aggregation prone is still an open question (Tompa, 2002; Zhang et al., 2013).

In the second part of this thesis, the semi-disordered heterogeneous nuclear ribonucleoprotein D-like (HNRPDL) and the globular yeast URN1 FF domain were used to investigate the molecular determinants of non-functional aggregation in two structurally different protein models; a brief description of these determinants analysed in the next chapters is given hereunder.

Semi-disordered proteins and β-aggregation

Human RNA binding proteins containing intrinsically disordered regions with an amino acid composition resembling those of yeast prion domains (PrLD) are being found to aggregate in different neurodegenerative disorders. These proteins include: (i) fused in sarcoma (FUS), TAR DNA-binding protein 43 (TDP-43), EWSR1 and TAF15, involved in amyotrophic lateral sclerosis (ALS) and/or some forms of frontotemporal lobar degeneration (FTLD) (Da Cruz & Cleveland, 2011; Geser et al., 2009; Couthouis et al., 2012; Neumann et al., 2011), (ii) hnRNPA2B1 and hnRNPA1, linked to familial inclusion body myopathy with Paget's disease of bone, frontotemporal dementia and ALS (Kim et al., 2013) and (iii) TIA1, a protein associated with Welander distal myopathy (Klar et al.,2013). Despite TDP-43 is perhaps the best characterized of these PrLDs-containing proteins it still not clear whether the pathological aggregates formed by this protein

have an amorphous or an amyloid nature. The difficulty of purifying soluble TDP-43 makes challenging to decipher this issue by means of classical *in vitro* aggregation studies. Chiti and co-workers have circumvented this limitation using bacteria to model intracellular TDP-43 aggregation. Interestingly, despite TDP-43 IBs were toxic to neuroblastoma, they did not exhibit amyloid signatures and were structurally amorphous (Capitini et al., 2014). This observation raises the question of whether it is indeed the lack of an ordered structure in these human prion-like protein aggregates the underlying cause of their toxicity in different pathological states.

Globular proteins and amyloid formation

In many occasions pathological protein aggregation and fibril formation result from the population of partially folded conformations (Vivekanandan et al., 2011; Bemporad et al., 2006; Sabate et al., 2012). In native structures, proteins locate most of their hydrophobic side-chains and main-chain hydrogen bond formers inaccessible to solvent, thus reducing anomalous intermolecular bonding and aggregation. In the opposite case, strong denaturing conditions usually render unfolded, but soluble protein conformations because the non-covalent inter- and intramolecular contacts that a protein may form are impeded. This explains why conditions promoting the population of partially folded conformations become more favourable to aggregation than native or unfolded states.

At mild concentrations, the fluorinated alcohol 2,2,2-trifluoroethanol (TFE) is a co-solvent that diminishes native tertiary interactions and induces partly folded states of polypeptides by disrupting hydrophobic contacts without preventing establishment of intermolecular interactions (Otzen, 2010b). In this respect, a TFE promoted partly folded state and an aggregation-prone conformational state of a pathological protein are similar. Most of the proteins for which TFE-induced amyloid formation has been studied consist of at least one preformed β-sheet element in their native structure. Intermolecular docking between these destabilized and partially exposed secondary structures in TFE-promoted partially folded states is thought to trigger the initial self-assembly reaction of these polypeptides (Srisailam et al., 2003; Gosal et al., 2004; Munishkina et al., 2003; Pallares et al., 2004). In contrast, little is known about the effect of TFE on the aggregation of all α-helical proteins, where preformed β-sheet conformations are not present in their aggregation-competent states.

As discussed previously, conditions that destabilize the native structure or increase unfolding rates, such as mutations or environmental conditions, are related to pathological states in different protein models (Foss et al., 2005; Booth et al., 1997). In this way, proteins have evolved disulfide bonds arguably in defence from harsh environments, such as the extracellular space, since these covalent bonds can strongly stabilize proteins reducing their conformational fluctuations (Arolas et al., 2005). Accordingly, it is described that disulfide bonds can decrease the entropy of the unfolded ensemble, making it less favourable compared with the folded conformation

(Pace et al., 1988). Moreover, disulfide bridge formation affects not only the stability of the protein at equilibrium but also, in many cases, its mechanism and/or kinetics of folding (Goldenberg & Creighton, 1984; Zhou, 2003). Hence, engineering of new disulfide bonds and analysis of the corresponding changes in folding kinetics represents a useful strategy to study the proximity of specific protein regions in the transition state ensemble and analyse the disulfide role in misfolding and aggregation phenomena (Graña-Montes et al., 2012; Dombkowski et al., 2014).

Other important cause that leads to the formation of aberrant intracellular protein deposits is the reactive oxygen species-mediated damage, considered as the main contributor to the aging process (Berlett & Stadtman, 1997). Reactive Oxygen Species (ROS) can modify irreversibly the proteins structure compromising their functional activity, conformation and stability (Sun et al., 1999; Zhou et al., 2014). The oxidation of a single amino acid alone may perturb specific inter-protein and protein-solvent interactions leading to unfavorable conformational rearrangements that result in more stable non-native minima (Jahn & Radford, 2005; Eichner & Radford, 2011). Therefore, these oxidative modifications can bring to different changes in both the aggregation propensity and/or fibrillazation kinetics as described for example in the case of the Apolipoprotein A-I (apoA-I) oxidation, where the reduced conformational stability and oligomeric self-association leads to misfolding of the protein and formation of amyloid-like fibrillar structures (Wong et al., 2010) or the amyloid Aβ-(1–42) peptide oxidation that significantly delays the rate of fibril formation altering the characteristic Aβ fibril morphology and constraining formation of protofibrils, which are key intermediate in β-amyloidosis and the associated neurotoxicity (Hou et al.,2002).

2.1 The prion-like RNA-processing protein HNRPDL forms inherently toxic amyloid-like inclusion bodies in bacteria

As discussed above, the aggregation of human RNA binding protein prion-like TDP-43 has been modelled in bacteria, showing that TDP-43 inclusion bodies (IBs) are amorphous but intrinsically neurotoxic (Capitini et al., 2014). In the present work, the IBs formed by heterogeneous nuclear ribonucleoprotein D-like (HNRPDL) were characterized to address whether the lack of ordered structure in protein deposits is a common property of pathogenic human PrLD-containing proteins.

HNRPDL displays a predicted amyloidogenic prion-like domain at the C-terminus

HNRPDL is a 420 residues long protein for which no structural information is available yet. Both SMART (http://smart.embl-heidelberg.de) and PFAM (pfam.sanger.ac.uk/) databases coincide to indicate the presence of two contiguous canonical RNA recognition motifs (RRM) including residues 149–221 and 234–306, occupying a central position in the protein (Figure IV-2.1.1).

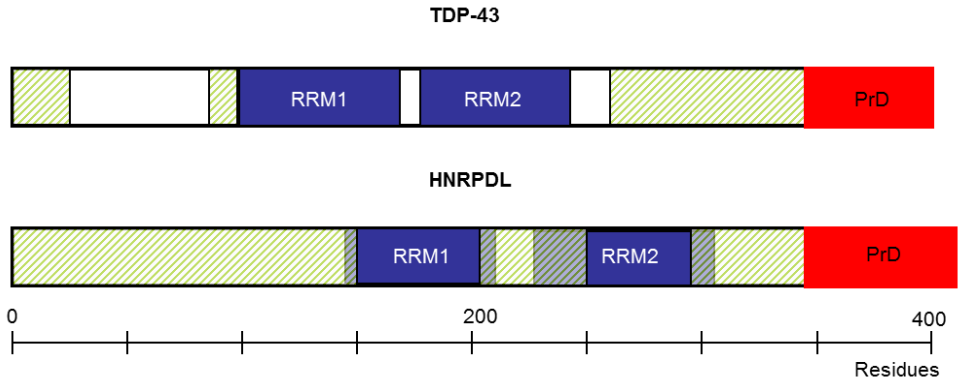

Figure IV-2.1.1| **TDP-43 and HNRPDL domain architecture.** Cartoons of proteins TDP-43 and HNRPDL show the domain architecture, where RRM accounts for RNA recognition motif and are represented in blue, and predicted disordered regions and prion domains (PrD) are shown in striped green and red, respectively. The places where RRM domains as assigned according to PFAM overlay with disordered predicted regions were assumed to correspond to canonical RRM domains.

Both the N- and C- terminal boundaries of these small domains are predicted to be low complexity regions without any associated function or structural motif. Disorder predictions using FoldIndex (Prilusky et al., 2005), FoldUnfold (Galzitskaya et al., 2006), and RONN (Yang et al., 2005) algorithms suggest that both the 1–149 and 306–420 sequence stretches are essentially disordered (Figure IV-2.1.1). The amino acid compositional bias of Q/N enriched prion domains has allowed the recent development of three different algorithms to identify the presence of PrLDs in protein sequences: PAPA (Toombs et al., 2012), PLAAC (Lancaster et al., 2014) and PrionScan (Espinosa Angarica et al., 2014). No prionic propensity is predicted with any of these

programs for the N-terminal segment, whereas all of them identify the C-terminal region as displaying a PrLD comprising residues 340–420. Overall, this domain architecture and PrLD location recapitulates that of TDP-43 (Figure IV-2.1.1; Table IV-2.1.1).

Table IV-2.1.1| **Prediction of PRLDs and their amyloid cores potency in the sequences of HNRPDL and TDP-43 RNA binding proteins**

	Uniprot	PrD			pWaltz score
		PAPA	PLAAC	Prion scan	
HNRPDL	O14979	337-417	339-418	341-420	82.27
TDP-43	Q13148	346-416	299-378	341-417	68.16

It has been recently shown by our group that the identification and evaluation of the potency of amyloid nucleating sequences in the context of disordered Q/N rich protein segments allows discrimination of genuine yeast prions from non-prionic sequences displaying very similar amino acid composition, a concept that was implemented in the pWALTZ algorithm (Sabate et al., 2015). The C-terminal PrLD of HNRPDL displays a pWALTZ score (82.27) higher than the corresponding PrLD in TDP-43 (68.16) (Table 1) and, strikingly, higher than those of Ure2p (73.99) and Sup35 (73.66) prion domains (Sabate et al., 2015), thus indicating the presence of an amyloidogenic sequence stretch comprising residues 342–362 in this Q/N rich disordered protein region.

Aggregation of HNRPDL into IBs in bacteria

The inherent aggregation propensity of human amyloid proteins results in most of them aggregating into insoluble IBs when they are produced in bacteria (Sabate et al., 2010). To test if this is the case of HNRPDL, the cellular distribution of the recombinant protein was analysed after its expression in *E. coli* at 37°C for 20 h. As assessed by SDS-PAGE, a new protein band of ≈50 kDa, corresponding to the expected HNRPDL molecular weight (47 kDa), could be detected in induced cells (Figure IV-2.1.2A). The bacteria cells were harvested, lysed and centrifuged and the resulting supernatant and pellet fractions were analysed by SDS PAGE. HNRPDL was found essentially in the insoluble fraction suggesting that it likely aggregated into IBs (Figure IV-2.1.2A). The protein remained in the insoluble fraction when protein expression was induced at either 25 or 18°C (data not shown).

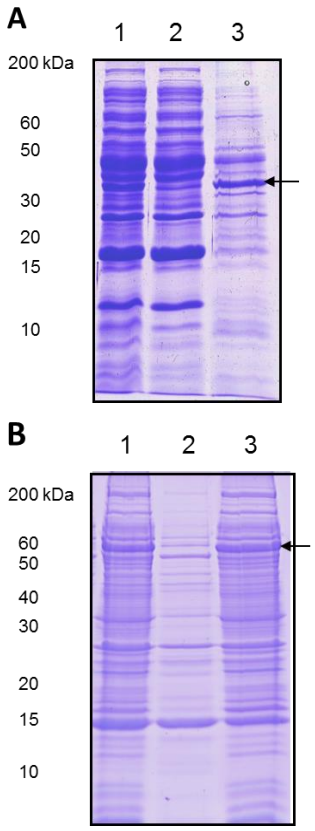

Figure IV-2.1.2| **Expression of recombinant HNRPDL protein in E. coli cells.** (A) Analysis on SDS-PAGE of *E. coli* cells extracts expressing HNRPDL protein. (B) SDS-PAGE analysis of cell extracts from cells expressing the GST-HNRPDL fusion. On both gels lane 1 shows total extract; lane 2, soluble fraction (supernatant), and lane 3 insoluble fraction (pellet). Arrows indicate the bands corresponding to HNRDPL protein.

Further, the HNRPDL cDNA was cloned downstream of the GST gene in a pETM30 vector and expressed the fusion protein at 20°C for 20 h. A new protein band of ≈75 kDa was observed for induced cells, corresponding to the sum of the molecular weights of GST (26 kDa) and HNRPDL (47 kDa) (Figure IV-2.1.2B). Fractionation indicated that despite the theoretical solubility provided by GST, the fusion was located in the insoluble fraction (Figure IV-2.1.2B) a localization that was maintained when protein expression experiments were performed at lower temperatures (data not shown). Because RRM domains are known to be soluble at high concentrations (Wang et al., 2014) and no aggregation-prone region is detected at the disordered N-terminal segment using predictive algorithms like AGGRESCAN (de Groot et al., 2012) or TANGO (Fernandez-Escamilla et al., 2004), it is likely that the predicted amyloidogenicity of the prion-like C-terminal region would account for the propensity of HNRPDL to form intracellular aggregates, either alone or when fused to GST.

HNRPDL IBs bind to thioflavin-S in living cells

As recently shown, the thioflavin-S (Th-S) staining of living bacterial cells can be used to detect the presence of intracellular amyloid-like structures as well as to find inhibitors that interfere with amyloid formation (Pouplana et al., 2013; Espargaro et al., 2012a). The staining of cells expressing HNRPDL was monitored using confocal microscopy. As it can be observed in Figure IV-2.1.3A, induced cells exhibited a green fluorescent background with strong fluorescent foci located at the cell poles, suggesting that HNRPDL adopts amyloid-like conformations in bacterial IBs.

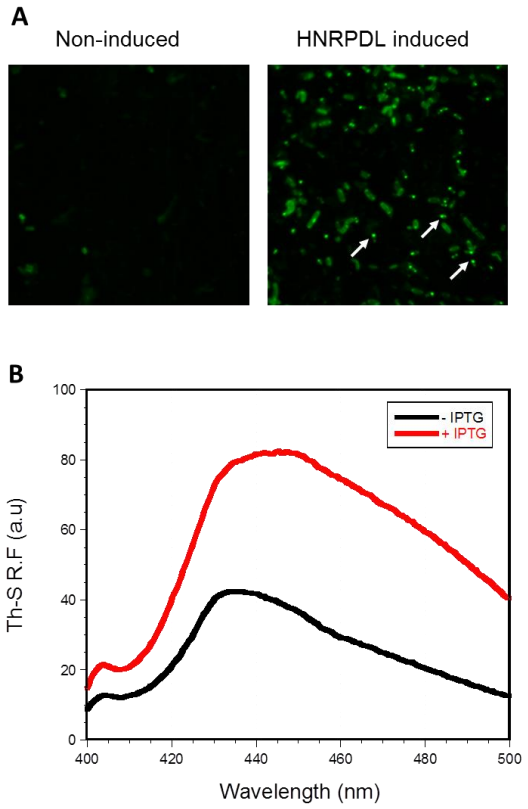

Figure IV-2.1.3| **Th-S staining of cells expressing HNRPDL.** (A) Fluorescent confocal microscopy images of non-induced E. coli cells and expressing HNRPDL IBs stained with Th-S at ×100 magnification. (B) Fluorescence spectra of Th-S in the presence of non-induced (−IPTG) and induced (+IPTG) living cells expressing HNRPDL. Arrows indicate the position of IBs.

In contrast, non-induced control cells exhibited only residual fluorescence. The presence of intracellular amyloid-like protein conformations in induced cells could also be monitored using fluorescence spectroscopy. As previously described for cells expressing Aβ42 (Espargaro et al., 2012a), the Th-S fluorescence maximum increases and red-shifts in the presence of living cells expressing HNRPDL, relative to the Th-S fluorescence maximum recorded in the presence of non-induced cells (Figure IV-2.1.3B).

Purified HNRPDL IBs bind to amyloid dyes

Next, the HNRPDL IBs were purified to characterize biophysically their amyloidogenic properties. Using SDS-PAGE densitometry it was calculated that HNRPDL constituted around 30% of all proteins in the purified IBs fraction (Figure IV-2.1.4). To evaluate the specific contribution of HNRPDL in the different assays, relative to that of other proteins present in this fraction, cells bearing the same plasmid without any insert were induced and the IBs fraction purified in the same manner than those containing the HNRPDL cDNA and used as negative control (Figure IV-2.1.4).

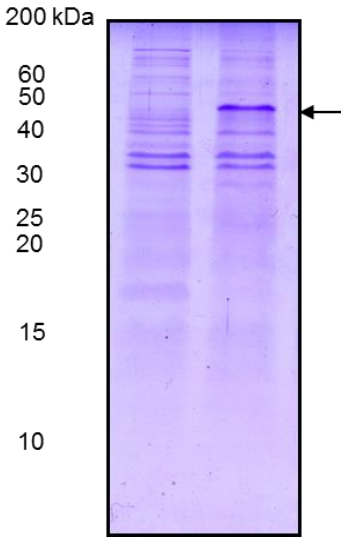

Figure IV-2.1.4| **Purification of recombinant HNRPDL IBs**. SDS-PAGE analysis of IBs purified from the insoluble fraction of induced cells grown at 37°C containing either an empty plasmid (lane 1) or a plasmid encoding for HRNPDL (lane 2). The arrow indicates the band corresponding to HNRDPL.

In addition, the IBs of cells expressing the yeast prion Ure2p and Aβ42 were purified using the same protocol and used as positive controls, since extensive characterization of the bacterial IBs formed by these two proteins have revealed that they possess an amyloid-like nature (Dasari et al., 2011; Espargaro et al., 2012b). Thioflavin-T (Th-T) fluorescence emission is enhanced in the presence of amyloid fibrils. Consistent with their amyloid properties, the same behaviour is observed upon incubation of Th-T with Aβ42 and Ure2p IBs. In the same way, the increase in Th-T fluorescence in the presence of HNRPDL IBs suggests the existence of amyloid conformations in the polypeptides embedded in these aggregates (Figure IV-2.1.5A). Although their impact in Th-T fluorescence is lower than that of Aβ42 IBs, it is quite similar to the one promoted by Ure2p IBs and remarkably different from that observed in the IBs fraction of negative control cells. The absorbance of the amyloid dye congo red (CR) redshifts in the presence of amyloid fibrils. The same effect was observed in the presence of Aβ42, Ure2p and HNRPDL IBs, consistent with the presence of amyloid-like structures in

these aggregates. The observed red-shift was smaller for HNRPDL than for the other two amyloid proteins, but still significantly different from that promoted by the IBs fraction of negative control cells (Figure IV-2.1.5B). Indeed, quantification of CR bound to IBs (see *Methods)* indicates that HNRPDL binds 2.4 times more dye than control IBs. The difference spectrum between the dye in the absence and presence of purified IBs allows the detection of the characteristic band at 540 nm, corresponding to the amyloid conformation in the three IBs (Figure IV-2.1.5C).

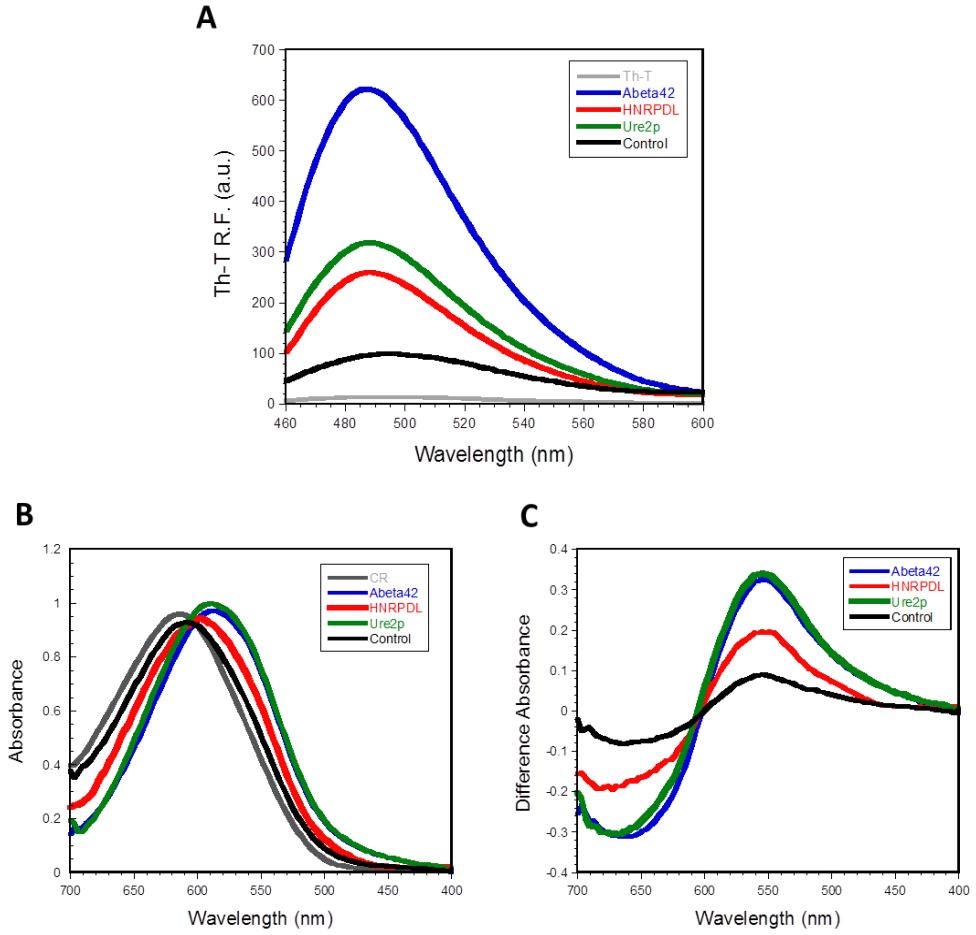

Figure IV-2.1.5| **Specific binding of amyloid dyes to HNRPDL IBs.** (A) Fluorescence emission spectra of Th-T in the absence or the presence of Aβ42, Ure2p, HNRPDL and control IBs. (B) Congo red (CR) absorbance spectra in the absence or the presence of Aβ42, Ure2p, HNRPDL and control IBs. (C) Difference absorbance spectra of CR in the presence and in the absence of IBs, showing the characteristic amyloid maximum at 540 nm.

HNRPDL IBs are enriched in intermolecular β-sheet structure

From a structural point of view, the formation of amyloid fibrils is always characterized by an enrichement in protein β-sheet content. Attenuated Total Reflectance–Fourier Transform Infrared spectroscopy (ATR-FTIR) is a powerful tool to investigate the secondary structure in protein aggregates (Shivu et al., 2013). In the present study, this

technique was used to analyse the conformational properties of the IBs (Figure IV-2.1.6; Table IV-2.1.2).

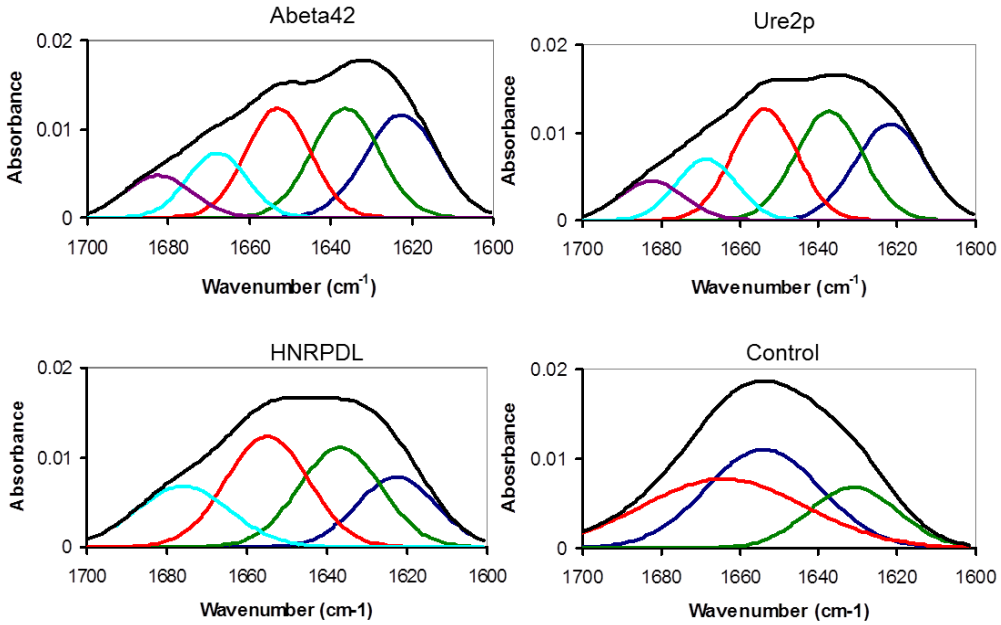

Figure IV-2.1.6| **Secondary structure content of HNRPDL IBs.** FTIR absorbance in the amide I region of the infrared spectrum (black) for Aβ42, Ure2p, HNRPDL and control IBs. Spectral components in the Fourier deconvoluted FTIR spectra are shown. The area and position of the correspondent bands are indicated in Table IV-2.1.2.

Deconvolution of the absorbance spectra in the amide I region allows to observe a signal at ≈1622 cm^{-1} common to the IBs formed by Aβ42, Ure2p and HNRPDL proteins, which is otherwise absent in negative control samples. This band is usually attributed to the presence of densely packed β-sheet structures, linked by short and strong hydrogen bonds, compatible with the intermolecular contacts in an amyloid fold (Natalello & Doglia, 2015). Aβ42, Ure2p and HNRPDL IBs also share a band at ≈1636 cm^{-1}, which has been typically assigned to intramolecular β-sheet; this band is also present in the negative control, but it contributes less to the total spectral area. In contrast, the negative control IBs exhibits higher contributions at ≈1653 cm−1 and ≈1665 cm^{-1}, which indicates an enrichment in helical, irregular and turn conformations, relative to Aβ42, Ure2p and HNRPDL IBs. Aβ42 and Ure2p IBs display a band at 1682 cm^{-1}, which is usually assigned to a high frequency β-sheet signal [66]. The lack of this signal, together with the presence of an exclusive band at ≈1676 cm^{-1}, attributed to turns (Goormaghtigh et al., 1990), suggests that despite sharing an amyloid nature, the fine structural properties of HNRPDL IBs differ from those formed by Aβ42 and Ure2p.

Table IV-2.1.2|**Contribution of secondary structure components to the absorbance FTIR spectra of Aβ42, Ure2p, HNRPDL and control IBs**

Aβ42		Ure2p		HRNPDL		Control		
Band (cm^{-1})	Area (%)	Band (cm^{-1})	Area (%)	Band (cm^{-1})	Area (%)	Band (cm^{-1})	Area (%)	
1622	26.20	1621	24.72	1622	19.40	-	-	Intermolecular β-sheet
1636	25.88	1637	26.70	1636	28.67	1630	19.32	Intramolecular β-sheet
1652	24.48	1653	25.89	1654	32.00	1653	40.44	α-helix/random
1668	13.48	1668	13.31	-	-	1667	40.22	Turns
-	-	-	-	1676	14.92	-	-	β-turns
1682	9.94	1682	9.38	-	-	-	-	Antiparallel β-sheet

HNRPDL IBs possess an inner amyloid core

Transmission Electronic Microscopy (TEM) was used to monitor the morphology of HNRPDL IBs. Freshly purified IBs displayed a typical electrodense amorphous appearance (Figure IV-2.1.7). However, upon incubation of purified IBs at 37°C for 12 h, the presence of fibrillar structures becomes already evident (Figure IV-2.1.7). The same behaviour has been reported for the amyloid-like IBs of other proteins and interpreted as the IBs containing densely packed bundles of amyloid fibrils inside cells that become relaxed and exposed upon *in vitro* incubation (Wang et al., 2008). This property can be qualitatively tested using proteinase K (PK), a protease usually used to map the protected core of amyloid fibrils because in spite of being highly active against peptidic bonds it cannot easily attack the highly packed backbones in amyloid β-sheet structures. Accordingly, as previously demonstrated, the PK digestion allows revealing the existence of a fibrillar core in Aβ peptide IBs (Morell et al, 2008). The same approach was used to assess if the presence of a similar fibrillar material might account for the amyloid conformational properties of HNRPDL IBs. Upon PK digestion, the presence of typical long and unbranched amyloid fibrils becomes evident. The fibrils are associated with apparently amorphous material and in some micrographs fibrils emerging from the preformed compact IBs are seen.

The elementary fibrils are ≈5 nm in diameter and tend to associate laterally into bundles, thus supporting that HNRPDL IBs constitute a bacterial reservoir of amyloid structures, that coexist with less ordered and PK susceptible protein regions, in good agreement with the deduced secondary structure content from FTIR analysis.

According to the presence of an amyloid core: (1) HNRPDL IBs are much more resistant towards PK digestion than negative control IBs (Figure IV-2.1.8A) and (2) HNRPDL IBs retain significantly higher Th-T binding in diluted solutions than negative control IBs even upon long time incubation (Figure IV-2.1.8B). These two properties recapitulate that of the amyloid-like IBs formed by Aβ40 and Aβ42 peptides in bacteria (Dasari et al., 2011).

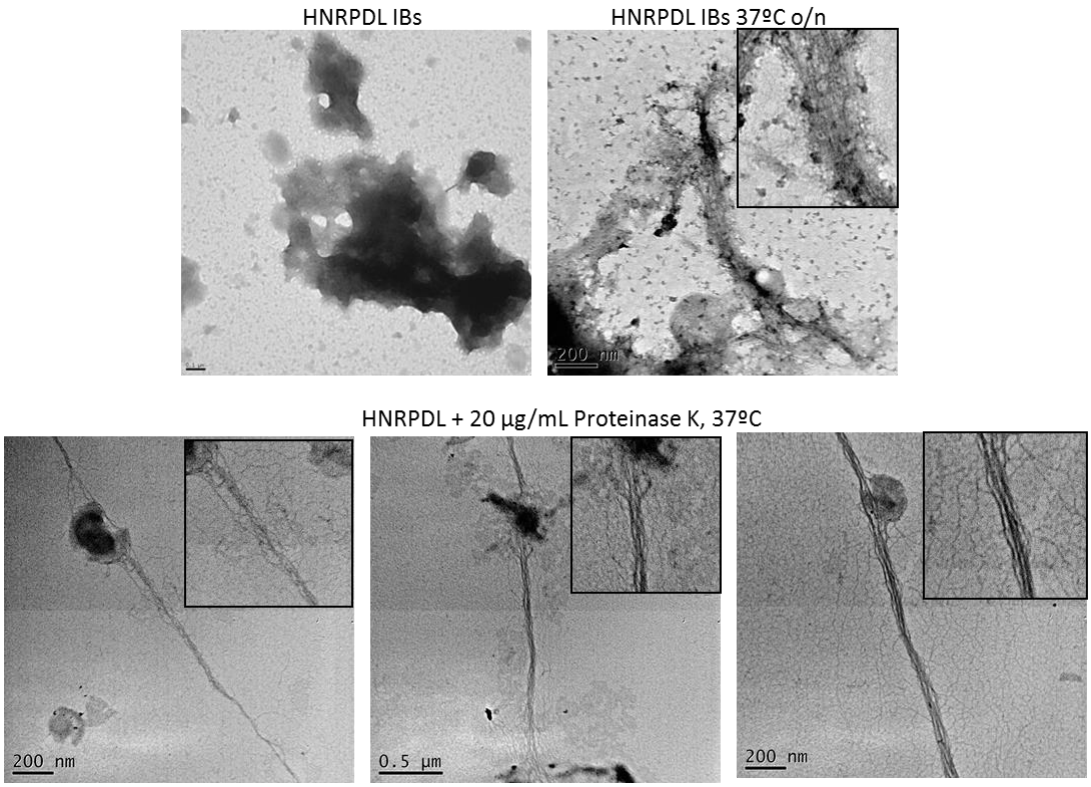

Figure IV-2.1.7| **HNRPDL IBs contain amyloid-like fibrils.** Negatively stained HNRPDL IBs visualized by TEM. The upper panel shows freshly purified HNRPDL IBs (left) and IBs incubated overnight at 37°C (right). The bottom panel displays representative micrographs of PK digested HNRPDL IBs.

HNRPDL IBs are toxic to cultured neuronal cells

It has been shown for different and unrelated proteins that the binding to ANS-like dyes correlates with the toxicity of amyloid species, suggesting that the exposure of hydrophobic patches is a critical characteristic of these pathogenic assemblies (Mannini et al., 2014). Thus, the binding of bis-ANS to Aβ42, Ure2p and HNRPDL IBs was analysed. In the presence of these aggregates, bis-ANS experienced the expected blue-shift and a strong increase in fluorescence maximum. The strongest spectral changes were promoted by the Aβ42, and the prion Ure2p IBs. However, HNRPDL IBs induced a significantly higher increase in bis-ANS fluorescence than negative control IBs (Figure IV-2.1.9). The aggregates formed by different human prion-like proteins have been shown to exert neurotoxicity (Blokhuis et al., 2013); therefore it was tested

if, in agreement with their bis-ANS binding ability, purified HNRPDL IBs could be toxic for cultured neuroblastoma SH-SY5Y cells.

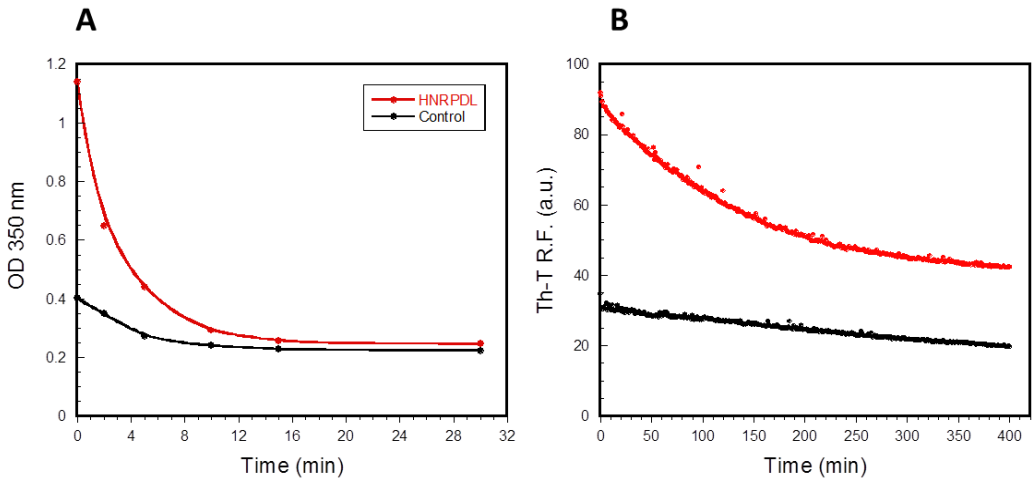

Figure IV-2.1.8| **Kinetic stability of HNRPDL and control IBs.** (A) Kinetics of HNRPDL and control IBs digestion by 20 µg/mL PK at 37 °C followed by the decrease in turbidity at OD_{350nm}. It is shown how the turbidity of control IBs decreases suddenly upon PK addition whereas in HNRPDL IBs the signal is lost more progressively, indicating higher resistance to proteolysis. (B) To monitor the Th-T fluorescence kinetics for HNRPDL and control IBs the initial stock solutions were diluted to yield a final OD_{350nm} of 0.05 in a 25 µM Th-T containing solution at 25 °C with continuous agitation. Control IBs bind little Th-T already at the beginning of the experiment and the signal decays linearly with time. In contrast the Th-T fluorescence of HNRPDL IBs shows a bi-exponential behaviour with measurable fluorescence still after 6 h, indicative of the presence of a persistent amyloid structure.

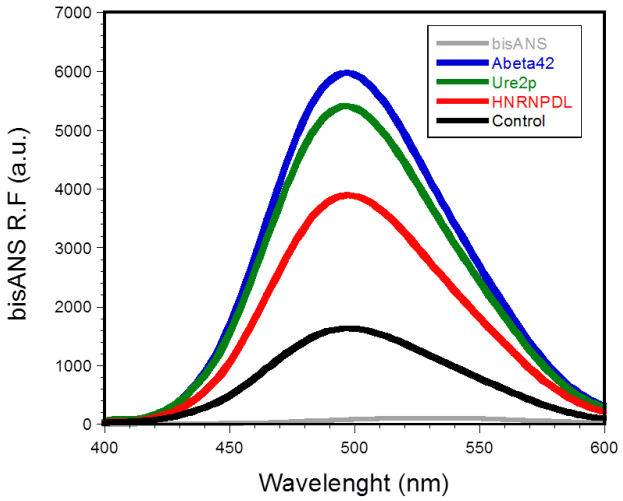

Figure IV-2.1.9| **Binding of bis-ANS to HNRPDL IBs.** Fluorescence spectra of bis-ANS in the absence and presence of Aβ42, Ure2p, HNRPDL and control IBs.

The combination of Hoechst and propidium iodide (PI) staining allows to asses cell viability by fluorescence microscopy, as viable cells are permeable to Hoechst and PI only enters cells with disintegrated membranes thus corresponding to dead cells. Cell morphology can be monitored as well to discriminate toxic and non-toxic aggregates in this assay. In samples treated with negative control IBs cell were attached to the culture plate at a confluent stage with only a reduced number of cells becoming stained with PI, indicating that they display low or no toxicity (Figure IV-2.1.10).

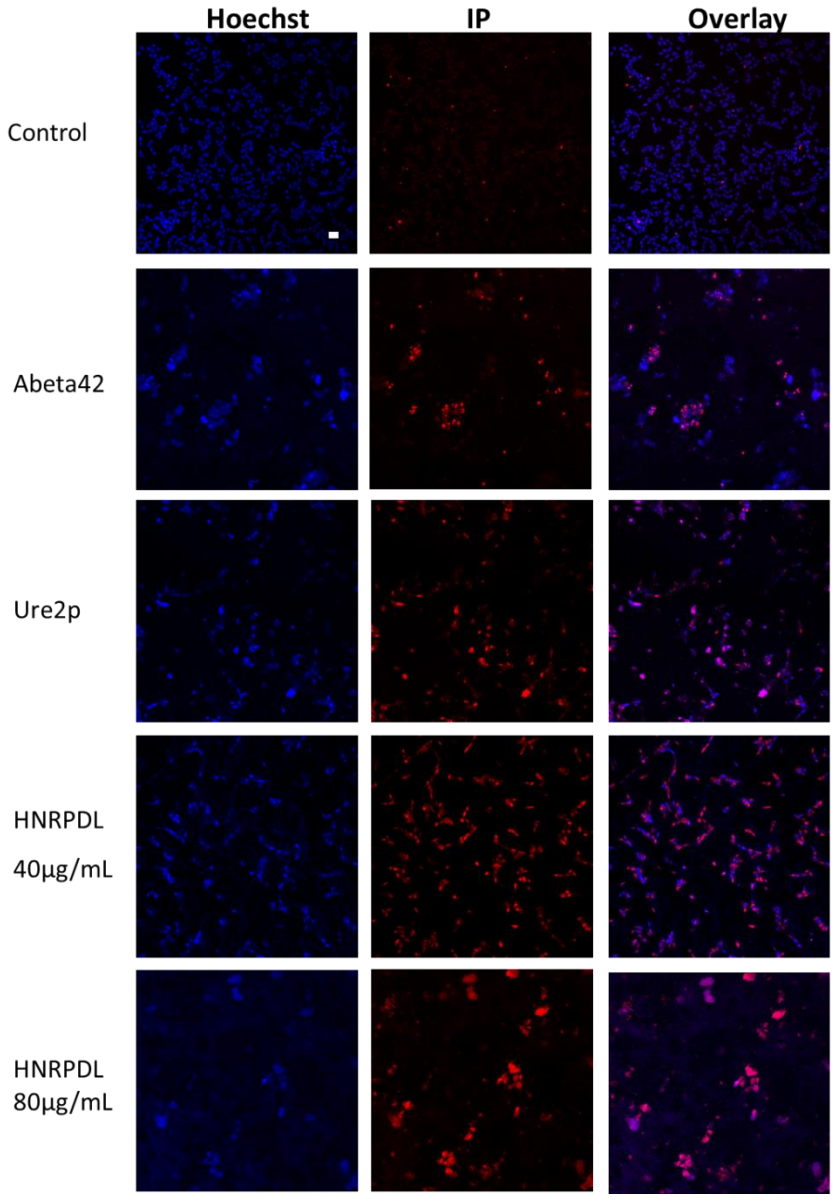

Figure IV-2.1.10| **Toxicity of HNRPDL IBs as visualized by confocal microscopy.** Representative confocal fluorescence microscopy images of SH-SY5Y cells stained with propidium iodide (IP) or Hoechst after incubation with Aβ42, Ure2p, HNRPDL and control IBs for 24 h at 37°C. The bar corresponds to 15 µM.

In contrast, the IBs formed by Aβ42 and Ure2p proteins were inherently toxic to neuronal cells as both induce positive PI staining in most cell nuclei. In the same manner, HNRPDL IBs turned to be highly neurotoxic, with a large majority of cells being stained by PI. Moreover, this effect was dose dependent, since cells incubated with 40 μg/mL of HNRPDL IBs kept attached, homogeneously distributed and displayed normal morphology, whereas cells treated with 80 μg/mL HNRPDL IBs lost completely their morphology becoming detached and agglutinated (Figure IV-2.1.10).

Discussion

The number of human proteins involved in neurodegenerative disorders is rapidly expanding, suggesting that there are likely numerous disease-associated proteins yet to be identified. Many of these disorders involve the formation of self-templating aggregates (Knowles et al., 2014). However, since most protein aggregates are not infectious, prion-based disorders have been always considered different from the rest of aggregation caused diseases. Nevertheless, increasing evidence indicates that the proteins involved in many neurodegenerative disorders, including Alzheimer's and Parkinson's, display a prion-like behaviour, exhibiting a cell-to-cell propagation (Peggion et al., 2014). In addition, different human proteins containing intrinsically disordered domains with an amino acid composition resembling those of the prion forming domains (PFDs) in yeast prions are being found connected to degenerative disorders (Falsone, A. & Falsone, S.F., 2015). Many of these disorder-linked PrLD-containing proteins are RNA-binding proteins typically containing one or more RRM domains (Pinkus et al., 2014). TDP-43 was the first identified protein of this class. It was initially found to be a major constituent of the protein aggregates in the spinal cord motor neurons, in the hippocampus and neocortex of ALS or FTLD patients, but it is also present in an aggregated form in other neurodegenerative disorders (Da Cruz & Cleveland, 2011). A majority of the mutations linked to ALS or FTLD map into the PrLD, implicating thus this domain in the disease (Falsone, A. & Falsone, S.F., 2015). HNRPDL is a less studied RNA-binding protein, which shares domain organization with TDP-43 (Figure IV-2.1.1), despite its precise three-dimensional structure is unknown. Interestingly, it has been shown that two mutations occurring in the PrLD of this protein, D378N and D378H, lead to limb-girdle muscular dystrophy 1G (Vieira et al., 2014). According to PrionScan, PLAAC and PAPA prion predictors these two mutations increase the prion propensity of the domain (Table IV-2.1.3). The structure of TDP-43 inclusions in ALS and FTLD patients is still unclear and whether these deposits have an amyloid nature or not is matter of debate. Due to the difficulty of purifying TDP-43 for the in vitro characterization of its aggregation process and because the intracellular aggregation of human amyloid proteins in bacteria has been shown to result into amyloid-like IBs, Chiti and co-workers characterized the nature of the IBs formed by TDP-43 in *E. coli* to approximate the conformational properties of its inclusions in ALS and FTLD (Capitini et al., 2014). They found out that TDP-43 aggregates present in E.

coli IBs did not possess any of the hallmarks of amyloid fibrils, allowing them to be classified as amorphous. However, they were shown to be toxic for cultured neuronal cells. This raises the question of whether this conformation is a generic property of the aggregates formed by human PrLD-containing proteins and whether it is indeed the lack of an ordered structure in the aggregates the underlying cause of their toxicity in pathological states.

Table IV-2.1.3|**Predicted prion propensity of wild type HNRPDL and mutants involved in limb-girdle muscular dystrophy 1G**

	PrionScan	PAPA	PLAAC
HNRPDL	42.904	0.12	30.301
D378H	44.280	0.14	31.395
D478N	46.922	0.15	33.013

The analysis of the conformational aggregates formed by HNRPDL in bacteria indicate that this is not the case, since these aggregates bind to amyloid dyes, are enriched in intermolecular β-sheet conformation and contain inner fibril-like structure; still they are neurotoxic. The amorphous nature of the aggregates formed by TDP-43 contrasts with those formed by the yeast PFDs to which its PrLD resembles, since these latter display clear amyloid properties, both *in vitro* (Sabate et al., 2011) and when expressed recombinantly in bacteria (Espargaro et al., 2012b). It has been shown that the presence of a short amyloidogenic stretch in PrLDs, as predicted with our algorithm pWALTZ, determines to a large extent its amyloid potential (Sabate et al., 2015). Q/N enriched yeast putative prion domains with pWALTZ scores higher than 73.55 all formed amyloid assemblies, whereas those falling below this threshold display lower amyloid propensity. Interestingly, the PrLDs of HNRPLD and TDP-43 display pWALTZ values above and below this threshold, respectively, which might account, at least in part, for their different intracellular amyloid propensity. Aggregation constraints the evolution of proteins and accordingly nature have evolved different strategies to minimize protein aggregation in sequences and structures (Navarro et al., 2014). In this context, the inherent aggregation of human proteins containing PrLDs and their link to disease, strongly suggest that these domain are conserved because they serve functional purposes. Increasing evidence indicates that in RNA-binding proteins, these disordered domains work in the reversible recruitment of the protein into RNA-P bodies or stress granules under cellular stress (Malinovska et al., 2013; Li et al., 2013). The amyloidogenic properties of HNRPLD constitute yet another example illustrating how the determinants for the establishment of functional interactions and those accounting for the formation of toxic amyloid assemblies overlap significantly (Castillo

& Ventura, 2009; Sabate et al., 2012a), suggesting that in PrLDs-containing proteins the formation of functional macromolecular complexes and the aggregation of their individual subunits might compete in the cell. This will explain, why point mutations in these domains or environmental changes, such as prolonged stress, enhance recruitment into stress granules (Kim et al., 2013; Wolozin, 2012), disrupting the reversibility of the assembly and finally leading to the accumulation of aggregates, triggering the onset of the disease. The present work illustrates the potency of microbial cell factories to model amyloid conformational conversion.

Methods

Protein expression and purification

Human HNRPDL cDNA was cloned into a pET28a (+) vector (Novagen, INC., Madison, WI, USA). The plasmids encoding for Aβ42 and Ure2p proteins were as previously described (Dasari et al., 2011; Espargaro et al., 2012b; Sabate et al., 2012b). The plasmids were transformed into E. coli BL21 (DE3) cells. Cells were grown aerobically in liquid Luria–Bertani (LB) medium containing appropriate antibiotics in a rotary shaker at 37°C and 250 rpm. Overnight cultures were diluted 100-fold in LB and allowed to grow to an OD 600 of 0.6. At the indicated OD600, protein expression was induced with 1 mM isopropyl β-d-1-thiogalactopyranoside (IPTG) and in the case of Aβ42 and Ure2p the culture was continued at 37°C for 4 h as previously described (Dasari et al., 2011; Sabate et al., 2012b). HNRPDL cells were cultured at 37°C 25°C or 18°C for 20h upon induction. To express HNRPDL-GST, the human HNRPDL sequence was cloned into a pETM-30 vector in order to produce a N-terminal fusion protein with a His tag followed by GST with a TEV protease cleavage site; the resulting construct was transformed into E. coliBL21(DE3) cells and grown as described above, inducing protein expression for 20h at 20°C or 16°C. As a negative control, E. coli BL21 (DE3) cells were transformed with an empty pET28a(+) vector, grown and induced in the same conditions than cells containing the HNRPDL encoding plasmid.

Inclusion bodies purification

Intracellular IBs were purified as previously described (Morell et al., 2008). Briefly, cell pellets from 5 mL induced cultures were resuspended in 140 µL of lysis buffer (10 mM Tris–HCl, pH 8.0, 1 mM EDTA, 150 mM NaCl), containing 0.8 µL protease inhibitor PMSF (17.4 mg/mL) and 3 µL lysozyme (10 mg/mL). The suspension was incubated for 30 min at 37°C under gentle agitation. Then cells were incubated with 1% (v/v) NP-40 for 50 min under mild agitation at 4°C. To remove nucleic acids, 3 µL of DNase I from a 1 mg/mL stock, 3 µL of 1 mg/mL of RNase and 3 µL of 1 M $MgSO_4$ were added and the resulting mixtures were further incubated at 37°C for 30 min. IBs were collected by centrifugation at 12,000×gfor 15 min at 4°C. Finally, IBs were washed with lysis buffer containing 0.5% Triton X-100 three times, twice with lysis buffer and finally stored at -

80°C until analysis. The purified IBs fraction was resolved on a 15% SDS–PAGE gel stained with Coomassie brilliant blue.

Thioflavin-S binding in living cells

Detection of cell-permeable thioflavin-S (Th-S) binding was performed in non-induced and induced living cells expressing HNRPDL protein. Bacterial cells were washed with PBS and diluted to an OD_{600nm} of 0.1. Cells were incubated for 1 h in the presence of 125 µM Th-S diluted in PBS and washed twice with PBS. Fluorescence emission spectra were recorded in a range of 400–500 nm using an excitation wavelength of 375 nm. Apertures of 5 nm were fixed in both excitation and emission slits. The analysis of fluorescence microscope images allowed the detection of accumulated amyloid deposits inside bacterial cells. Cells were placed on top of a microscope slide and covered with a cover slip. Photographs were acquired using a 488-nm argon laser and emission collected in a 515–540 nm range.

Thioflavin-T binding

Thioflavin-T (Th-T) binding was analyzed for IBs purified from cells expressing Aβ42, Ure2p or HNRPDL and from control cells, resuspended in PBS at pH 7.0 and OD_{350nm} of 0.1 in the presence of 25 µM Th-T. Fluorescence emission spectra were recorded from 460 to 600 nm with an excitation wavelength of 440 nm, using a slit width of 5 nm for excitation and emission in a Jasco FP-8200 spectrophotometer (Jasco corporation, Japan). Each trace represents the average of 3 accumulated spectra. Th-T fluorescence kinetics for HNRPDL and negative control IBs were analyzed from diluted IBs at a final OD_{350nm} of 0.05 in PBS at pH 7. Samples were incubated for 400 min under agitation (800 rpm) at 25°C, in the presence of 25 µM Th-T. The kinetic traces were measured exciting at 440 nm and emission was recorded at 475 nm, slit width of 5 nm were used for excitation and emission in a Jasco FP8200 spectrophotometer (Jasco corporation, Japan).

Congo red binding

Congo red (CR) interaction with IBs purified from cells expressing Aβ42, Ure2p or HNRPDL and from control cells was tested using a Cary-400 UV/Vis spectrophotometer (Varian Inc.). IBs samples were diluted to a final OD350nm of 0.1 in PBS at pH 7.0 and 20 µM of CR was added. After 5 min of equilibration, the absorbance spectra were recorded from 400 to 700 nm. The differential CR spectra in the presence and absence of protein were calculated to detect the typical amyloid band at ≈540 nm. CR binding was quantified by the equation: CR Bound $=Abs_{540nm}/25{,}295 - Abs_{477nm}/46{,}306$.

Bis-ANS binding

Binding of 4,4'-bis[1-anilinonaphthalene 8-sulfonate](bis-ANS) to purified Aβ42, Ure2p, HNRPDL IBs and the negative control extract was evaluated by registering bisANS

fluorescence between 400 and 600 nm after excitation at 370 nm in a Jasco FP-8200 spectrophotometer (Jasco corporation, Japan), with excitation and emission slit widths of 5 nm. 25 µM of bis-ANS was added to IBs at a final OD_{350} of 0.1 in PBS. Spectra were registered at 25°C as the accumulation of three consecutive scans, after equilibration of the sample for 5 min.

ATR FTIR spectroscopy

ATR FTIR spectroscopy analyses of purified Aβ42, Ure2p, HNRPDL and control IBs were performed with a Bruker Tensor 27 FTIR Spectrometer (Bruker Optics Inc.) with a Golden Gate MKII ATR accessory. Spectrum acquisitions consisted of 16 independent scans, measured at a resolution of 2 cm^{-1}within the 1,800–1,500 cm^{-1} range. Spectra were acquired, background subtracted, baseline corrected 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. All FTIR spectra were fitted to overlapping Gaussian curves using PeakFit package software (Systat Software) and the maximum and the area of each Gaussian were calculated.

Limited proteinase K digestion

HNRPDL and negative control IBs were resuspended at a final OD_{350} of 1 in PBS buffer at pH 7.0. Digestion was initiated by adding proteinase K (PK) at a final concentration of 20 µg/mL and the reaction was carried out for 30 min at 37°C under agitation (500 rpm). PK proteolysis was monitored at 350 nm using a Cary-400 UV/Vis spectrophotometer.

Transmission electron microscopy (TEM)

Purified HNRPDL IBs (100 µg/mL) were digested with 20 µg/mL proteinase K (PK) and incubated at 37°C at different digestion times. Proteolytic mixtures were centrifuged and pellets were resuspended in water. Then 10 µL of purified and PK digested HNRPDL IBs solutions were placed on carbon-coated copper grids and allowed to stand for 5 min. For negative staining, grids were washed with distilled water and stained with 2% (w/v) uranyl acetate for 1 min. The samples were imaged using a JEM-1400 transmission electron microscope operating at an accelerating voltage of 120 kV.

Cell viability assay

Human SH-SY5Y cells were cultured in F-12 medium supplemented with 10% FBS on glass slides at 70% confluence and maintained at 37°C in a 5% CO_2 atmosphere. Cell cultures were incubated in the absence (control) and the presence of Aβ42, Ure2p and HNRPDL IBs resuspended in sterile PBS for 24 h. Cells were counterstained with 0.5 µg/mL Hoechst and 10 µg/mL PI (Molecular Probes) for 15 min at 37°C and washed twice with PBS buffer. Cell morphology and viability were analyzed by confocal

fluorescence microscopy (Olympus Fluoview 1000) with an UPlansApo 10x objective using an orange diode (588–715 nm emission collected) and a UV laser (excited at 350 nm and collected at 405 nm).

2.2 Trifluoroethanol Modulates Amyloid Formation by the All α-helical URN1 FF Domain

In the present work, it was investigated the impact of TFE-induced structural rearrangements on the process of amyloid formation by the all α-helical FF protein.

Conformational Properties of URN1-FF at Low pH in the Presence of TFE

At pH 2.5 the URN1-FF forms a Molten Globule (MG) state that retains most of the α-helical secondary structure content characteristic of the protein native state (Castillo et al., 2013). Accordingly, in these conditions, the far-UV CD spectrum indicates that most of the polypeptide chain adopts an α-helical conformation and the spectrum is similar to that of the native protein at pH 5.7 (Figure IV-2.2.1A). The URN1-FF contains two buried Trp residues at positions 27 and 56. The intrinsic fluorescence spectrum of the acid induced URN1-FF MG displays higher intensity and is slightly red-shifted compared with that of the native state suggesting an opening of the globular structure with a concomitant exposure of hydrophobic clusters to the solvent at low pH (Figure IV-2.2.1B). This extent is confirmed by the observed increase in the fluorescence emission of the bis-ANS dye and the blue-shift of its maximum in the presence of the protein at low pH, when compared with the spectrum of bis-ANS in the presence of the native protein (Figure IV-2.2.1C). Despite the fact that the acid induced URN1-FF MG does not exhibit any evidence of the population of β-sheet conformations, this state is aggregation competent (Castillo et al., 2013), which contrasts with the native state, where the protein remains soluble for years. Two main reasons might account for the different solubility of the two states: (i) opening of the tertiary structure and associated exposition of previously hidden non-polar residues would promote the establishment of hydrophobic intermolecular contacts leading to the formation of aggregation nuclei; (ii) fluctuations of the helical secondary structure would result in the transient formation of disordered protein regions able to establish intermolecular backbone hydrogen bonds, leading to formation of β-sheet enriched oligomers. The first mechanism has been proposed to be involved in the aggregation of proteins belonging to different structural classes, whereas the second one has been proposed to apply for proteins enriched in α-helical structures such us myoglobin (Fandrich et al., 2003). To decipher the contribution of these two effects to the aggregation of URN1-FF the conformational and aggregational properties of the protein were analysed in the presence of 15% and 25% TFE (v/v). The far-UV CD of URN1-FF at pH 2.5 in the presence of both 15% and 25% TFE (v/v) correspond to that of an all α-helical protein (Figure IV-2.2.1A). Accordingly, the analysis of the CD spectra with the K2D3 algorithm predicts a secondary structure content consisting of ≈95% α-helix, without any significant β-sheet component. Both spectra display higher ellipticity at 210 nm and 222 nm than that of the MG state at the same pH, although only 25% TFE (v/v) promotes higher ellipticity than that in the native state. At pH 2.5, the presence of TFE

induces an increase and a red shift in the intrinsic Trp fluorescence of URN1-FF (Figure IV-2.2.1B), indicating that the presence of the cosolvent likely promotes additional opening of the structure. However, the spectra are still significantly different from that of denatured URN1-FF, where a strongest shift and reduction in the fluorescence emission is observed, suggesting that in the presence of TFE the protein presents at least a residual compact structure.

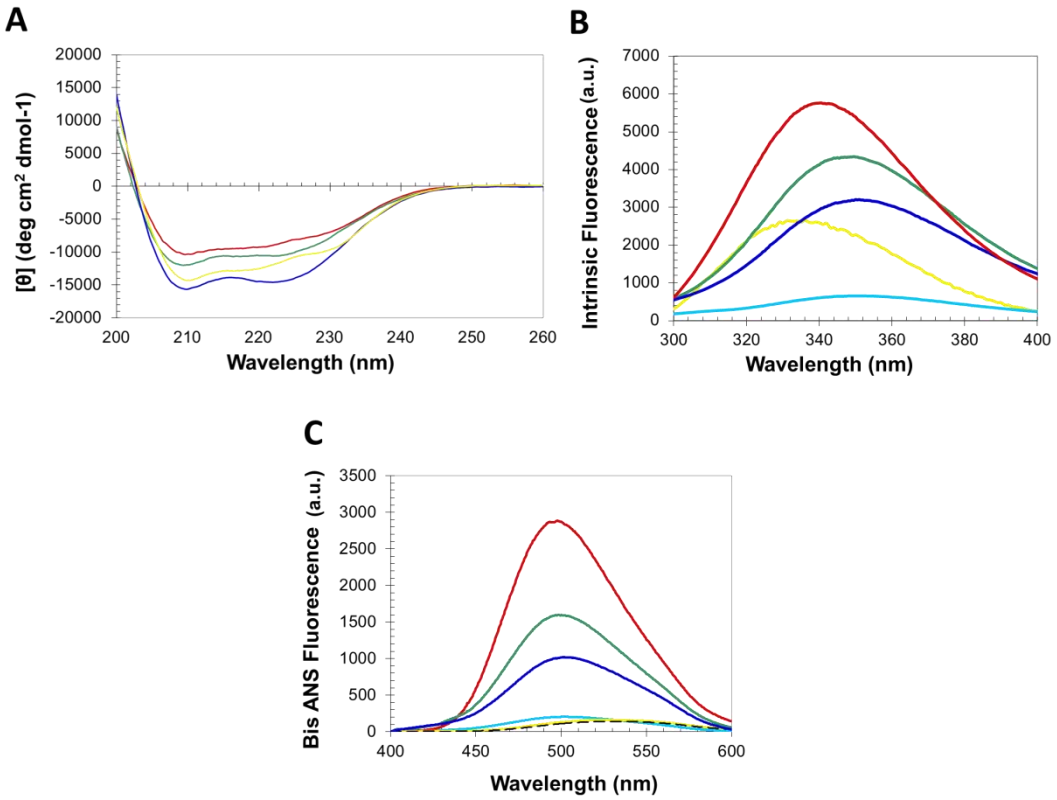

Figure IV-2.2.1| **2,2,2-trifluoroethanol (TFE) dependence of URN1-FF molten globule (MG) conformational properties.** Protein samples were prepared at 20 µM and pH 2.5 and immediately measured by (A) far-UV CD, (B) tryptophan intrinsic fluorescence and (C) 4,4'-Dianilino-1,1'-Binaphthyl-5,5'-Disulfonic Acid (bis-ANS) fluorescence at 25 °C. The fluorescence emission spectrum of bis-ANS in the absence of protein is represented as a dotted line. URN1-FF species were measured at0% (red), 15% (green) and 25% (blue) of TFE (v/v). The spectra of the native protein at pH 5.7 are shown in yellow. In (B) and (C) the spectrum of heat denatured URN1-FF at 90 °C is shown in light blue.

In agreement with the Trp spectral data, the presence of TFE induces a reduction in bis-ANS fluorescence (Figure IV-2.2.1C), suggesting the disruption of certain hydrophobic interactions still present in the MG state. However, this reduction in bis-ANS binding is moderate if we compare it with that of the denatured state where the bis-ANS fluorescence decreases dramatically and red shifts. Overall, the data suggest that the presence of TFE promotes a conformation with a well-defined α-helical

secondary structure, as observed for other amyloidogenic peptides (Brender et al., 2011), but only rudimentary tertiary contacts and highly accessible hydrophobic side-chains. These properties would allow us to study if it is the presence of solvent exposed non-polar residues or alternatively the weakening of the secondary structure that is mainly responsible for the aggregation of the URN1-FF domain when incubated at acidic pH.

Effect of TFE on Amyloid Fibril Formation by the URN1-FF Domain

The MG formed by URN1-FF at pH 2.5 is a stable species that remains apparently soluble for weeks at 20 µM concentration. However, it self-assembles into amyloid fibrils upon incubation at higher concentrations. The FF domain was incubated at 100 µM for 7 days at acid pH in the absence or the presence of 15% and 25% TFE (v/v) and analysed the presence of amyloidogenic aggregates by monitoring the binding to Thioflavin-T (Th-T) of these protein solutions (Figure IV-2.2.2A). Despite all the samples promoting an increase in Th-T fluorescence emission indicative of the presence of amyloid-like protein aggregates, the presence of 15% and 25% TFE (v/v) reduced Th-T emission by 2.5 and 6 fold, respectively. The changes induced in Congo Red absorbance spectrum were monitored to explore the amyloid properties of the incubated protein solutions. All the samples promoted the characteristic amyloid-induced red-shift and an increase in absorbance in the dye spectrum (Figure IV-2.2.2B). However, the binding to CR decreases in the presence of 25% (v/v) TFE. The presence and morphology of protein aggregates was further analysed using Transmission Electron Microscopy (TEM) (Figure IV-2.2.2C). Both in the absence and presence of 15% TFE (v/v) all the detected aggregates correspond to long fibrillar species. The fibrils in both solutions were long and unbranched and consist of linear thin and thick fibrils. Linear thin fibrils displayed a diameter of ≈7.0 nm consistent with that of the amyloids formed by disease-linked polypeptides, which diameters in the 4–10 nm range (Merz et al., 1983). Thick fibrils had a diameter of ≈14 nm, which suggest that they likely result from the association of two thin fibrils and in fact the presence of these individual fibrils could be observed in some images. In the absence of TFE thin fibrils dominate the solution, whereas in the presence of 15%TFE thick fibrils become more abundant species. The presence of 25% (v/v) almost complete abrogated the presence of long fibrils and short protofibrilar assemblies become the main aggregated species. The aggregation of URN1-FF at acidic pH follows this kinetic scheme but exhibits a very short lag phase of only few minutes (Figure IV-2.2.3) suggesting that the formation of amyloid-like intermolecular interactions occurs rapidly in the aggregation process. The presence of 15% TFE (v/v) has a dramatic effect on the kinetics of URN1-FF amyloid formation extending the lag phase up to 15 h. Moreover, no evident aggregation could be observed during the time of the experiment in the presence of 25% TFE. Overall, the presence of TFE at moderate concentrations has an inhibitory effect on the process of amyloid fibril formation by the URN1-FF domain at low pH.

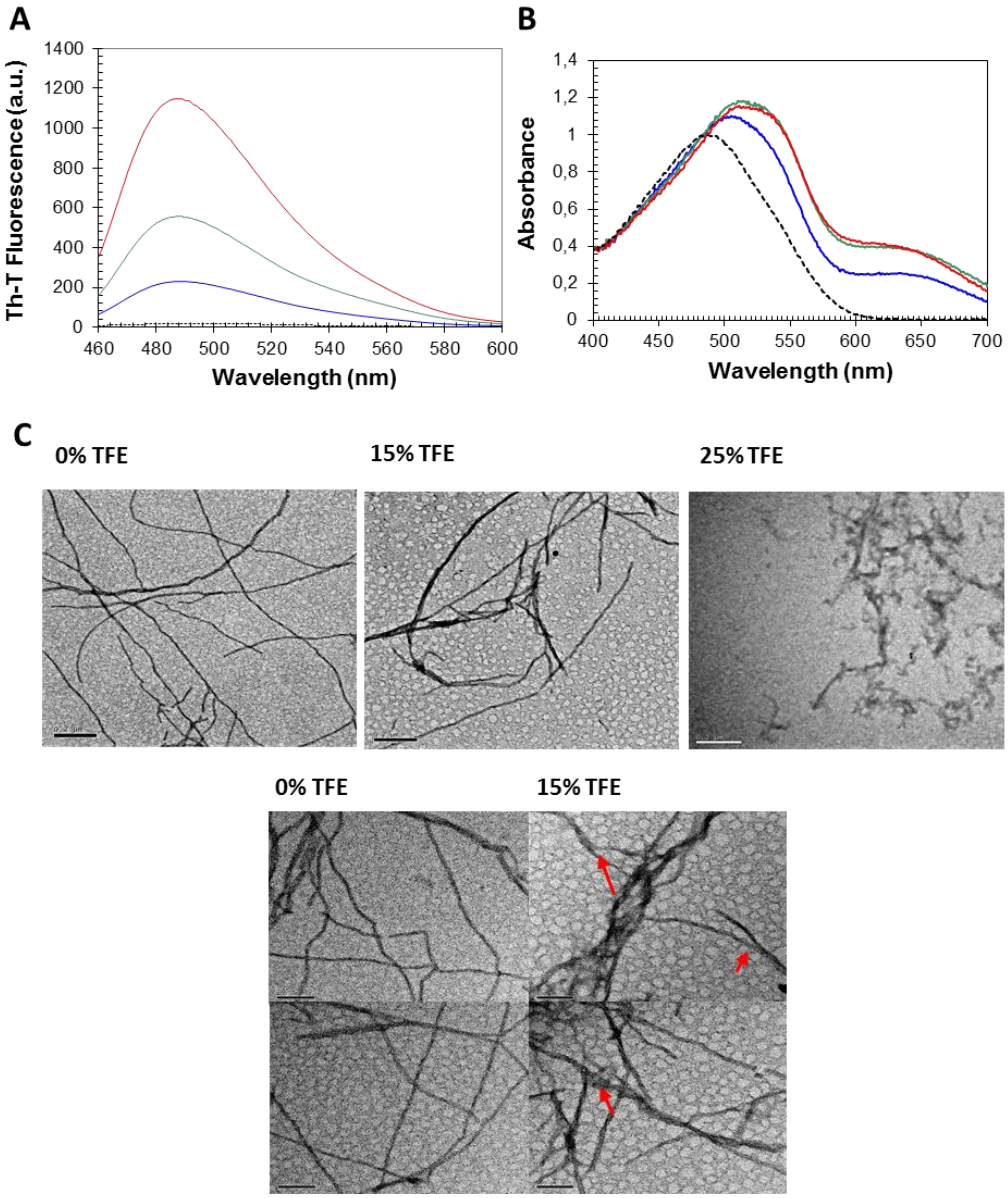

Figure IV-2.2.2| **Amyloid-like properties of URN1-FF aggregates formed at different TFE concentrations.** The protein domain was incubated at pH 2.5 and 100 µM for one week. (A) Fluorescence emission spectra of Th-T (25 µM) in the absence (dotted line) and in the presence of 10 µM of protein aggregates. (B) Absorption spectra of CR (20 µM) in the absence (dotted line) and in the presence of 12 µM of URN1-FF aggregates, in (A) and (B) protein aggregates formed at 0%, 15% and 25% TFE (v/v) concentrations are represented in red, green and blue, respectively. (C) Representative TEM images of URN1-FF solutions incubated at the indicated TFE concentrations. The bar corresponds to 200 nm and 100 nm in upper and lower panels, respectively. The arrows indicate the assembly of thin fibrils to form ticker fibrillar structures.

To monitor if the presence of TFE affects the kinetics of URN1-FF amyloid fibril formation, the increase in Th-T fluorescence emission was followed during 2000 min (Figure IV-2.2.3). The kinetics of amyloid fibril formation can be usually adjusted to a sigmoidal curve, reflecting the existence of a nucleation-dependent growth reaction.

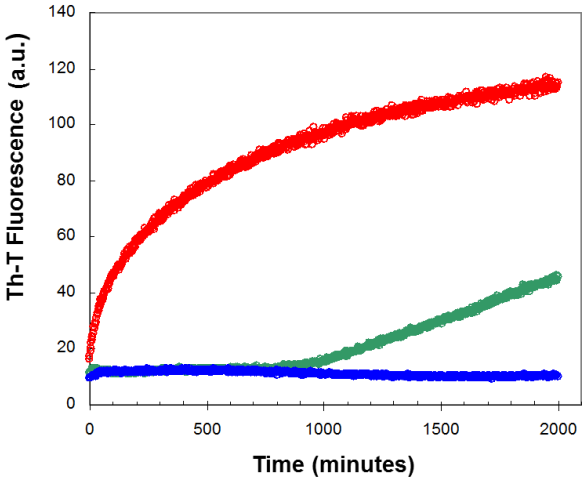

Figure IV-2.2.3| **TFE dependence of URN1-FF aggregation kinetics.** Change in Thioflavin-T (ThT) fluorescence (25 µM) during the aggregation of URN1-FF at 100 µM. Reactions at 0%, 15% and 25% TFE (v/v) concentrations are represented in red, green and blue, respectively.

Seeding Properties of URN1-FF Domain Aggregates

The conformational properties of the amyloid structures formed in the absence and presence of 15% of TFE were monitored in order to analyse if, as with the soluble protein, they exhibit differences in these two conditions. To this aim protein solutions incubated for at least 21 days were used, since quantification of the amount of aggregated protein by sample fractionation using sedimentation at 100,000g for 1 h indicated that >95% of the total protein was located in the insoluble fraction both in the absence and presence of the cosolvent. The fibrils formed in the absence and presence of 15% TFE (v/v) will be referred to as f0% and f15% fibrils, respectively. The far-UV CD spectra of these protein solutions indicate that in both cases a transition from the original helical structure towards a β-sheet enriched conformation has occurred, since both spectra present a characteristic minimum at ≈217 nm, without any remaining α-helical component (Figure IV-2.2.4A). It was also monitored the Trp intrinsic fluorescence emission in both fibrilar states. In contrast to what happens in the soluble states, the f15% fibrils display lower fluorescence emission, suggesting that likely their hydrophobic residues, including Trp, are more buried (Figure IV-2.2.4B). This is also consistent with their lower binding to bis-ANS, relative to the f0% fibrils (Figure IV-2.2.4C). Although speculative, it is tempting to propose that these conformational differences might be related to the different persistence of thick fibrils in the two solutions, in such a way that the association of two thin fibrils might occur through hydrophobic residues that become therefore protected from the solvent.

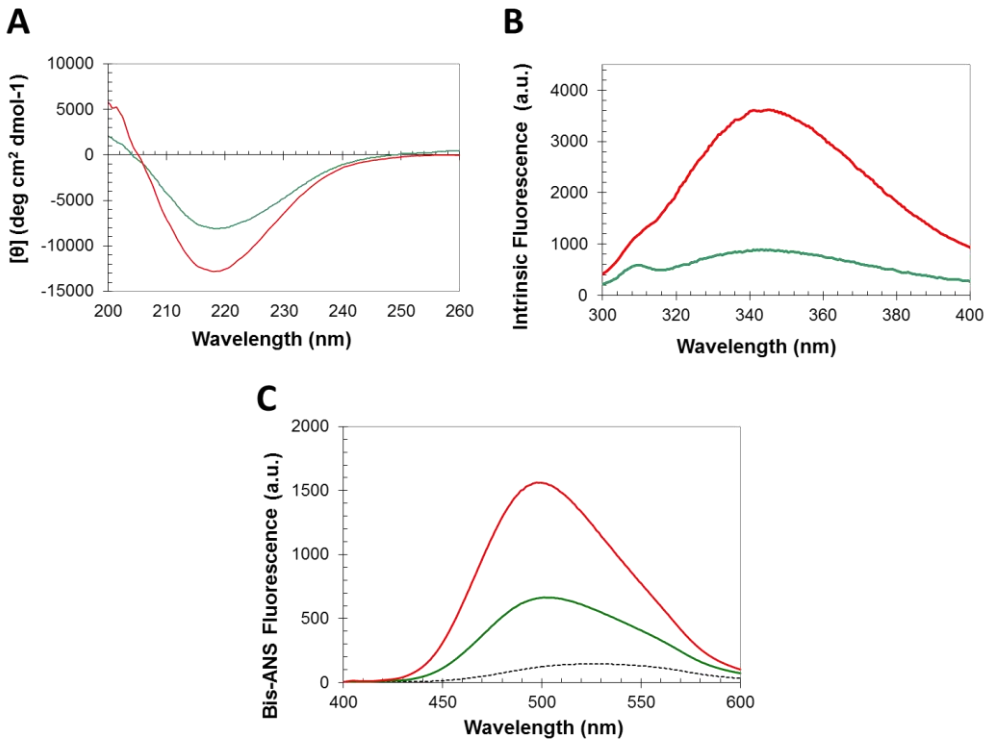

Figure IV-2.2.4|**TFE dependence of the conformational properties of URN1-FF fibrils.** The protein domain was incubated at pH 2.5 and 100 µM for at least 21 days. (A) Far-UV CD spectra using a final concentration of 20 µM, (B) Tryptophan intrinsic fluorescence, (C) Fluorescence emission spectra of bis-ANS (25 µM) collected in the absence (dotted line) and in the presence of fibrils (10 µM).In all the cases, f0% fibrils in red and f15% fibrils in green.

The nucleation step of the amyloid assembly is shortened in the presence of preformed amyloid fibrils of the same protein that can act as seeds for the polymerization reaction. This reaction is not only sensible to the protein sequence but also to the protein conformation and it has been shown for different models that fibrils of the same protein formed under different conditions might have differential seeding properties. In contrast, in specific cases, proteins of dissimilar sequences might cross-seed (Ramamoorthy, 2013). So, the f0% and f15% URN1-FF fibrils were tested to address if the amyloid aggregation reaction of this domain at low pH and aqueous solution can in fact be seeded with fibrils preformed in the same conditions. The addition of 10% sonicated preformed fibrils at the beginning of the polymerization reaction results in a significant acceleration of the reaction at early stages indicating that they exert the nucleating effect characteristic of amyloids (Figure IV-2.2.5). Next, it was tested if preformed f15% fibrils were able to seed the reaction when added to an URN1-FF solution devoid of the cosolvent. As shown in Figure IV-2.2.5A, despite the accelerating effect is lower than that of the fibrils formed in the absence it is still very significant. Thus, despite the different spectroscopic features of the two types of aggregates it is likely that they share some common molecular details that allow them

to cross-seed. Next, it was addressed if f0% and f15% URN1-FF preformed fibrils would be able to short the long lag phase of the aggregation reaction when it occurs in the presence of the 15% TFE (v/v) (Figure IV-2.2.5B). The f15% fibrils efficiently seeded the aggregation of the soluble protein in the presence of the cosolvent. Surprisingly, the f0% fibrils did not exert any accelerating effect on the aggregation reaction in this condition indicating that presence of TFE not only acts by modulating the formation of the initial nucleus but also induces structural diversity in URN1-FF fibrils.

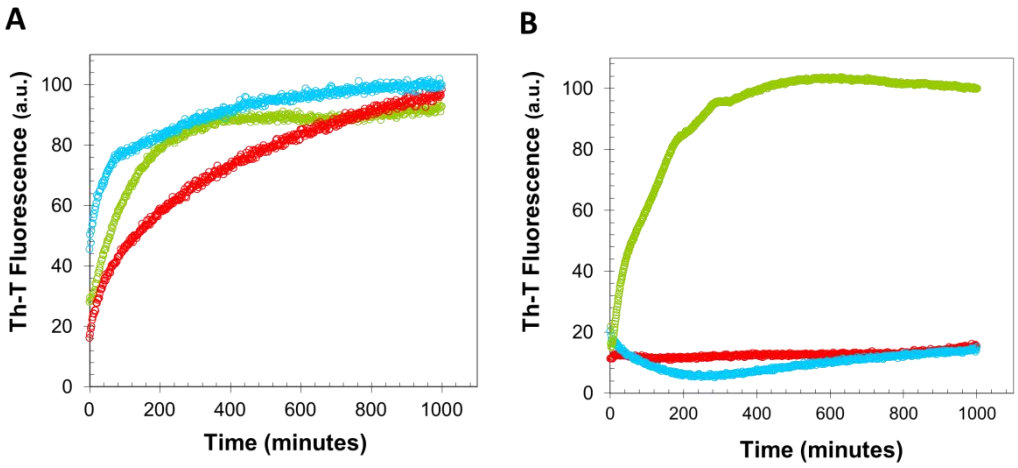

Figure IV-2.2.5|**TFE dependence of the self-propagation of URN1-FF amyloid fibrils originating from various concentrations of TFE.** Change in ThT fluorescence (25 µM) during the aggregation of URN1-FF at 100 µM. 10% of preformed fibrils were used for seeding and cross-seeding assays. (A) Aggregation reactions in the absence of TFE and (B) the presence of 15% TFE (v/v). Unseeded reactions are shown in red, reactions seeded with f0% and f15% fibrils are shown in blue and green, respectively.

Discussion

Biophysical characterization of protein aggregation usually exploits co-solvents to trigger changes in the protein conformation or environment that allows the study of both the conformers involved in the self-assembly reaction and its kinetics in a controlled manner (Seeliger et al., 2013; Grudzielanek et al., 2005). The fluorinated alcohol 2,2,2-trifluoroethanol (TFE) is perhaps the most used co-solvent to investigate protein aggregation (Otzen, 2010b). TFE promotes a gradual coil-to-helix transition in peptides and disordered proteins (Luo & Baldwin, 1997; Myers et al., 1998) which attain maximal ellipticity at ~30%–40% v/v TFE. At the same concentrations, TFE usually results in globular proteins denaturation, inducing the formation of non-native α-helices. The presence of TFE at lower concentrations, typically 10%–40%, v/v has been shown to induce the aggregation of different globular proteins (Calamai et al., 2005; Srisailam et al., 2003; Gosal et al., 2004; Munishkina et al., 2003; Pallares et al., 2004; Anderson & Webb, 2012).

In the present work, it was analysed how the TFE-induced structural rearrangements can affect the process of the amyloid formation in the FF domain.

The data reported above demonstrate that the FF molten globule (MG) in presence of TFE displays a well-defined α-helical secondary structure but a rudimentary tertiary structure with significant exposure of hydrophobic regions. The analysis of the aggregates proceeding from solutions incubated with different concentration of TFE shows that, as in the case of the MG in absence of the cosolvent, is detectable the formation of amyloid fibrils which display, however, marked tinctorial and morphological differences when compared with the fibrils formed without TFE. Additionally, the study of the aggregation kinetics reveals that the presence of the cosolvent dramatically affects the reaction rate. This effect contrasts with the pro-aggregational effect shown for TFE in the same concentration range for a wide range of proteins, indicating that the effect of this cosolvent is not generic and depends on the conformational properties of the target protein. The kinetic assays indicate that this fluoroalcohol acts at the early stages of the aggregation reaction making more difficult or impeding the structural transition from an initially monomeric and soluble form to aggregation-prone species. The conformational characterization of the protein in the absence and presence of the cosolvent suggests that this effect is exerted through a reinforcement of the global helical propensity of the protein. TFE is a solvent less polar than water and a weaker hydrogen bond competitor. Accordingly, it has been suggested that its ability to promote the formation of ordered aggregates depends on an strengthening of the interactions that stabilize intermolecular β-sheet structure as well as intramolecular β-turns; however, our data suggest that in the case of all α-helical proteins, such as URN1-FF, it also promotes a reinforcement of the backbone hydrogen bonds sustaining the helical structure, which competes with the formation of the intermolecular hydrogen bonds necessary to align the main chain in β-sheet enriched conformations. In contrast to what is observed in other protein models containing preformed β-strands in their native state, by favoring stronger interactions, TFE changes the balance factors that favor solubility relative to aggregation, since in α-helical proteins they result in the formation of native-like contacts, whereas for β-sheet containing proteins its effect is disruptive and promote non-native intermolecular binding. In URN1-FF, this over-stabilizing effect of the native secondary structure can overcome the disruption of the tertiary structure and thus the presence of hydrophobic residues exposed to solvent. In support of this view, it has been shown that mutations in proteins that stabilize α-helical structure can slow down the process of aggregation from their denatured states (Taddei et al., 2001).

The different seeding and cross-seeding ability displayed for the f0% and f15% URN1-FF fibrils upon the amyloid formation in absence or presence of TFE suggests that this cosolvent affects not only the nucleating step but it also leads to structural changes in the formed fibrils. These results are consistent with recent data obtained for barstar protein, in which the protofibrils formed in the presence and absence of

trifluorethanol were shown to differ not only in the orientation and number of β-sheet motifs but also in the number of protein residues involved in the maintenance of the protofibrillar structure (Jha et al., 2012). The lack of seeding capability of f0% fibrils when the reaction occurs in the presence of TFE indicates that they differ in conformation to f15% fibrils; however, since both f0% and f15% fibrils are able to seed the aggregation reaction in aqueous solution, it is difficult to argue that they will not share that f0% fibrils depolymerize fast in the presence of TFE, which would preclude them to act as effective seeds in this condition. In support of this view, when f0% fibrils were diluted in 15% TFE (v/v) the sequential/structural determinants responsible for amyloid propagation. Yet another possibility is the Th-T fluorescence decreased rapidly towards the control intensity obtained without fibrils, indicative of a high rate of depolymerization in the presence of the cosolvent (Figure IV-2.2.6). Despite the fact that f15% fibrils also exhibit a decay in Th-T binding upon dilution, the resulting reaction is 6-fold slower and the final Th-T fluorescence stronger, indicating that a higher population of f15% fibrils remained at equilibrium. Similarly, the fibrils formed by the amyloidogenic β2-microglobulin protein under different TFE concentrations differ in their conformational properties, in such a way that the fibrils with slower depolymerization reactions are those more formed at higher concentrations of TFE, exhibiting at the same time the slower aggregation kinetics (Chatani et al., 2012). This apparent paradox can be solved if one considers α-helix and cross-βsheet as two sides of the same coin, both structures being stabilized by extended networks of hydrogen bonds, the formation of which is facilitated in the presence of TFE.

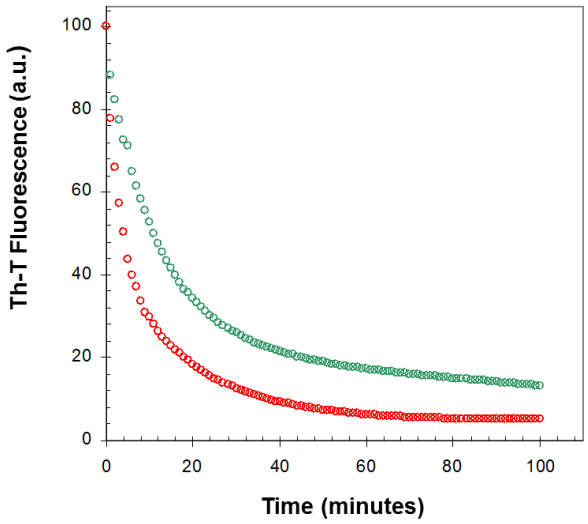

Figure IV-2.2.6|**Depolymerization of URN1-FF fibrils in the presence of 15% TFE.** Change in Th-T fluorescence (25 μM) after 5-fold dilution of URN1-FF f0% (red) and f15% (green) fibrils in 15% TFE.

As a result, in proteins like the URN1-FF domain, displaying intrinsic α-helical propensity, the presence of TFE would result in a high energy barrier of fibrillation. This

barrier would be much lower in proteins containing regions with low helical propensity, such as β-strands, explaining why, in those cases, TFE act at moderate concentrations as a pro-aggregational compound.

The toxicity of amyloid fibrils is related to their conformational properties. It has been shown for different and unrelated proteins that the binding to ANS-like dyes correlates with the toxicity of amyloid species, suggesting that the exposure of hydrophobic regions is a critical characteristic of these pathogenic assemblies. Although the aggregation of URN1-FF is not associated to any known disease, the striking different binding of the fibrils formed in the absence and presence of TFE to bis-ANS suggest that the polarity of the microenvironment might be an important determinant of the toxicity of the resulting fibrils, an effect that could be relevant inside the cell where aggregation might occur in the cytosol but also at the membrane surface, which lipid molecules resemble in their chemical properties to TFE, a cosolvent usually employed as a membrane mimic.

Methods

Protein Expression and Purification

Competent *E. coli* BL21(DE3) cells were transformed with a pETM-30 plasmid encoding the URN1-FF domain that corresponds to residues 212–266 of yeast URN1; it was cloned as an N-terminal fusion protein with a His tag followed by GST and a TEV protease cleavage site. Transformed cells were incubated in Luria Bertani medium overnight at 310 K and then diluted to 1/100 (v/v). After growing to 0.6 optical density they were induced with 1 mM IPTG overnight at 298 K. A His-tag column was used to isolate the URN1-FF fusion. The GST protein was removed by TEV cleavage and a final gel filtration on a HiLoad™ Superdex™ 75 prepgrade column (Amersham Pharmacia Biotech AB, Uppsala, Sweden). The protein was dialyzed against water and lyophilized. Protein concentration was determined by UV absorption using a ε value of 1.948 $mg^{-1}{\cdot}mL{\cdot}cm^{-1}$.

Protein Preparation for URN1-FF Conformational Assays

Lyophilized URN1-FF protein was prepared at 20 µM in 50 mM glycine at pH 2.5 and different amounts of TFE were added. Protein solution was filtered through a 0.22 µm filter and immediately analyzed at 298 K.

Protein Preparation for URN1-FF Aggregation Assays

Lyophilized URN1-FF protein was dissolved at 100 µM in 100 mM glycine at pH 2.5 in the presence or absence of TFE and filtered through a 0.22 µm filter. The samples were incubated under agitation at 400 rpm and 310 K.

Circular Dichroism and Intrinsic Tryptophan Fluorescence

Monomeric and aggregated URN1-FF species were dissolved at 20 µM and measured immediately. Far-UV CD spectra were measured in a Jasco-710 spectropolarimeter (Jasco Corporation, Hachioji, Japan) thermostated at 298 K. Spectra were recorded from 260 to 200 nm, at 0.2 nm intervals, 1 nm bandwidth, and a scan speed of 200 nm/min. For eachspectrum were averaged twenty accumulations. Tryptophan intrinsic fluorescence was measured at different temperatures in a Cary-100 Varian spectrofluorometer (Varian, Inc., Palo Alto, CA, USA) using an excitation wavelength of 280 nm and recording the emission from 300 to 400 nm. Three averaged spectra were acquired and slit widths were
typically 5 nm for excitation and emission.

bis-ANS Binding Assay

Aggregated samples were diluted at 10 µM in phosphate buffer pH 7.5 containing 25 µM of bis-ANS. To study soluble URN1-FF species, samples were prepared at 20 µM containing 25 µM of bis-ANS and analyzed immediately. The excitation wavelength was 370 nm and the emission spectra were recorded between 400 and 600 nm, using excitation and emission slit widths of 5 nm. Three spectra were accumulated after 5 min of equilibration at different temperatures in a CARY-100 Varian Spectrophotometer (Varian, Inc., Palo Alto, CA, USA).

Binding to Th-T

Aggregated URN1-FF was diluted to 10 µM in phosphate buffer at pH 7.5 in the presence of 25 µM of ThT. The sample was excited at 440 nm and fluorescence emission was acquired between 460 and 600 nm, using excitation and emission slit widths of 5 nm. Each trace was the average of 3 accumulated spectra at 298 K in a CARY-100 Varian spectrophotometer (Varian, Inc., Palo Alto, CA, USA).

Binding to Congo Red

Thirty microliters of aggregated URN1-FF were added to 220 µL of CR (20 µM) in 5 mM phosphate, 150 mM NaCl pH 7.4 buffer at 25 °C. After 5 min of equilibration, optical absorption spectra were recorded from 400 to 700 nm and accumulated for 3 times with a Jasco V-630 spectrophotometer (Jasco Corporation, Hachioji, Japan). Solutions containing only protein and only CR were analyzed to eliminate the protein scattering and dye contribution to the spectra.

Electron Microscopy (EM)

The analysis was performed using a HITACHI H-7000 transmission electron microscope (Hitachi, Tokyo, Japan) operating at an accelerating voltage of 75 kV. Ten microliters of incubated samples were diluted tenfold with water and were placed on carbon-coated

copper grids. After 5 min the grids were washed with distillated water and stained with 2% (w/v) uranyl acetate for 1 min.

Aggregation Kinetics and Seeding Assays

URN1-FF protein was prepared at 100 µM in 100 mM glycine at pH 2.5 in the presence of 25 µM of Th-T and 0%, 15% and 25% TFE (v/v). Immediately after equilibrating the sample at 310 K during 5 min, Th-T intrinsic fluorescence was measured every 2 min. The sample was excited at 440 nm and emission was recorded at 475 nm for Th-T. Slitwidths of 5 nm and 10 nm were used for excitation and emission respectively in a CARY-100 Varian spectrophotometer (Varian, Inc., Palo Alto, CA, USA). For seeding assays, 10% (v/v) of preformed fibrils formed under different conditions were added to soluble URN1-FF at 100 µM in the absence or presence of TFE, immediately before incubation; the kinetics were followed as described above.

Depolymerization Assays

Ten percent (v/v) of preformed fibrils formed under different conditions were diluted 5-fold in 100 mM glycine at pH 2.5 containing 15% of TFE (v/v). The sample was excited at 440 nm and emission was recorded at 475 nm for Th-T. Slit widths of 5 nm and 10 nm were used for excitation and emission respectively in a CARY-100 Varian spectrophotometer (Varian, Inc., Palo Alto, CA, USA). Th-T intrinsic fluorescence was measured every minute.

2.3 Polypeptide chain cross-linking by a novel disulfide bond affects the stability, folding and amyloid fibril formation capability of the all-α FF domain

In this work, the URN1-FF domain was engineered to obtain a mutational variant with a novel disulfide bond in order to investigate the covalent crosslink impact upon protein folding and amyloid formation in an all α helical protein model.

Mutagenesis of *yeast* URN1 FF domain and disulfide bond formation

The FoldX algorithm (Schymkowitz et al., 2005a) was used to design an URN1-FF domain variant in which the wild type Cysteine residue (Cys57) is linked to a selected mutated residue by a disulfide bridge. After computational analysis, Lys23 was chosen as optimal candidate for mutation into a Cys residue, considering their proximity (distance between the β-carbons = 3.2Å), face-to-face orientation, and proper dihedral angles (Figure IV-2.3.1A). The FF mutant (FFK23C-SS) was cloned and produced in *E. coli*. The designed covalent bond was spontaneously formed *in vivo* as confirmed by the absence of detectable alkylation by vinylpyridine in the recombinant purified protein (FF WT≈ FF K23C-SS≈ 7090 Da) (Figure IV-2.3.1B).

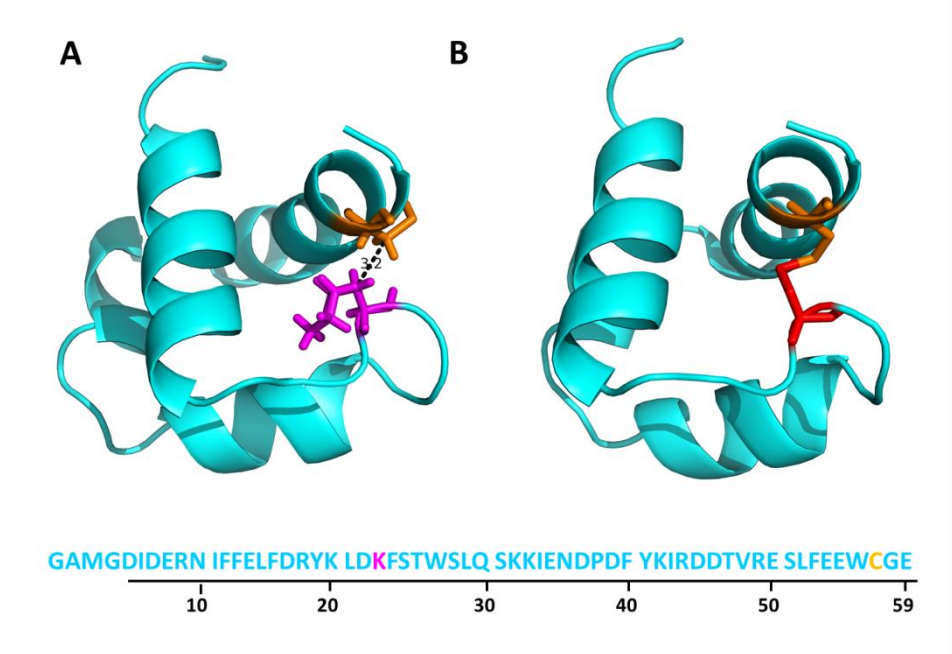

Figure IV-2.3.1| **Structure representation of FF WT and FF K23C-SS mutant.** (A) Ribbon representation of wild type FF and Lys23 (magenta) and Cys57 (orange) β-carbons distance are shown. (B) Structural model of FFK23C-SS mutant showing disulfide bond formed by mutated Cys23 (red) and wild type Cys57 (orange).The Protein Data Bank accession code for the structures is 2JUC. The figure was prepared with PyMOL (www.pymol.org).

Conformational properties of FF domains under physiological conditions

The far-UV CD spectrum of the mutant domain (FF K23C-SS) under native conditions corresponds to that of a canonical α-helical protein as observed in the wild type domain (FF WT) displaying, however, an slightly more intense peak at 210 nm. The reduced form of the mutant (FF K23C-SH_2) also exhibits the typical α-helical pattern, but displays lower ellipticity (Figure IV-2.3.2A). The analysis of the intrinsic fluorescence contributed by the two tryptophan (Trp) residues at positions 27 and 56 of the URN1 FF domains shows that the mutant FF K23C-SS emission spectrum displays lesser intensity and is slightly blue-shifted compared with that of the FF WT suggesting deeper burying of Trp residues into the globular structure. Conversely, the reduced mutant FF K23C-SH_2 spectrum shows an increase of fluorescence intensity with a red-shift respect to the wild type suggesting a major exposure of Trp residues and a certain opening of the globular structure (Figure IV-2.3.2B).

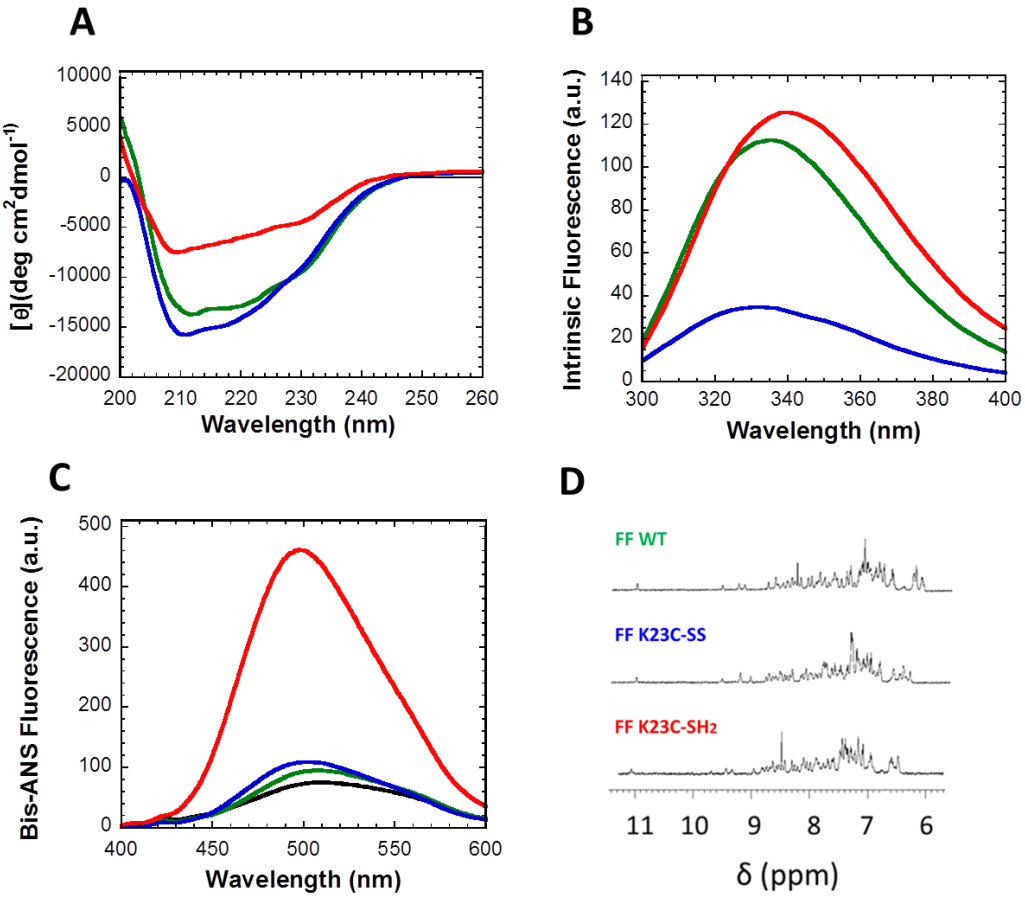

Figure IV-2.3.2| **Conformational analysis of FF domains**. FF protein solutions were studied at 298 K monitoring (A) far-UV CD signal, (B) tryptophan intrinsic fluorescence and (C) bis-ANS fluorescence. Free Bis-ANS emission spectrum is represented with a black continued line. FF-WT, FF K23C-SS and FF K23C-SH_2 are shown in green, blue and red, respectively. (D) One-dimensional NMR (^{1}H-NMR) spectra of FF proteins were recorded at 298 K and 600 MHz, using a protein concentration of 50 µM.

This behaviour is confirmed by the bis-ANS binding ability of FF K23C-SH_2 which shows a higher intensity in the fluorescence emission with a blue-shift of its maximum compared with that of the wild type domain indicating an increased exposure of hydrophobic clusters. FF K23C-SS bis-ANS emission spectrum indicates a similar binding to that of FF WT indicating low exposure of hydrophobic regions (Figure IV-2.3.2C). The one-dimensional NMR (^{1}H-NMR) spectrum of the three domains under native conditions (298K and pH 5.7) displays a wide signal dispersion of resonances with good peak sharpness, characteristic of folded molecules (Figure IV-2.3.2D).

Thermal denaturation of URN1 FF domains

Thermal stability of FF domains was analysed at pH 5.7 by circular dichroism (CD), intrinsic fluorescence and differential scanning calorimetry (DSC) (Figure IV-2.3.3). The thermal denaturation curves were followed by far-UV CD at 222 nm (IV-2.3.3CA) and by intrinsic fluorescence emission at 360 nm (IV-2.3.3B) showing a single cooperative transition that could be fitted to a two-state unfolding model for FF WT and FF K23C-SH_2.

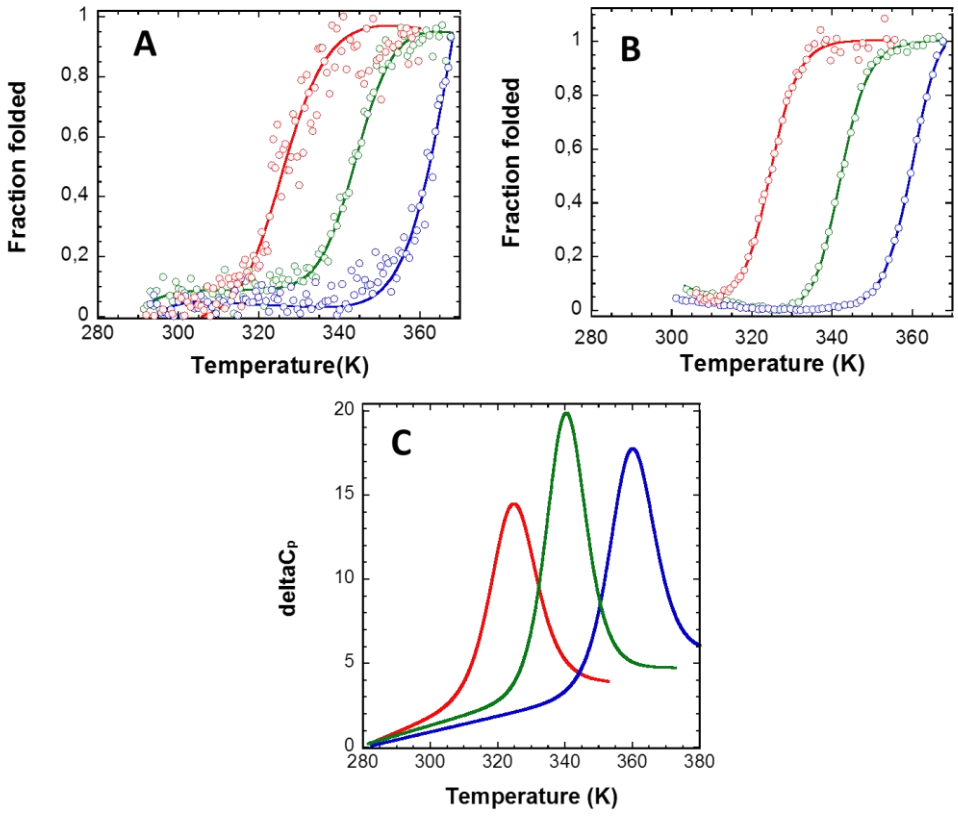

Figure IV-2.3.3| **Thermal denaturation of FF domains.** Thermal stabilities were analysed by (A) far-UV CD signal change at 222 nm, (B) intrinsic fluorescence emission at 360 nm and (C) differential scanning calorimetry. FF-WT, FF K23C-SS and FF K23C-SH_2 are shown in green, blue and red, respectively.

FF K23C-SS turned to be highly stable, not being completely denatured even at 95 ˚C, which precluded fitting of the thermal curves monitored both by CD and fluorescence. DSC scans of the three FF domains exhibit a single cooperative transition corresponding to a two-state unfolding process without a significant population of intermediate partially unfolded states, according to the ratio between the van't Hoff and calorimetric enthalpies (Figure IV-2.3.3C). The thermal stability of FF domain is drastically affected by the Lys23 mutation as revealed by the extraordinary increase of the FF K23C-SS melting temperature (Tm), ≈20K higher than the wild type domain, and the significant decrease of FF K23C-SH_2 Tm, ≈ 15K lower than the wild type (Table IV-2.3.1). When they could be measured, the transition unfolding temperatures obtained by the three independent methods were very similar (Table IV-2.3.1) indicating that the reduced FF mutant is less stable compared to the FF WT, and FF K23C-SS is considerably more stable towards thermal denaturation than the wild-type. The temperature-induced unfolding of FF domains was also monitored by ^{1}H-NMR (Figure IV-2.3.4). FF WT displays native signal dispersion and peak sharpness until 328 K; above this temperature the intensity of the signals gradually decays until at 348K the spectrum collapses and the resonances at low fields are hardly detectable, indicating the absence of a preferential folded conformation. In the case of FF K23C-SS, the spectra remain essentially unchanged until 343K and even at 348K the spectrum presents good signal dispersion, although with decreased peak intensity. FF K23C-SH_2 spectrum displays a good signal dispersion just until 320K and a rapid loss of the signal is observed above this temperature. The analysis of the spectra at 338K demonstrates clearly the different thermal stability of the three FF domains in the present study (Figure IV-2.3.4D).

Table IV-2.3.1| **Thermodynamic parameters obtained from thermal denaturation at pH 5.7**

	ΔH (T_m)(kcal mol^{-1})[a]	**ΔC_p (kcal K^{-1} mol^{-1})[a]**	**Tm(K)[a]**	**Tm(K)[b]**	**Tm(K)[c]**
FF-WT	59.05 ± 0.18	0.31 ± 0.05	340.24 ± 0.07	340.9 ± 0.3	342.6 ± 0.1
FF K23C-SS	57.84 ± 0.48	0.17 ± 0.09	359.9 ± 0.2	-	-
FF K23C-SH_2	46.60 ± 0.24	-	325.0 ± 0.3	324.9 ± 0.3	324.3 ± 0.1

[a] DSC
[b] CD
[c] Intrinsic Fluorescence

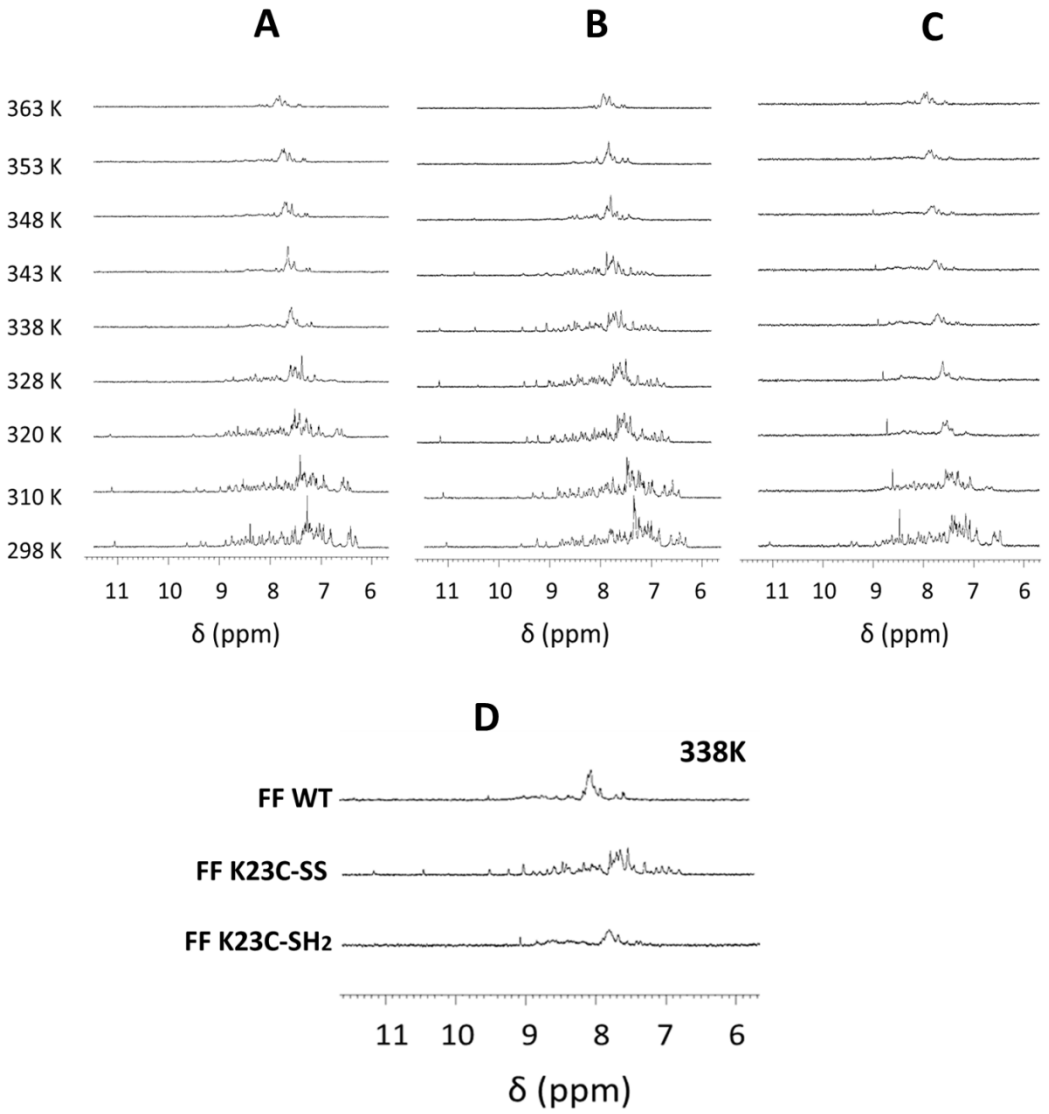

Figure IV-2.3.4| **^{1}H-NMR spectra of FF-WT (A), FF K23C-SS (B) and FF K23C-SH$_2$ at the indicated temperatures. (D) Comparative spectra of the three FF domains at 338K.**

Equilibrium unfolding of FF domains

The resistance against chemical denaturation with guanidinium chloride (GuHCl) was used to analyse differences in the stability of the three different FF domains. The chemical denaturation curves at equilibrium were followed at 298K and pH 5.7 monitoring the changes in molar ellipticity at 222 nm (Figure IV-2.3.5A) and in intrinsic fluorescence at 360 nm (Figure IV-2.3.5B) under increasing GuHCl concentrations. All three FF domains display a single visible transition indicative of the unfolding process cooperativity. The main thermodynamic parameters of the unfolding reaction were calculated from the equilibrium curves assuming a two-state model (R> 0.99) (Table IV-2.3.2). The stability of URN1 FF WT calculated from fluorescence and CD measurements is similar with $\Delta G^{(H20)}_{U\text{-}F} \approx 5.5$ kcal/mol and $[GuHCl]_{50\%}$ of 2.8M. In

agreement with the thermal denaturation data, FF K23C-SH_2 results to be less stable than the wild type domain, exhibiting a $\Delta G^{(H20)}_{U\text{-}F}$ of ≈3.6 kcal/mol and $[GuHCl]_{50\%}$ of 2.29M. Conversely, the value of $[GuHCl]_{50\%}$ for FF K23C-SS is too high (>5 M) and a final baseline for the fully unfolded state cannot be accurately measured. Therefore, the FF K23C-SS stability was studied analysing the equilibrium denaturation with the strong chaotropic agent guanidinium thiocyanate (GITC) (Figure IV-2.3.5C). Despite wild type is not stable enough to be analysed in the presence of GITC, it is clear that crosslinking notably stabilizes FF K23C-SS in front of chemical denaturation as confirmed by the calculated thermodynamic values (Table IV-2.3.2). Interestingly, the dithiol form of the proteins shows a marked destabilization relative to the FF WT suggesting that the mutated residue is not sterically or electrostatically acceptable in the reduced state.

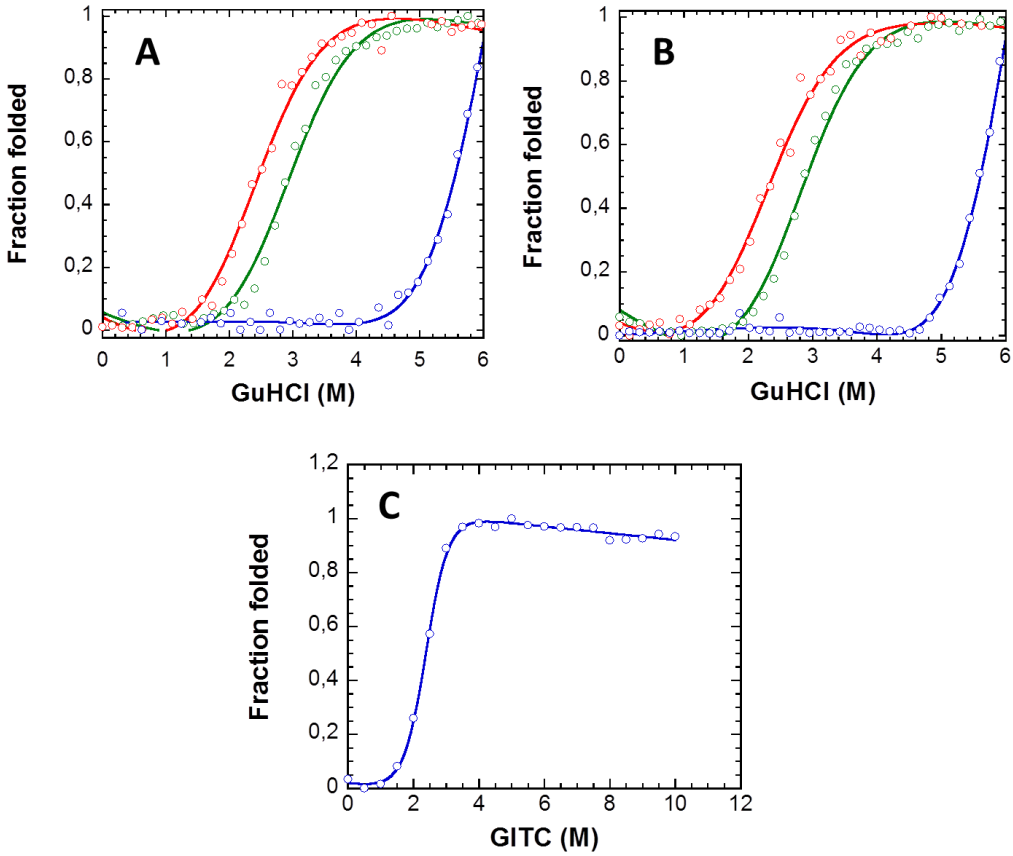

Figure IV-2.3.5| **Chemical denaturation of FF domains.** Chemical equilibrium curves with guanidinium chloride (GuHCl) were analyzed at 298 K by (A) far-UV CD at 222 nm and by (B) tryptophan intrinsic fluorescence at 350 nm. (C) Equilibrium curve with guanidinium thiocyanate (GITC) monitored by intrinsic fluorescence was performed for FF K23C-SS. FF WT, FF K23C-SS and FF K23C-SH_2 are shown in green, blue and red, respectively.

Table IV-2.3.2| **Thermodynamic parameters of FF domains equilibrium unfolding**

	[a] $\Delta G^{(H_2O)}_{U-F}$ (Kcal/mol)		[b] m (Kcal/mol·M)		[c] [GuHCl]50% (M)	
	CD	**Intrinsic Fluorescence**	**CD**	**Intrinsic Fluorescence**	**CD**	**Intrinsic Fluorescence**
FF WT	6.20 ± 0.33	5.09 ± 0.22	2.14 ± 0.10	1.82 ± 0.07	2.79 ± 0.02	2.80 ± 0.02
FF K23C-$(SH)_2$	3.83 ± 0.43	3.30 ± 0.53	1.57 ± 0.14	1.43 ± 0.18	2.43 ± 0.05	2.29 ± 0.07
FF K23C-SS	-	-	-	-	> 5	> 5
FF K23C-SS[d]	-	4.14 ± 0.36	-	1.74 ± 0.12	-	2.37 ± 0.04

[a] Gibbs energy of unfolding at [denaturing agent]=0

[b] m value, dependence of free energy of unfolding with denaturating agent

[c] The denaturating agent concentration required to unfold 50% of the protein molecules

[d] Parameters obtained by the equilibrium unfolding with Guanidinium thiocyanate (GITC)

Folding and unfolding kinetics of FF domains

The kinetics of folding and unfolding of the FF domains were determined by stopped-flow at pH 5.7 and 298 K under a wide range of denaturant conditions. In all cases, the folding and unfolding traces by fluorescence fit well into single exponential functions. The chevron plots appear to be linear in the complete range of denaturant concentrations studied, indicating the lack of detectable intermediates, according to a two-state model (Figure IV-2.3.6). The rate constants for folding (K_f) and unfolding (K_u) and their dependence on the denaturant concentration (m_f and m_u) are shown in Table IV-2.3.3. The kinetic data of FF domains are in good agreement with the equilibrium stabilities reported above. The FF K23C-SH_2 folding rate appears very similar to that of FF WT, while its unfolded rate results highly affected, suggesting that the mutated Lys plays an important role in maintaining the native conformation even though is not likely involved in formation of the folding nucleus as indicated by the calculated ϕ_{F-U} value of 0.12 (Figure IV-2.3.6A). Due to its high resistance to urea, FF K23C-SS kinetic measurements were performed with guanidinium chloride (GuHCl) and compared to those of URN1 FF WT (Figure IV-2.3.6B). Remarkably, FF K23C-SS shows a folding rate extraordinary faster than the wild type k_f. This amazing acceleration of the folding reaction can be univocally attributed to the presence of the disulfide bridge, since the thiol form exhibits a folding rate close to that of FF WT, as discussed above. As expected, the crosslinking acts also upon the unfolding rate which results significantly slower than the wild type domain indicating the disulfide bond contribution to the maintenance of the native conformation. The calculated ϕ_{F-U} value of 0.84 for the FF

K23C-SS mutant suggests that the establishment of the disulphide bond might participate actively in the formation of the folding nucleus of the crosslinked protein.

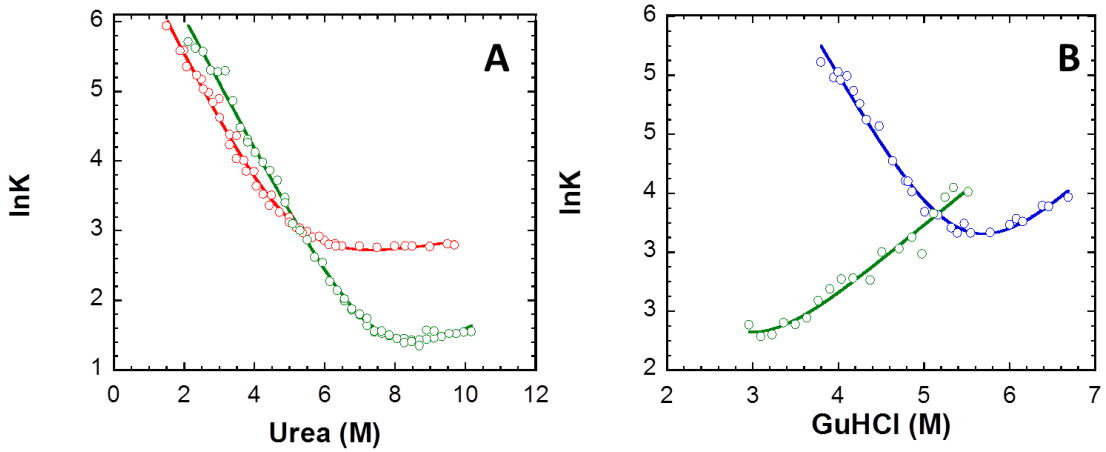

Figure IV-2.3.6| **Folding and unfolding kinetics of FF domains.** (A) Dependence of the folding and unfolding rate constants (k) on urea concentration of FF WT (green) and FF K23C-SH_2 (red). (B) Dependence of the folding and unfolding rate constants (k) on guanidinium chloride concentration of FF WT (green) and FF K23C-SS (blue).

Table IV-2.3.3| **Folding kinetic parameters for FF domains**

	K_f (s^{-1})	K_u (s^{-1})	m_f (Kcal/mol·M)	m_u (Kcal/mol·M)	[a] $m_{U\text{-}F}$ (Kcal/mol·M)	[b] $[\,]_{50\%}$ (M)	[c] $\Delta G_{U\text{-}F}$ (Kcal/mol)	[d] $\Delta\Delta G_{U\text{-}F}$ (Kcal/mol)
FF WT (Urea)	2819 ± 198	0.34± 0.12	0.94 ± 0.02	0.26 ± 0.04	0.71 ± 0.06	7.48	5.34	–
FF K23C-(SH)2 (Urea)	1846 ± 145	7.8 ± 1.5	1.00 ± 0.03	0.08 ± 0.02	0.64 ± 0.05	5.03	3.24	-2.1
FF K23C-SS (GuHCl)	167110 ± 69447	0.16 ± 0.16	1.67 ± 0.10	0.14 ± 0.02	1.5 ± 0.25	5.43	8.21	4.97

[a] Dependence of the Gibbs energy of unfolding with urea/GuHCl

[b] The urea/ GuHCl concentration required to unfold 50% of the protein molecules

[c] Gibbs energy of unfolding with urea/GuHCl determined from the kinetic parameters

[d] The difference in free energy between the mutant protein and WT protein

FF aggregation under native conditions

The FF domains were incubated at pH 5.7 and 310 K during one week to test their ability to self-assembly into macromolecular structures. The presence of amyloid aggregates was preliminary analysed using the amyloid specific dye Th-T binding (Figure IV-2.3.7A).

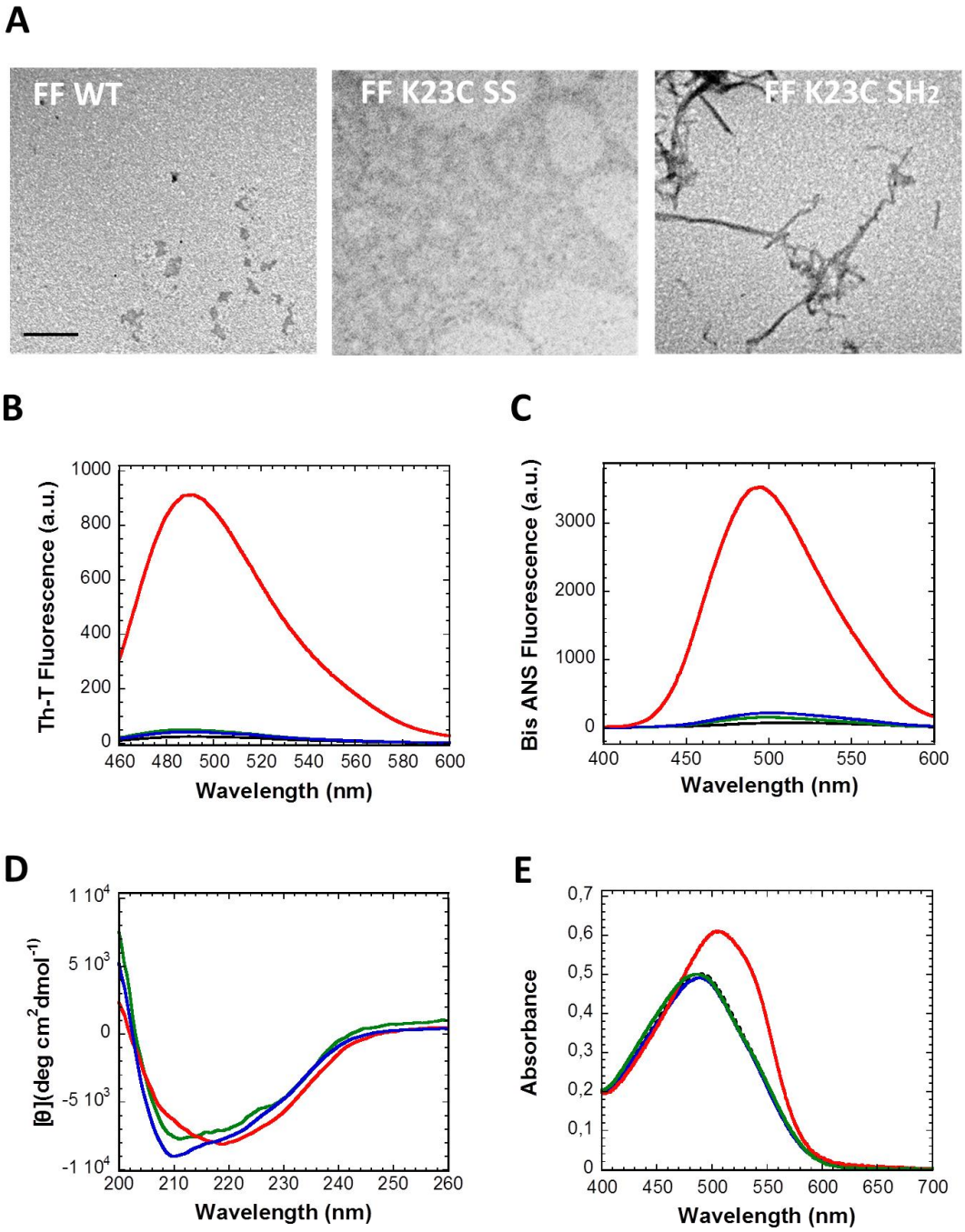

Figure IV-2.3.7| **Morphological, structural and tinctorial properties of FF aggregates at pH 5.7.** FF aggregates were analyzed by (A) Th-T fluorescence, (B) Congo Red union (C) far-UV CD spectroscopy, (D) binding to Bis ANS dye, and (E) TEM microscopy,. FF-WT, FF K23C-SS and FF K23C-SH_2 samples are represented in green, blue and red, respectively. Free Congo red spectrum is indicated with a black dotted line. The scale bar represents 200 nm.

The analysis of FF K23C-SH_2 solution in presence of Th-T promotes a significant increase in the fluorescence emission, while both wild type and FF K23C-SS solutions do not bind the amyloid dye. Accordingly to these results, only for the FF K23C-SH_2 aggregates is observable Congo Red (CR) binding, with the characteristic red-shift at ≈

540 nm and increase in the absorbance maximum of the dye (Figure IV-2.3.7B). The conformational change of the native soluble monomer into aggregated species was monitored by far-UV CD (Figure IV-2.3.7C). The FF α-helical native spectrum with the characteristic minima at 210 and 222 nm is observed for the incubated FF WT and FF K23C-SS proteins. In contrast, the formation of a β-sheet enriched conformation is clearly detectable for the incubated FF K23C-SH_2 domain. Then, the bis-ANS binding was used to monitor the exposure of hydrophobic clusters to the solvent in the putative aggregated states showing an increase of bis-ANS emission only for the FF K23C-SH_2 (Figure IV-2.3.7D). The morphological analysis of the FF aggregates by transmission electron microscopy (TEM) confirms the presence of amyloid-like fibrillar structure only for the FF K23C-SH_2, with the FF WT and FF K23C-SS solutions lacking aggregates (Figure IV-2.3.7E).

The aggregation rate of FF domains was analyzed at pH 5.7 and 310 K monitoring the time course changes in the fluorescence of the amyloid staining dye Th-T (Figure IV-2.3.7).

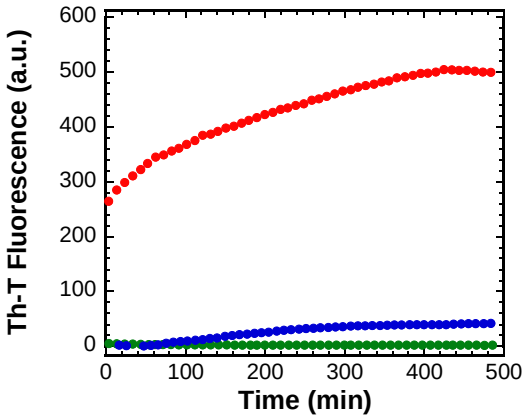

Figure IV-2.3.8| **Aggregation kinetics of FF domains at pH 5.7.** Aggregation kinetics of FF-WT (green), FF K23C-SS (blue) and FF K23C-SH_2 (red) were monitored at 310 K following Th-T emission fluorescence at 475 nm.

The kinetics of amyloid fibril formation generally follows a sigmoidal curve reflecting a nucleation-dependent growth mechanism. In the case of FF K23C-SH_2 domain, the aggregation process shows a kinetic scheme where the initial lag phase is completely abolished indicating that the dithiol form aggregation occurs very rapidly under these conditions. FF WT and FF K23C-SS domains do not exhibit significant Th-T binding according to the data reported above.

Thus, in agreement with previous reported data (Castillo et al., 2013), FF WT does not form amyloid-like aggregates under physiological conditions. Interestingly, the mutation of the Lys23 to a cysteine drastically accelerates the amyloid formation when it is in its reduced form, while its oxidation into a disulfide bond precluding this pro-aggregative effect.

Conformational and stability analysis of FF domain under low pH

As discussed in section IV-2.2 the URN1 FF protein, at pH 2.5, forms a Molten Globule (MG) state that retains most of the α-helical secondary structure content characteristic of the protein native state (Castillo et al., 2013; Marinelli et al., 2013). The FF K23C-SS was analyzed under acidic conditions to study the effect of the disulfide introduction upon the MG state. The far-UV CD analysis of FF K23C-SS at pH 2.5 corresponds to that of an α-helical conformation, as in the native state but with slightly reduced ellipticity, as in the case of the wild type (Figure IV-2.3.9A).

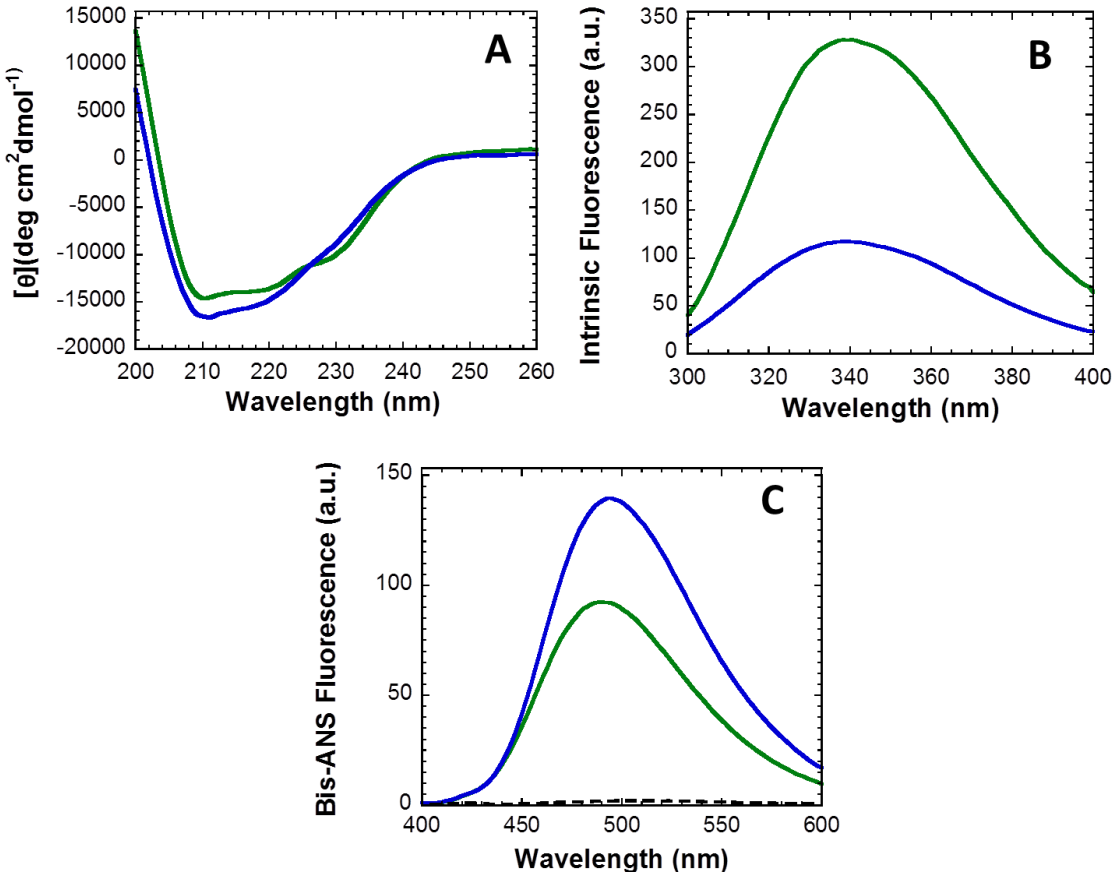

Figure IV-2.3.9| **Conformational analysis of FF domains at pH 2.5**. FF protein solutions analysed at pH 2.5 (solid line) and at pH 5.7 (dotted lines) were monitored at 298 K by (A) far-UV CD signal, (B) tryptophan intrinsic fluorescence and (C) bis-ANS fluorescence. Free Bis-ANS emission spectrum is represented with a black solid line. FF-WT and FF K23C-SS are shown in green and blue, respectively. FF samples measured at pH 5.7 are shown with dotted lines.

The intrinsic fluorescence spectra of both acid induced FF domain states show an increase of signal emission and a slight red-shift compared with that of their native states according to a higher opening of their globular structures at low pH (Figure IV-2.3.9B). The FF WT spectrum displays, however, higher fluorescence emission intensity than the FF K23C-SS domain showing a major exposure of Trp residues in the wild type molten state than in the conformation attained by the FF mutant. These data suggest

that the covalent bond introduction might regulate the compactness of the globular structure at low pH as well as at physiological conditions. As expected, both FF domains promote an increase in the bis-ANS fluorescence emission compared with the native proteins confirming the exposure of hydrophobic regions under acidic conditions (Figure IV-2.3.9C). Interestingly, the FF K23C-SS spectrum shows a higher bis-ANS intensity emission than the wild type domain indicating that a different conformational rearrangement occurs in presence of the disulfide bond.

As reported previously by our group, the URN1 FF thermodynamic stability is drastically affected at lower pHs (Castillo et al., 2013). To test the pH effect upon the FF mutant stability, the thermal unfolding of FF domains was analysed at pH 2.5 by intrinsic fluorescence and by differential scanning calorimetry (DSC) (Figure IV-2.3.10).

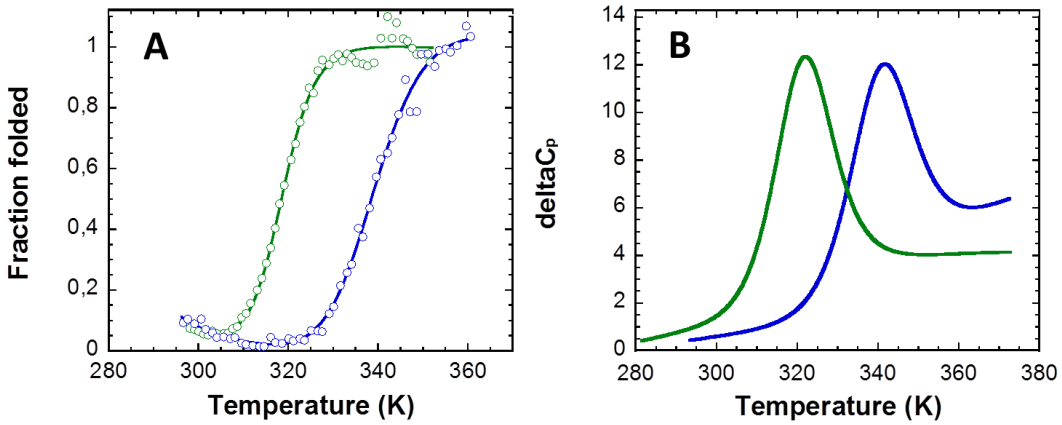

Figure IV-2.3.10| **Thermal unfolding of FF domains at pH 2.5.** Thermal stabilities under acidic conditions were analysed by (A) intrinsic fluorescence emission at 360 nm and (B) differential scanning calorimetry. FF-WT and FF K23C-SS are shown in green and blue, respectively.

The thermal curves of the FF WT and FF K23C-SS were followed monitoring the emission changes at 360 nm. In both cases a single cooperative transition was observed and the data could be fitted accurately to a two-state temperature-induced unfolding model ($R > 0.99$). The two-state unfolding process without intermediate partially unfolded states was confirmed by the DSC analysis. As suggested by the similar transition temperatures of unfolding obtained by the two independent methods, the FF K23C-SS domain is significantly more stable toward thermal denaturation than the wild-type under low pH as reflected by its melting temperature (T_m), ≈ 20K higher than that of the FF WT domain (Table IV-2.3.4). Indeed, the stability of FF K23C-SS at pH 2.5 resembles that of the wild-type at pH 5.7.

Table IV-2.3.4| **Thermodynamic parameters obtained from thermal unfolding at pH 2.5**

	ΔH (T_m)(kcal mol^{-1})[a]	**ΔCp (kcal K^{-1} mol^{-1})[a]**	**Tm(K)[a]**	**Tm(K)[b]**
FF-WT	43.43 ± 0.22	0.38 ± 0.04	321.04 ± 0.15	318.44 ± 0.10
FF K23C-SS	44.69 ± 0.71	0.55 ± 0.12	339.7 ± 0.5	338.87 ± 0.61

[a] DSC
[b] Intrinsic Fluorescence

Amyloid fibril formation by FF domains under acidic conditions

In order to study the pH-effect upon FF domains amyloid formation, the FF proteins aggregation kinetics were analyzed at pH 2.5 and 310K monitoring the changes over time in the fluorescence of the amyloid staining dye Th-T (Figure IV-2.3.11A). Interestingly, the FF K23C-SS amyloid assembly is faster and higher than that of the wild type, which might owe to the higher exposure of hydrophobic regions observed in the FF K23C-SS soluble monomer under acidic conditions (Figure IV-2.3.9C). The FF aggregation reactions were also followed at different time points by transmission electron microscopy (TEM) (Figure IV-2.3.11B). Abundant early FF K23C-SS aggregates were visible after 60 min of incubation, whereas few wild type aggregates were observed, in agreement with the Th-T intensity in the respective kinetics. At 240 min, however, the wild type domain forms ordered fibrils while only short assemblies are observable for the FF K23C-SS domain. These differences in morphology are also observed after 1000 min of incubation indicating that the disulfide introduction may interfere in fibril structuration. FF mature fibrils incubated for one week at pH 2.5 and 310K were analysed by Th-T, CR and bis-ANS binding (Figure IV-2.3.12A,B,C). Both FF fibrils result positive to amyloid dyes displaying, however, different binding strength indicating a probable distinct conformational rearrangement of the fibrils, as also suggested by the dissimilar emission intensity shown in bis-ANS binding spectra. The observation of FF mature fibrils by transmission electronic microscopy (TEM) confirms the presence of two distinct morphological species (Figure IV-2.3.12D). The analysis of FF WT mesh shows long, unbranched and twisted fibrils with a diameter of ≈14 nm as previously reported for the FF wild type domain under acidic conditions (Castillo et al., 2013). On the other hand, unbranched straight linear fibrils with diameters of 22-30 nm are observed for FF K23C-SS sample. The evaluation of the FF fibrils secondary structure by Attenuated Total Reflectance Fourier Transform Infrared spectroscopy (ATR-FTIR) reveals similar amyloid structural components for both fibrils (Figure IV-2.3.12E). As shown in Table IV-2.3.5, deconvolution of the absorbance spectra in the amide I region shows an intense signal at ≈1615 cm^{-1} and ≈1618 cm^{-1} for wild type and mutant fibrils, respectively, indicating the presence of predominant intermolecular amyloid β-sheets (Sarroukh et al., 2013; J. Zurdo et al., 2001; Villegas et al., 2000).

Common bands around 1641 cm^{-1} associated to disordered structures were also observed for both fibrils. While, a small signal at ≈1658 cm^{-1} corresponding to α-helical structure is only detected for the FF K23C-SS fibrils. These data suggest that despite the presence of a common β-sheet signature, the different proportion of disordered and α-helical components might modulate the resulting fibrils morphology.

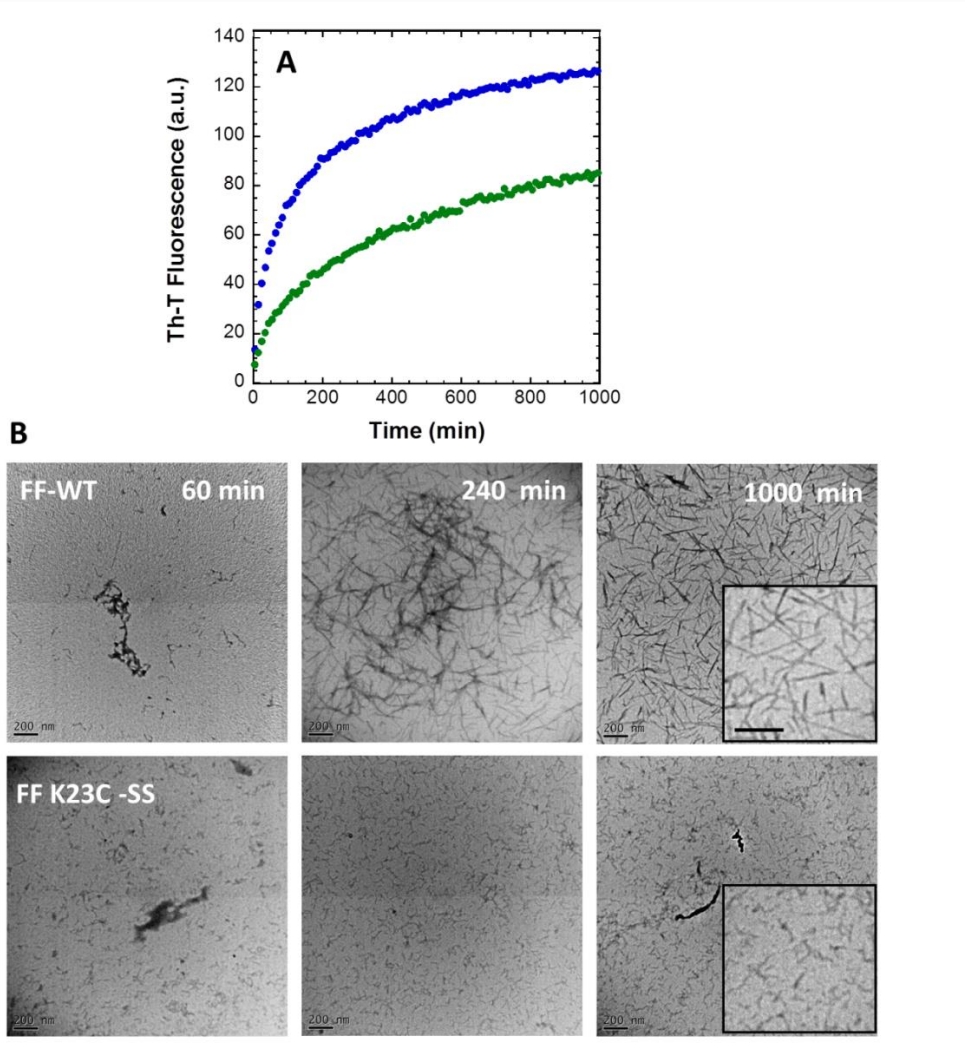

Figure IV-2.3.11| **Aggregation kinetics of FF domains under acidic conditions.** Protein solutions were incubated with agitation at pH 2.5 and 310 K. Aggregation kinetics were followed by (A) Th-T fluorescence. (B) Transmission electron microscopy images of negatively stained aggregates at different time points are shown. FF-WT and FF K23C-SS are indicated in green and blue, respectively. Bar scale represents 200 nm.

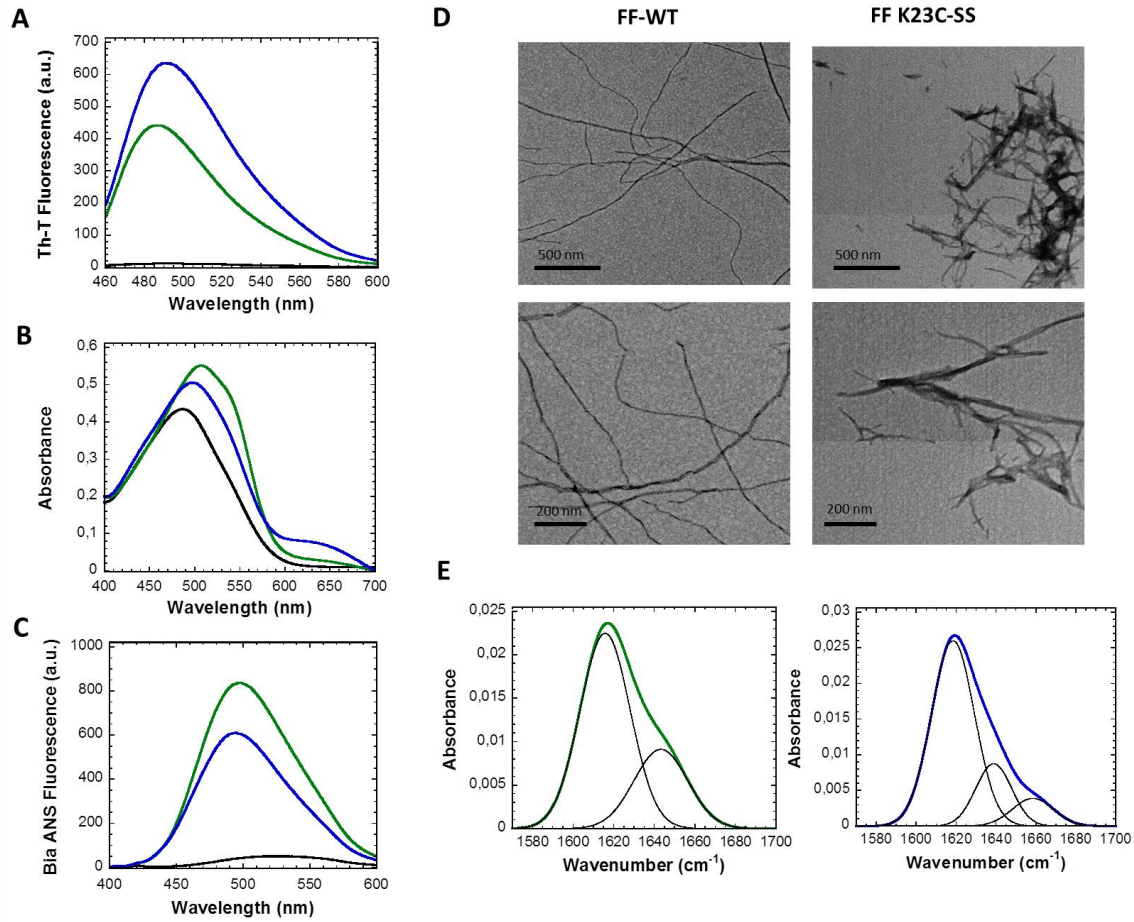

Figure IV-2.3.12| **Tinctorial, morphological and structural properties of mature FF fibrils formed at pH 2.5.** FF protein solutions were prepared at 100 µM at pH 2.5 and incubated for one week at 310 K under agitation. The fibrils were analyzed by (A) Th-T fluorescence,(B) Congo Red union and (C) binding to Bis-ANS. FF-WT and FF K23C-SS are shown in green and blue, respectively. Free Congo red spectrum is shown with a solid black line. D) Negatively stained URN1 FF fibrils were visualized by TEM microscopy at different magnification. Scale bar represents 500 (upper panels) and 200 nm (lower panels). (E) URN1 FF fibrils were characterized by ATR-FTIR showing the absorbance spectra of the amide I region (colored thick line) and the component bands (continuos thin lines). The sum of individual spectral components after Fourier self-deconvolution closely matches the experimental data.

Table IV-2.3.5| **Secondary structure bands in the deconvoluted absorbance FTIR spectra of FF fibrils formed at pH 2.5.**

	Bands (cm^{-1})	% Area	Structure
URN1 FF WT	1615	69.8	β - sheet (inter)
	1643	30.2	random
URN1 FF K23C-SS	1618	70.2	β – sheet (inter)
	1639	19.7	random
	1658	10.1	α-helix

Discussion

Protein stability is considerably increased by naturally occurring disulfide cross-links since they can globally increase up to 5-6 kcal/mol the stability of folded proteins generally reducing conformational fluctuations (Matsumura & Matthews, 1991; Betz, 1993; Darby & Creighton, 1995; Arolas et al., 2005). Remarkably, the presence of disulfide bonds in numerous precursor proteins (55%) involved in amyloidosis suggests an important role of these covalent bonds in the aggregation kinetics and in the structure and toxicity of the formed aggregates (Mossuto et al., 2011). In the present work, the introduction of a novel disulfide bridge in the yeast FF domain was analyzed in attempt to study the resulting impact upon the protein stability and aggregation of this all α helical protein model.

Computational calculation of the mutation impact on protein stability using the FoldX algorithm (Schymkowitz et al., 2005a) ($\Delta\Delta G_{total}= \Delta G_{mut} - \Delta G_{wt}$),rendered a $\Delta\Delta G= -1.28$ kcal/mol for the FF mutant in its dithiol form while, for its disulfide form, a ΔΔG value corresponding to -10.89 kcal/mole was obtained suggesting a potential over-stabilizing effect of the covalent bond. Experimental data confirm the differential effect exerted by the Lys23 to Cys mutation in the reduced and oxidized state of this residue on the FF domain stability. However, in contrast to theoretical predicted data, the dithiol form results to be significantly less stable than the wild-type domain. The lower ΔH value in FF K23C-SH_2 than in FF WT, calculated by DSC experiments, suggests that this destabilization may be caused by the loss or weakening of intramolecular and protein–solvent interactions in the native state and the further increase of the ΔS value, indicating an opening of the globular structure and hydrophobic regions exposure as experimentally observed in the dithiol form. Interestingly, in agreement with the FoldX prediction, the crosslinking of the FF mutant stabilizes the domain against both thermal and chemical denaturation, with an increase in T_m of ≈34 K and a $\Delta\Delta G_{U-F}$ at 298 K of ≈5 kcal/mol, relative to the dithiol form. The slight decrease of the

ΔH value compared with that of the wild type protein suggests that the enormous stabilization conferred by the covalent bond is due essentially to the entropy reduction of the unfolded state accordingly to the chain entropy model, which states that the major effect of the crosslink on the stability results from a decrease in the conformational entropy of the unfolded molecule (Pace et al., 1988; Fass, 2012). Additionally, a decrease in the change of heat capacity of unfolding (ΔCp) was described in numerous disulfide mutants as the cytochrome c protein (Betz & Pielak, 1992). The ΔCp between the folded and unfolded states of proteins seems to be associated with the exposure of buried side-chains to solvent. Hence, the 45% decrease in the ΔCp of FF K23C-SS likely results from a more compact unfolded state with restricted hydration of side chains, consistent with the conformational data for this analysis. It is worth to mention that, to the best of our knowledge, the over-stabilization attained in K23C-SS is the highest obtained in a small domain with a single point mutation.

The conformational entropy reduction in the unfolded state due to disulfide bond linkage is generally responsible for the acceleration in the folding rate. Decrease in unfolding rates depends on whether the regions involved in the crosslinked are or not native-like structured in the transition state. In this work, the crosslinking in the FF K23C-SS leads to an extraordinary increase of the folding rate, more 50-fold than that of the wild type, approaching the folding speed limit (Eaton et al., 2000). Moreover, the decrease in the unfolding rate suggests that the opening of the crosslinked regions may represents a rate-limiting step in the unfolding process as confirmed by the elevated unfolding rate observed for the dithiol form. Accordingly, the lower ϕ value calculated for the reduced form of the FF mutant indicates that the disulfide-forming regions are unstructured in the transition state of the wild type domain, as described for the structurally homologous human HYPA/FBP11 FF domain, which displays the C-terminal region largely disordered in its folding transition state (Jemth et al., 2005). Therefore, the different ϕ value obtained for the disulfide mutant and the corresponding dithiol form suggest that the introduction of the covalent bond may promote the involvement of the C-terminal of the protein in the folding nucleus "locking-in" the crosslinked regions.

Disulfide bonds are present in 15 % of the human proteome, in 65 % of secreted proteins, and in more than 50 % of those implied in amyloidosis (Smith & Radford, 2001; Mossuto et al., 2011). However, their effect upon the amyloid fibril formation is still far to be completely understood. The high stability and compactness of the FF K23C-SS structure at pH 5.7 prevents the protein from aggregation. By contrast, the low stability of its dithiol form leads to a very fast amyloid fibrillation in agreement with the elevate exposure of hydrophobic regions exhibited under physiological conditions. In order to understand the disulfide mutant ability to form amyloid fibrils in a denaturing environment, the FF mutant aggregation was analyzed at pH 2.5. Despite its thermal stability is very similar to that of the wild type domain at pH 5.7, the FF

K23C-SS promotes amyloid fibrils formation which is in agreement with its conformational characteristics at this pH compatible with molten globule-like conformation despite its native-like stability (Baldwin & Rose, 2013). The disulfide bond imposes a topological restraint in the helix 3 and helix1-2 loop leaving probably the helix 1 sterically free to initiate the amyloid formation. This region of the protein exhibits the highest aggregation propensity at pH 2.5 and indeed the dissected helix suffices to form amyloid fibrils (Castillo et al., 2013). Therefore, this work illustrates how, the global stabilization of a protein does not suffices to prevent amyloid formation unless it protect the specific regions that nucleate the aggregation process providing strong support to the so called amyloid-stretch hypothesis. Our data show, however, that once the aggregation process is initiated by this short stretch, the structural/topological characteristics of the aggregation prone state of the proteins play a crucial role in the kinetics, structure, morphology and likely the toxicity of the final amyloid state.

Methods

Design of disulfide bridge

The yeast URN1 FF-WT protein consists of 59 amino acidic residues and presents only a cysteine residue on position 57. To design the disulfide bond, FoldX algorithm (version 3b6) was used to screen the best position for the mutation of the second cysteine (Schymkowitz et al., 2005a). The disulfide bond formation was modeled using the Disulfide by Design 2.0 (DbD2) program, a web-based tool that predicts potential disulfide links analyzing the B-factor (temperature factor) of protein regions involved in predicted disulfide bonds generating 3D mutant models with potential disulfide bonds formed (Craig & Dombkowski, 2013). The relative difference in stability between the wild-type and mutant structure was calculated employing the FoldX BuildModel command by rotating the same residues in the wild-type and mutant structure according to the FoldX rotamer database (Schymkowitz et al., 2005a). The URN1 FF-WT Protein Data Bank (PDB) 2JUC was used as input.

URN1 FF cloning, mutagenesis and expression

cDNA of yeast URN1 FF domain was cloned into a pETM-30 vector as an N-terminal fusion protein followed by His tag, GST protein and a TEV protease cleavage site (Bonet et al., 2008). The FF containing plasmid was transformed in BL21 (DE3) *E. coli* cells that were grown at 310 K and induced with 1 mM IPTG overnight at 298 K. Affinity chromatography was used to separate His-tag proteins from cellular extract. TEV digestion of purified FF proteins and further gel filtration on HiLoad™ Superdex™ 75 column (GE healthcare Life Sciences) allowed the separation of the GST protein from FF sample that was finally lyophilized after an extensive dialysis against water. FF K23C mutant was obtained by PCR site-directed mutagenesis and the resulting sequenced

plasmid was produced and purified as described above. SDS-PAGE was used to test the purity of the samples and their concentrations were determined by UV absorption (theoretical ε = 1.948 mg^{-1} mL cm^{-1}).

Disulfide crosslinking and reduction

To check the formation of the disulfide bond, 0.3 mg/mL FF K23C-SS was incubated in 0.1 M Tris-HCl buffer, pH 8.5, containing 0.1M 4-vinylpyridine for 45 min in the dark at room temperature. After a 1:10 dilution in 0.1% aqueous trifluoroacetic acid the sample was analyzed by MALDI-TOF mass spectroscopy in an Ultraflex spectrometer (Bruker). When required, the disulfide bond was reduced by sample incubation with 50 mM DTT or 10 mM Tris (2- carboxyethyl) phosphine (TCEP) at 353 K during 30 min.

Sample preparation for conformational assays

URN1 FF samples were dissolved at 20 µM in 50 mM sodium acetate at pH 5.7 and, when required, in 50 mM glycine buffer at pH 2.5. Protein solutions were filtered through a 0.22 µm filter. Far-UV Circular Dichroism (CD) spectra were acquired from 260 to 200 nm, at 0.2 nm intervals, 1 nm bandwidth and a scan speed of 100 nm/min in a Jasco-710 spectropolarimeter (Jasco Corporation, Japan) thermostated at 298 K. Twenty accumulations were averaged for each spectrum. Tryptophan intrinsic fluorescence was monitored at 298 K in a Jasco FP8200 spectrofluorometer (Jasco Corporation, Japan). Three averaged spectra were recorded using an excitation wavelength of 280 nm and recording the emission from 300 to 400 nm; slit widths were 5 nm for both excitation and emission. To study the binding to 4,4'-dianilino-1,1'-binaphthyl-5,5'-disulfonic acid (bis-ANS), protein solutions were prepared at 20 µM in presence of 10 µM of bis-ANS and analysed immediately. Three spectra were recorded in a Jasco FP8200 spectrofluorometer (Jasco Corporation, Japan). An excitation wavelength of 370 nm and emission between 400 and 600 nm were used with a 5 nm excitation and emission slit widths.

Thermal denaturation

Thermal stabilities of URN1 FF proteins were studied by far-UV vircular dichroism (CD) spectroscopy, intrinsic fluorescence intensity and differential scanning calorimetry (DSC). Samples were prepared at 20 µM in 50 mM sodium acetate buffer at pH 5.7 to follow the molar ellipticity at 222 nm in a Jasco-710 spectropolarimeter (Jasco Corporation, Japan), each 0.1 K with 2 min of temperature equilibrium between measures. Thermal denaturation was monitored in a Jasco FP8200 spectrofluorometer (Jasco Corporation, Japan) by intrinsic fluorescence intensity recording the emission signal at 360 nm after excitation at 280 nm. The fluorescence emission was monitored each 0.5 K with 2 min of temperature equilibrium between measures, using slit widths of 5 nm for both excitation and emission wavelength. Two-state transition curve was

used to fit experimental data using the nonlinear least squares algorithm provided with Kaleidagraph (Abelbeck Software).

The thermal unfolding of the FF domains was also analyzed by differential scanning calorimetry (DSC) in a capillary-cell microcalorimeter from MicroCal (Northampton, MA, USA) at a range 0.4–0.7 mg mL^{-1} protein concentration with scan rate of 2 K/min . The partial molar heat capacity curves (Cp) were calculated from the DSC data and analyzed using Origin 8.5 (OriginLab, Northampton, MA). The Cp curves were fitted according to the two-state unfolding model to determine the unfolding thermodynamic parameters of the domain under each condition.

NMR spectroscopy

URN1 FF solutions were prepared at final concentration of 50 µM in 20 mM sodium phosphate buffer, 130 mM NaCl at pH 5.7, using a 9:1 H_2O/D_2O ratio. One-dimensional NMR spectra were collected at designed temperatures on a Bruker AVANCE 600-MHz spectrometer using solvent suppression WATERGATE techniques. Spectra were processed and analyzed using the TopSpinv2.0 software packages from Bruker Biospin.

Chemical denaturation

For chemical denaturation, samples were prepared at 20 µM in presence of different denaturant agent concentrations (0-6 M guanidinium chloride; 0-10 M guanidinium thiocyanate). The reaction was left to equilibrate for at least 3h at room temperature and analyzed by CD at 222 nm and followed by fluorescence intrinsic intensity exciting at 280 nm and monitoring fluorescence emission at 360 nm. Experimental data were fitted to a two-state transition curve as discussed above for thermal denaturation.

Folding and unfolding kinetics

Kinetics of folding and unfolding reactions were followed in a Bio-Logic SFM-3 stopped-flow instrument at 298 K using excitation at 280 nm and a 320 nm fluorescence cut-off filter. A 20 µM protein solution was prepared and appropriate volumes of the same buffer containing concentrated denaturing agent (urea or guanidinium chloride) were added to initiate the unfolding reaction. In order to study the refolding reaction, selected volumes of free denaturant buffer were added to the initial protein solution at 20 µM containing concentrated denaturing agent. Kaleidagraph program (version 4.0, Synergy Software) was used to fit kinetic constants to the equation of a two-state folding protein. Kinetic and free energy values were determined using the following equations:

$$\Delta G_{U\text{-}F} = -RT \ln (k_u / k_f); \quad m_{U\text{-}F} = RT (m_f + m_u)$$

where *kf* and *ku* are the rates of folding and unfolding, respectively, and the m_f and m_u values represent the slopes of the respective folding and unfolding regions. The

differences in unfolding free energy ($\Delta\Delta G_{F-U}$) between the wild-type protein and mutant (*) are calculated as:

$$\Delta\Delta G_{F-U} (\Delta G_{F-U}) - (\Delta G_{F-U})^* = \Delta\Delta G_{\ddagger-U} - \Delta\Delta G_{\ddagger-F}$$

$$\Delta\Delta G_{\ddagger-U} = RT \ln (k_f / k_f^*) \qquad \Delta\Delta G_{\ddagger-F} = RT \ln (k_u / k_u^*)$$

$$\phi_F = \Delta\Delta G_{\ddagger-U} / \Delta\Delta G_{F-U}$$

ϕ_F value analysis describes the effects of the mutation upon the free energy of the individual native, denatured and transition state.

Sample preparation for aggregational assays

Aggregation of URN1-FF proteins were studied under physiological conditions using sodium acetate buffer at pH 5.7 and under acidic conditions with 100 mM glycine buffer at pH 2.5. Protein samples were prepared in the appropriate buffer at final concentration of 100 µM and incubated at 310 K under agitation during one week.

Aggregation kinetics

To monitor the aggregation kinetics, protein solutions were prepared in the distinct buffer conditions in presence of 25 µM Th-T. The samples were equilibrated at 310 K during 5 minutes and Th-T fluorescence intensity was monitored at the interval of one minute in a Jasco FP8200 spectrofluorometer (Jasco Corporation, Japan). Th-T excitation and emission wavelengths of 440 and 475 nm, respectively, were used with slit widths of 5 nm.

Transmission Electron Microscopy (TEM)

FF aggregates were diluted tenfold with water and 10 µL were placed on carbon-coated copper grids during 5 min. The grids were then washed with distilled water and stained with 2% (w/v) uranyl acetate for 1 min. HITACHI H-7000 transmission electron microscope operating at an accelerating voltage of 75 kV was used to analyze the aggregates.

Th-T and Congo Red (CR) binding

Aggregated samples were diluted to 10 µM in sodium acetate buffer at pH 5.7 in presence of 25 µM Th-T and analysed in a Jasco FP8200 spectrofluorometer (Jasco Corporation, Japan). Three Th-T spectra were accumulated exciting at 440 nm and acquiring fluorescence emission between 460 and 600 nm with excitation and emission slit widths of 5 nm. To study Congo Red (CR) union, URN1-FF aggregates were diluted to 10 µM in presence of 20 µM of CR and optical absorption spectra were recorded from 400 to 700 nm at 298 K. Each trace was the average of 3 accumulated spectra in a Cary-400 UV/Vis spectrophotometer (Varian Inc.). Protein and CR spectra alone were acquired to remove protein scattering and dye contribution.

Circular Dichroism (CD) and bis-ANS binding to study FF aggregates

Aggregated URN1-FF solutions were diluted at 10 µM in sodium acetate buffer at pH 5.7 and analysed by far-UV CD and bis-ANS binding as described above for the soluble proteins.

2.4 Cysteine specific post-translational sulfonation reduces FF domain stability and promotes amyloid fibrils formation

Hydrogen peroxide (H_2O_2) is an oxidizing agent largely used to model ROS effects since it is produced in cellular environments under non-pathogenic conditions (Shacter, 2000). In the present work, we analysed the hydrogen peroxide-induced oxidation effect on the URN1 FF domain stability and aggregation. In particular, it was studied the oxidation impact on the stable FF wild type domain and the FF mutant Arg18Ala (R18A) (Figure IV-2.4.1), which was generated during an alanine scanning analysis and characterized as a destabilized but properly folded FF domain (Castillo, et al; unpublished).

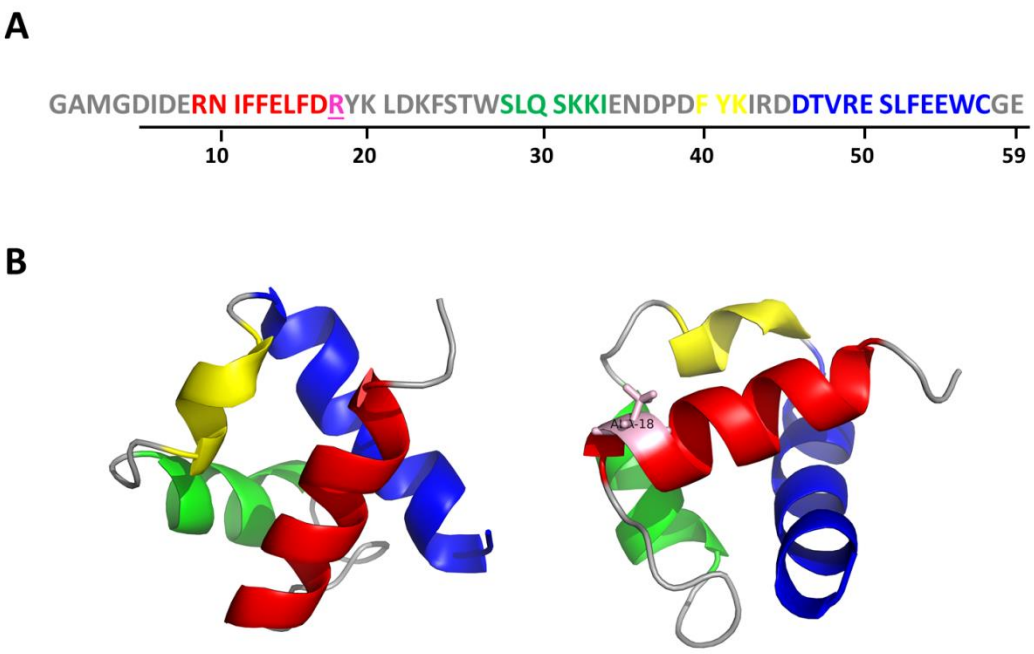

Figure IV-2.4.1|**Amino acidic and 3D structure of the URN1 FF domain.** (A) Amino acidic sequence and (B) ribbon representation of URN1 FF wild type (left) and R18A mutant (right) are shown. Red, green, yellow and blue regions correspond to α-helix 1, α-helix 2, helix 3_{10} and α-helix 3, respectively. The mutated residue (Arg18Ala) is represented in pink color. The Protein Data Bank accession code for the structure is 2JUC.This figure was prepared with PyMOL (www.pymol.org).

URN1 FF Oxidation modifies specifically Cys57 residue forming a cysteine sulfonic acid

FF domains were incubated in presence of H_2O_2 and their oxidation state was analysed by mass spectroscopy. As shown in Figure IV-2.4.2, an increase of ≈ 67 Da in the molecular weight is detectable for both oxidized proteins corresponding to the introduction of 4 oxygen atoms.

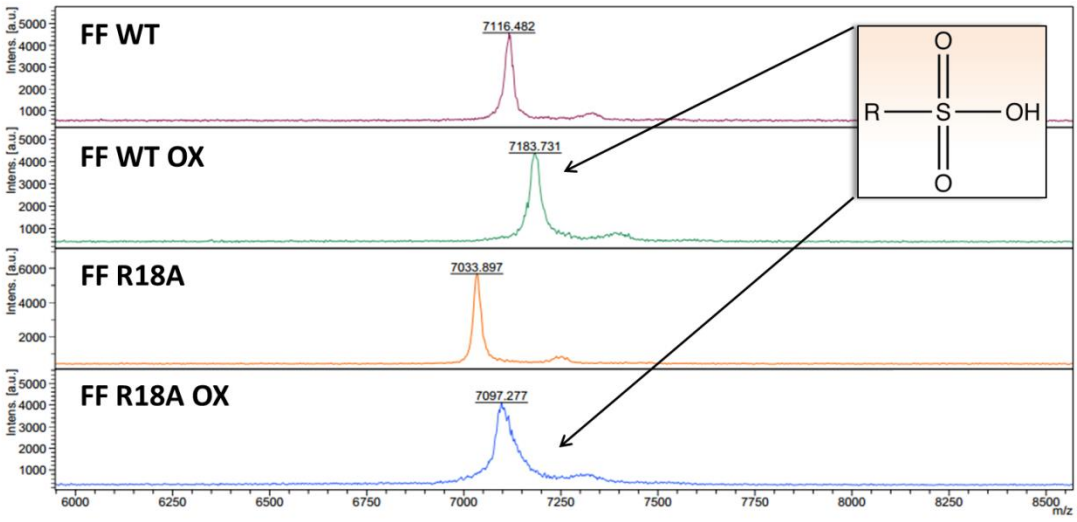

Figure IV-2.4.2| **Mass spectrometry identification of the sulfonation modification in FF domains.** MALDI-TOF spectra of oxidized FF domains and their respective non-oxidized controls are shown. In box the sulfonic group recognized for oxidized species is displayed, R corresponds to Cysteine residue (Cys57).

The further tryptic digestion coupled with MALDI–TOF analysis allowed the identification of two specific oxidation sites at Met3 and Cys57 positions (Table IV-2.4.1). Specifically, a methionine sulfoxide (MSO) accounting for one oxygen atom addition and a sulfonic cysteine (CYS SO_3) acid incorporating 3 oxygens (≈48 Da) are detected. As shown in Table IV-2.4.1, carboxamidomethyl (CAM≈ 57 Da) incorporation is observable only for peptides derived from non-oxidized samples indicating the presence of free thiol groups in their cysteine residues. The molecular mass detected for oxidized peptides was ≈ 9 Da less than the one expected for a CAM Cys adduct indicating the lack of alkylation due to the sulfonic cysteine replacement.
Therefore, the H_2O_2-induced FF oxidation results in the formation of two irreversibly modified specific residues where the sulfonated cysteine is expected to have the main impact on protein structure since the methionine is located in the disordered N-terminal tail and is not involved in the establishment of any specific interaction with the rest of the domain.

Conformational characterization of Sulfonated URN1 FF domains

The conformational properties of the oxidized proteins (FF WT OX and FF R18A OX) were studied at pH 5.7 and 298K by far UV-CD spectroscopy (Figure IV-2.4.3A), intrinsic fluorescence (Figure IV-2.4.3B) and bis-ANS binding (Figure IV-2.4.3C), and compared with those of the non-oxidized FF domains (FF WT and FF R18A). The typical α-helical pattern with two minima at 210 and 222 nm was observed for both non-oxidized protein spectra with a slight decrease in the α-helical signal of the FF R18A domain. The FF WT OX retains the native secondary structure content showing, however, a more accentuated reduction in the signal intensity.

Table IV-2.4.1| **Mass accuracy of FF domain fragments obtained after tryptic digestion and analysed by MALDI-TOF spectroscopy.** The data were processed using the FindPept tool (http://web.expasy.org/findpept/).

	User mass[a] (Da)	DB mass[b] (Da)	Δmass[c] (Da)	peptide	position	Modifications
FF WT	2141.117	2140.960	-0.156	(K)/IRDDTVRESLFEEWCGE	43-59	CYS_CAM*
	3262.709	3262.463	-0.245	(K)/IENDPDFYKIRDDTVRESLFEEWCGE	34-59	CYS_CAM
	3390.818	3390.558	-0.259	(K)/KIENDPDFYKIRDDTVRESLFEEWCGE	33-59	CYS_CAM
	3800.153	3799.879	-0.274	GAMGDIDERNIFFELFDRYKLDKFSTWSLQSK/(K)	2-32	-
FF WT OX	2132.115	2131.924	-0.191	(K)/IRDDTVRESLFEEWCGE	43-59	**CYS SO_3****
	3253.704	3253.427	-0.277	(K)/IENDPDFYKIRDDTVRESLFEEWCGE	34-59	**CYS SO_3**
	3381.817	3381.522	-0.295	(K)/KIENDPDFYKIRDDTVRESLFEEWCGE	33-59	**CYS SO_3**
	3816.200	3815.874	-0.326	GAMGDIDERNIFFELFDRYKLDKFSTWSLQSK/(K)	2-32	MSO***
FF R18A	2141.182	2140.960	-0.221	(K)/IRDDTVRESLFEEWCGE	43-59	CYS_CAM
	3262.792	3262.463	-0.328	(K)/IENDPDFYKIRDDTVRESLFEEWCGE	34-59	CYS_CAM
	3390.901	3390.558	-0.342	(K)/KIENDPDFYKIRDDTVRESLFEEWCGE	33-59	CYS_CAM
	3715.173	3714.815	-0.358	GAMGDIDERNIFFELFDAYKLDKFSTWSLQSK/(K)	2-32	-
FF R18A OX	2132.146	2131.924	-0.222	(K)/IRDDTVRESLFEEWCGE	43-59	**CYS SO_3**
	3253.753	3253.427	-0.326	(K)/IENDPDFYKIRDDTVRESLFEEWCGE	34-59	**CYS SO_3**
	3381.864	3381.522	-0.342	(K)/KIENDPDFYKIRDDTVRESLFEEWCGE	33-59	**CYS SO_3**
	3731.187	3730.810	-0.377	GAMGDIDERNIFFELFDAYKLDKFSTWSLQSK/(K)	2-32	MSO

[a] The peptide mass signal observed experimentally.

[b] The mass of the peptide calculated using the masses database (DB).

[c] The mass difference between the user mass and DB mass.

* Carboxyamidomethyl cys resulting by the artefactual modification through iodocaetamide (57.021Da).

** Sulfonic cysteine derived from the oxidation of cysteine (47.985 Da) (Figure 2).

***Methionine sulfoxide proceeding by the oxidation of the residue (15.995 Da).

FF WT, FF WT OX, FF R18A and FF R18A OX are used to indicate the wild type domain, the oxidized wild type domain, the mutant Arg18Ala and its oxidized form, respectively.

In contrast, the dramatic left-shift observed for the FF R18A OX spectrum reveals the conversion of the native α-helical structure towards a more disordered secondary structure with a minimum at ≈205 nm. In agreement with these data, the FF WT OX emission spectrum shows higher fluorescence emission intensity compared with that of the FF WT suggesting a larger exposure of Trp residues. The spectrum of FF R18A OX

is significantly red-shifted and higher than that of the non-oxidized mutant, consistent with the observed changes in secondary structure and likely associated with the opening of the globular structure and the concomitant exposure of hydrophobic regions, as indicated by the considerable enhancement of bis-ANS fluorescence. The analysis of bis-ANS binding also indicates conformational changes occurring in FF WT OX, with a significant increase of its emission when compared with wild type domain. In agreement with these data, the one-dimensional NMR (^{1}H-NMR) analysis at 298 K (Figure IV-2.4.5) of the oxidized FF domains shows, in both cases, a clear alteration of the native signal dispersion and peak sharpness indicating structural rearrangements relative to the parental proteins.

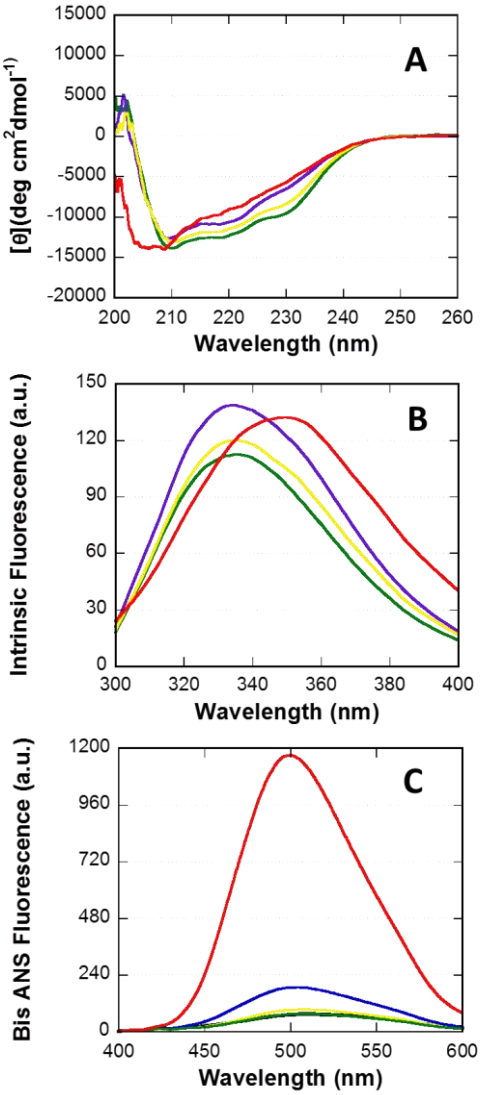

Figure IV-2.4.3|**Structural and conformational properties of oxidized and non-oxidized FF domains**. URN1 FF proteins were measured by (A) far-UV CD spectroscopy, (B) tryptophan intrinsic fluorescence and (C) bis ANS fluorescence, at 298 K. Free Bis ANS emission spectra is represented with a black continued line. Oxidized species are shown in blue and red for wild

type and R18A mutant while non-oxidized controls are represented in green and yellow, respectively.

Sulfonation effect upon FF domain stability

Thermal unfolding of the URN1 FF domains was followed at pH 5.7 by circular dichroism (CD) and tryptophan intrinsic fluorescence monitoring the molar ellipticity and emission changes at 222 and 360 nm, respectively (Figure IV-2.4.4 A,B). The obtained thermal denaturation curves show a single cooperative transition and the data could be fitted to a two-state unfolding model for all analysed FF domains except for FF R18A OX, due to its excessively low thermostability. The stability of both FF domains results considerably perturbed by the oxidative modifications as indicated by the decrease in the FF WT OX T_m value ≈ 20 K respect that of the wild type domain and a reduction of more than 20 K in the case of the R18A OX T_m, relative to the non-oxidized R18A mutant (Table IV-2.4.2).

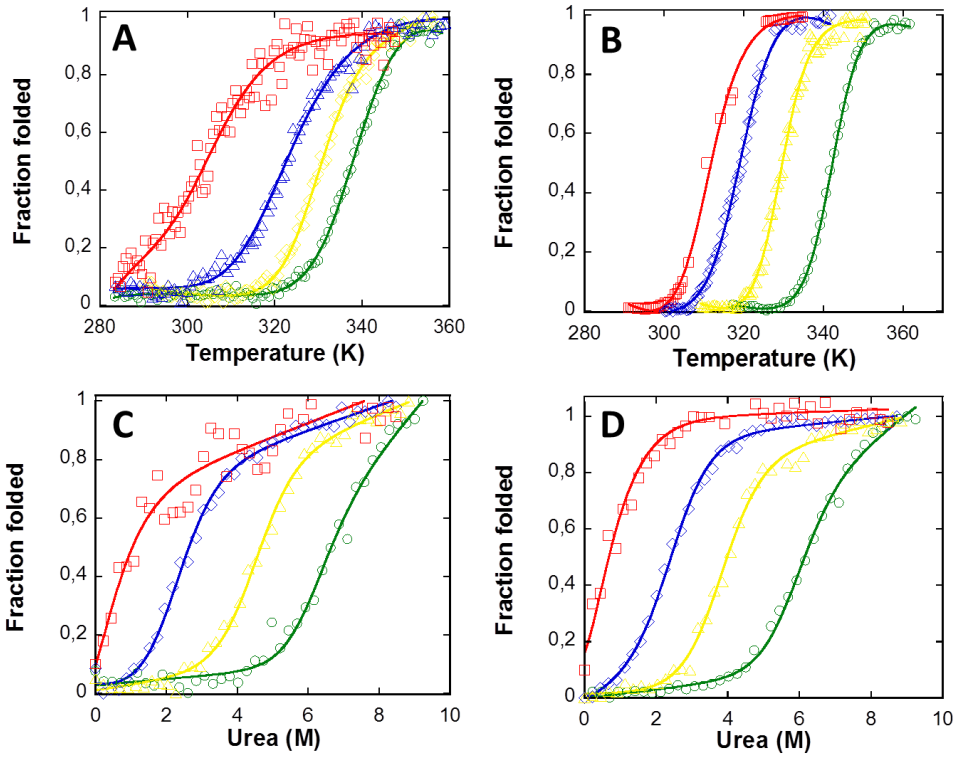

Figure IV-2.4.4| **Thermal and chemical stabilities of FF domains.** Thermal stabilities were analysed by (A) far-UV CD signal change at 222 nm and (B) intrinsic fluorescence emission at 350 nm. Chemical equilibrium curves with urea were followed at 298 K by (C) far-UV CD at 222 nm and by (D) tryptophan intrinsic fluorescence at 360 nm. Oxidized FF wild type and R18A mutant are shown in blue and red, while non-oxidized FF controls are represented in green and yellow, respectively.

These data were also corroborated by the thermal denaturation followed by ^{1}H-NMR. The main loss of the native signal dispersion is observed at 317 K for the FF WT OX and

at 341 K for the FF WT spectrum, while the FF R18A and FF R18A OX spectra display a decrease in the peak intensities at 329 and 308 K, respectively (Figure IV-2.4.5). Interestingly, all FF domains show reversible thermal denaturation behaviour as indicated by the complete recuperation of the native spectrum at 298 K after cooling down the sample.

Table IV-2.4.2| **Thermal denaturation parameters of FF domains**

	Tm(K)[a]	**Tm(K)[b]**
FF WT	340.9 ± 0.3	342.6 ± 0.1
FF R18A	335.5 ± 0.5	330.0 ± 0.3
FF WT OX	323.9 ± 0.6	318.0 ± 0.2
FF R18A OX	–	–

[a] CD
[b] Intrinsic Fluorescence

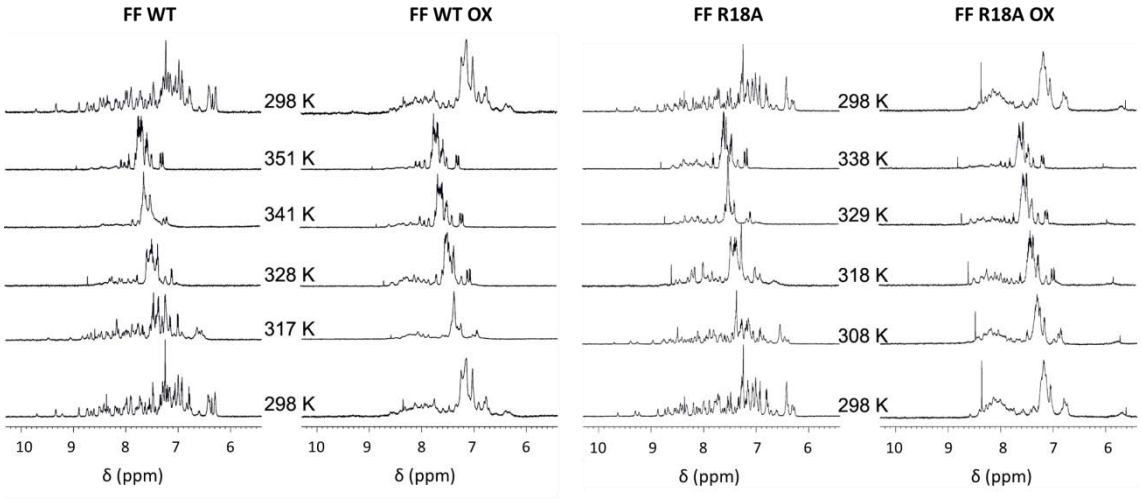

Figure IV-2.4.5| **Thermal stability of URN1-FF followed by one-dimensional NMR (^{1}H-NMR) spectroscopy.** FF wild type and R18A mutant were followed by ^{1}H-NMR at different indicated temperatures. 50 µM of oxidized and non-oxidized forms were analysed at 600 MHz.

The chemical unfolding of the FF domains was analysed at pH 5.7 and 298 K. The urea denaturation curves at equilibrium were obtained recording the changes in molar ellipticity at 222 nm (Figure IV-2.4.4C) and intrinsic fluorescence at 360 nm (Figure IV-2.4.4D) at increasing denaturant concentrations. The thermodynamic values were calculated assuming a two-state model according to the unfolding cooperativity shown by all FF domains (Table IV-2.4.3). The stability of FF WT OX measured by fluorescence

and CD spectroscopy corresponds to ≈ 1.5 kcal/mol with a $[GuHCl]_{50\%}$ of ≈ 2 M, while reliable thermodynamic values could not be approximated for FF R18A OX since it unfolds at very low urea concentrations. Therefore, in agreement with the thermal denaturation data, both oxidized species result to be significantly destabilized relative to non-oxidized FF domains.

Table IV-2.4.3| **Thermodynamic properties of equilibrium unfolding of oxidized FF domains**

	[a] $\Delta G^{(H20)}_{U\text{-}F}$ (Kcal/mol)		[b] m (Kcal/mol·M)		[c] [Urea]50% (M)	
	CD	**Intrinsic Fluorescence**	**CD**	**Intrinsic Fluorescence**	**CD**	**Intrinsic Fluorescence**
FF WT	5.3 ± 0.3	5.9 ± 0.2	0.79 ± 0.04	0.88 ± 0.03	6.8 ± 0.6	6.8 ± 0.3
FF R18A	3.17 ± 0.24	4.82± 0.56	0.71 ± 0.07	1.43 ± 0.18	4.46 ± 0.09	3.82 ± 0.06
FF WT OX	1.00 ± 0.28	1.66 ± 0.17	0.68 ± 0.05	0.79 ± 0.03	1.47 ± 0.28	2.09 ± 0.11
FF R18A OX	–	–	–	–	–	–

[a] Gibbs energy of unfolding at [Urea]=0

[b] m value, dependence of free energy of unfolding with urea

[c] The urea concentration required to unfold 50% of the protein molecules

Dissecting the role of cysteine oxidative modification

The data reported above demonstrate that the H_2O_2 induced oxidation causes important structural and conformational changes in both analysed FF proteins, which lead to a serious destabilization of the globular domains. This action is caused by the presence of a sulfonic group in cysteine at position 57 (Figure IV-2.4.6 A,B). However, it is not clear if the observed FF destabilization results from the generic loss of the wild type cysteine residue at this specific position or rather it is caused by the specific introduction of the sulfonic group in this position. Thus, in order to understand the molecular mechanism responsible for FF domain destabilization, the wild type cys57 amino acid was mutated to mimic the cysteine change in this position. On the one hand, a Cys57Ser (C57S) mutant was designed to investigate selectively the effect caused by the loss of the thiol reactive group in the cysteine side chain, since serine shares a very similar structure with cysteine differing only in the presence of an oxygen instead of a sulphur atom (Figure IV-2.4.6 A, C). On the other hand, the cysteine residue was replaced with aspartic acid to obtain the Cys57Asp (C57D) mutant where the aspartic amino acid may act as a "sulfonic cysteine mimic", emulating the negative charge of the oxidized cysteine side chain and its spatial position (Figure IV-2.4.6 B, D). The changes in thermodynamic stability promoted by the mutations were predicted with the FoldX algorithm (Schymkowitz et al., 2005a) ($\Delta\Delta G_{total} = \Delta G_{mut} - \Delta G_{wt}$). In

agreement with the properties expected for the mutants, ΔΔG values of -0.46 and 5.29 kcal/mol were predicted for the FF C57S and FF C57D mutant, respectively.

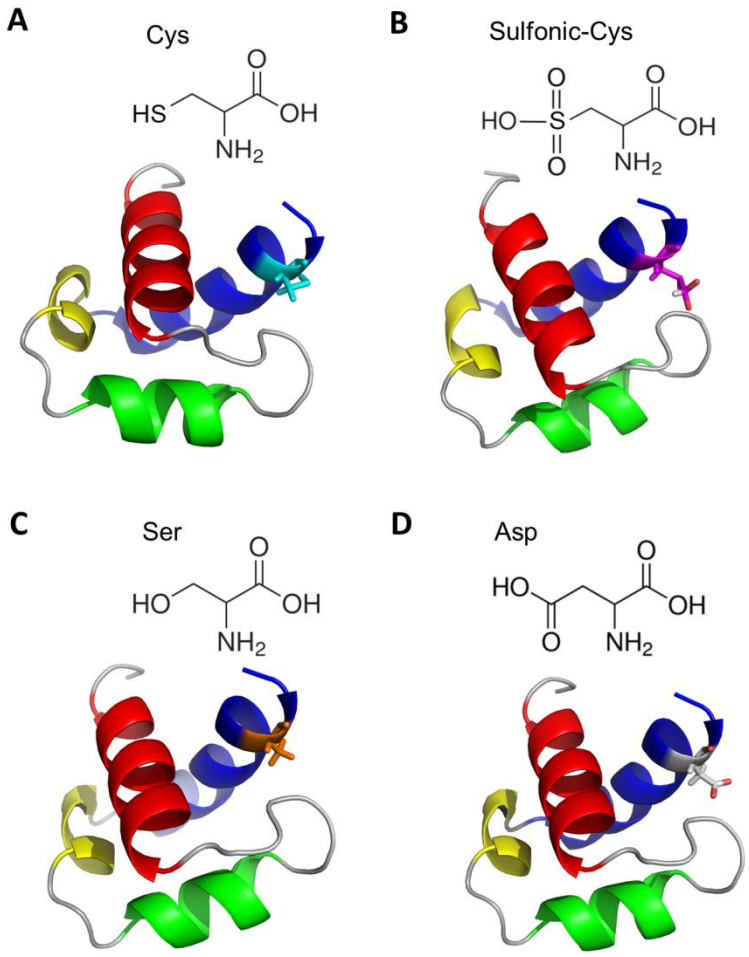

Figure IV-2.4.6| **3D structures of FF domain and its variants.** Ribbon representation of (A) FF wild type showing cys57 residue (cyan) and (B) its oxidized species with sulfonic-cys57 in magenta are shown. (C) URN1 FF C57S and (D) URN1 FF C57D mutants are represented, too. The Protein Data Bank accession code for the structures is 2JUC. The figure was prepared with PyMOL (www.pymol.org).

Study of conformational properties of URN1 FF cysteine mutants

The secondary and tertiary structure of FF Cys57Ser (FF C57S) and Cys57Asp (FF C57D) mutants were analysed by far UV-CD spectroscopy (Figure IV-2.4.7A), intrinsic fluorescence (Figure IV-2.4.7B) and bis-ANS binding (Figure IV-2.4.7C). The CD spectra of both FF mutants display a classical α-helical structure with signal intensities very similar to that of the wild type spectrum especially in the case of the FF C57S, while FF C57D mutant shows a slight decrease in the signal, resembling the sulfonated FF spectrum. The increase in intrinsic fluorescence emission recorded for the FF C57D mutant suggests a certain opening of the globular structure as in the case of the sulfonated FF domain. In contrast, the FF C57S tryptophan emission spectrum

resembles that of the wild type. The intense fluorescence emission registered for the FF C57D in the presence of bis-ANS confirms the above mentioned conformational changes and the exposure of hydrophobic regions as for the sulfonated species. As expected, the FF C57S mutant does not show significant bis-ANS binding with a fluorescence intensity equivalent to that of the wild type domain.

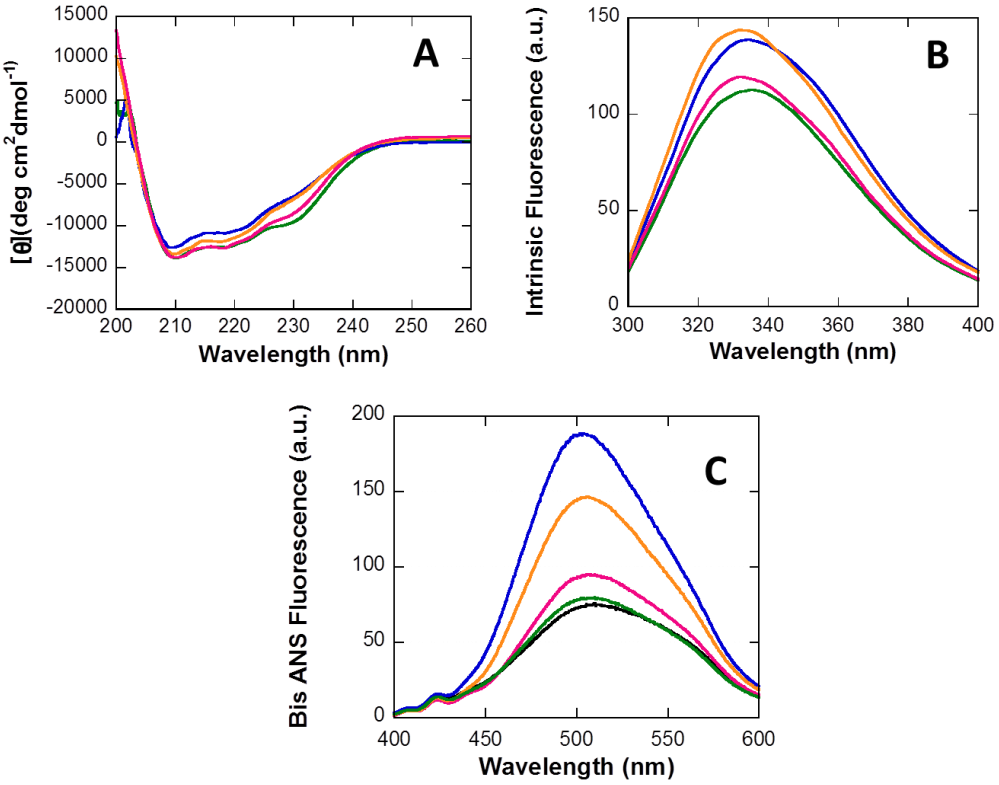

Figure IV-2.4.7| **Structural and conformational properties of sulfonated FF domain compared to wild type and its FF cysteine mutants**. URN1 FF protein solutions were measured by (A) far-UV CD spectroscopy, (B) tryptophan intrinsic fluorescence and (C) bis-ANS fluorescence, at 298 K. Free Bis-ANS fluorescence spectrum is represented with a black continued line. The wild type FF domain and the sulfonated FF protein are shown in green and blue, instead FF C57S and FF C57D are represented in pink and orange, respectively.

Thermal and chemical unfolding of FF cysteine mutants

The thermal and chemical stability of FF mutants were analysed at pH 5.7. The resulting data were compared with thermodynamic values obtained for wild type and sulfonated domains (Figure IV-2.4.8). Thermal and chemical denaturation curves were monitored by far-UV CD at 222 nm and by tryptophan fluorescence emission at 360 nm displaying, in both FF mutants, a single cooperative transition that can be fitted to a two-state unfolding model. As shown in Table IV-2.4.4, the melting temperatures obtained by CD and intrinsic fluorescence are very similar, coinciding to indicate that FF C57D displays a thermal stability close to that of the sulfonated FF, while the FF C57S thermostability seems to approach that of the wild type domain. An analogous

behaviour was observed with chemical unfolding of the FF mutants, showing for the FF C57D mutant a stability of ≈2 kcal/mol and $[GuHCl]_{50\%}$ of ≈ 3 M and, for the FF C57S mutant, a stability of ≈4 kcal/mol and $[GuHCl]_{50\%}$ of ≈ 6 M. Thus, taken together, these data indicate a similar destabilization for FF C57D and sulfonated species when compared to FF WT, whereas the FF C57S mutant, despite displaying a certain degree of destabilization, approaches the thermodynamic stability of the wild type domain.

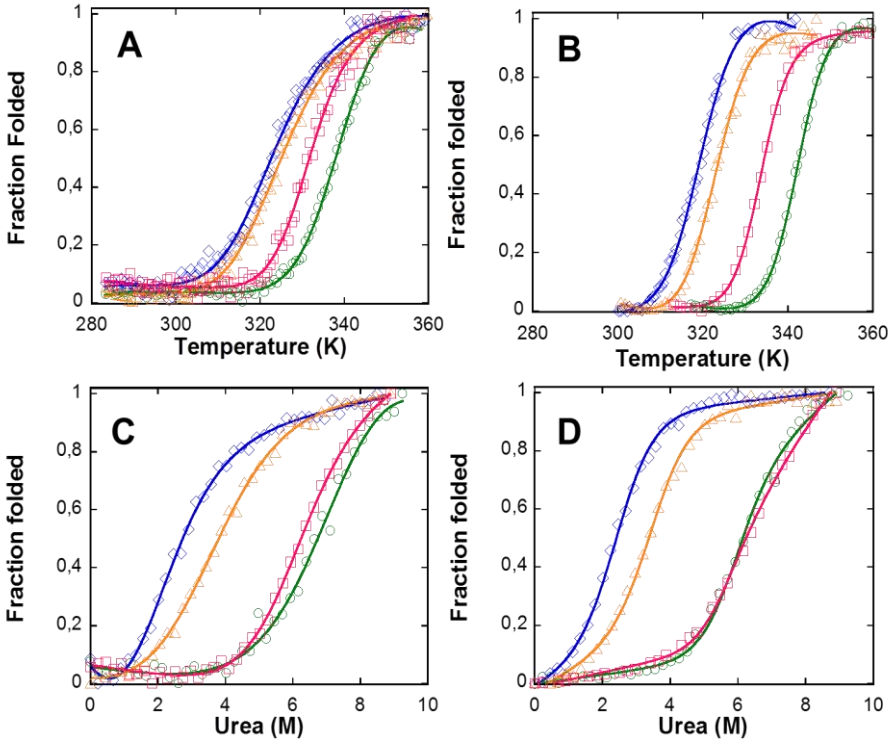

Figure IV-2.4.8| **Thermal and chemical stabilities of FF cysteine mutants.** Thermal stabilities were measured by (A) far-UV CD signal change at 222 nm and (B) intrinsic fluorescence emission at 350 nm. Chemical equilibrium curves with urea were monitored at 298 K by (C) far-UV CD at 222 nm and by (D) tryptophan intrinsic fluorescence at 360 nm. Sulfonated FF protein and wild type domain are shown in blue and green, while FF mutants C57S and C57D are represented in pink and orange, respectively.

Table IV-2.4.4| **Thermal denaturation parameters of FF mutants**

	Tm(K)[a]	Tm(K)[b]
FF WT	340.9 ± 0.3	342.6 ± 0.1
FF WT OX	323.9 ± 0.6	318.0 ± 0.2
FF C57D	324.8 ±0.3	323.6 ± 0.3
FF C57S	331.8 ± 0.2	333.7 ± 0.1

[a] CD
[b] Intrinsic Fluorescence

Table IV-2.4.5| **Thermodynamic properties of equilibrium unfolding of FF mutants**

	[a] $\Delta G^{(H20)}_{U-F}$ (Kcal/mol)		[b] m (Kcal/mol·M)		[c] [Urea]50% (M)	
	CD	**Intrinsic Fluorescence**	**CD**	**Intrinsic Fluorescence**	**CD**	**Intrinsic Fluorescence**
FF WT	5.3 ± 0.3	5.9 ± 0.2	0.79 ± 0.04	0.88 ± 0.03	6.8 ± 0.6	6.8 ± 0.3
FF WT OX	1.00 ± 0.28	1.66 ± 0.17	0.68 ± 0.05	0.79 ± 0.03	1.47 ± 0.28	2.09 ± 0.11
FF C57D	1.87 ± 0.31	2.18 ± 0.19	0.55 ± 0.06	1.00 ± 0.04	3.38 ± 0.18	3.19± 0.09
FF C57S	4.09 ± 0.18	3.94 ± 0.08	0.69 ± 0.04	0.62 ± 0.02	5.91 ± 0.15	6.26 ± 0.13

[a] Gibbs energy of unfolding at [Urea]=0

[b] m value, dependence of free energy of unfolding with urea

[c] The urea concentration required to unfold 50% of the protein molecules

Molecular dynamics (MD) simulations

To describe in atomistic detail the structural stability of WT and FF variants, molecular dynamics (MD) simulations have been carried out in collaboration with Elena Papaleo (Danish Cancer Society Research Center, Copenhagen, Denmark). We performed 1 microsecond of conventional MD for FF C57S, C57D and the WT variant. We could observe a slight destabilization induced by the C57D mutation, for example attested by an increase of fluctuations in the values of radius of gyration (Rg) and root mean square deviation of the main chain respect to the experimental WT structure (data not shown). Those simulations were hard to evaluate, especially if we aim at quantifying the differences between the different conformational substates of the protein in solution. We thus turned our attention to another MD-based technique, using an enhanced sampling approach based on metadynamics that has been demonstrated very accurate in describing the effects of point mutations on protein conformation (Papaleo et al., 2014). Metadynamics is a method that has been developed to sample rare conformational transitions in proteins and to estimate with high accuracy the associated free energy changes while keeping an atomistic description of both the solvent and the protein. It is however a computationally expensive approach, which requires more than 128 cores in parallel for a protein of the size of an FF domain. Metadynamics simulations have been carried out with a state-of-the-art force field (CHARMM22*) in explicit solvent for more than 400 ns and using as reaction coordinates (collective variables) properties that account for the overall structures of the FF domain, i.e: the radius of gyration and the helical content in the three α-helices of the FF fold. In agreement with the experimental data, the C57D mutation has destabilizing effects in the metadynamics simulations. Indeed, we can observe that the mutation induces the appearance of minima with larger Rg population in the free

energy landscape of the protein (Figure IV-2.4.9A), which also correlates with the destabilization of helix 1 (Figure IV-2.4.9B) and helix 3 (Figure IV-2.4.9D) for which a lower helical content is observed. Helix 2 seems not to be affected by the mutations and shows comparable free energy profiles for wt and mutant variant (Figure IV-2.4.9C).

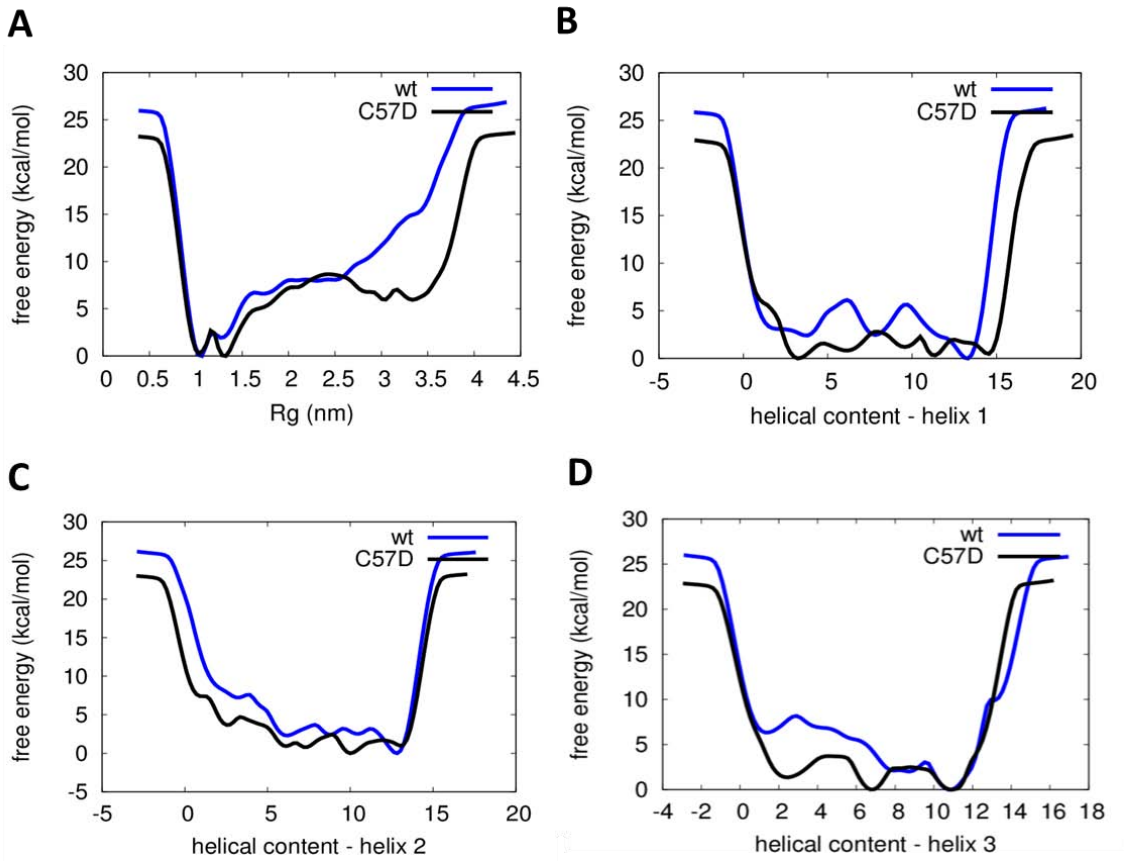

Figure IV-2.4.9| **Metadynamics simulation of FF domains.** (A) Rg populations in the free energy landscape of the FF WT and FF C57D and (B, C, D) the mutation impact on helical content. FF WT and FF C57D proteins are shown with blue and black line, respectively.

Folding and unfolding kinetics of URN1 FF

The kinetics of folding and unfolding of the FF cysteine mutants, FF sulfonated and FF wild type domain, were analysed by stopped-flow at pH 5.7 and 298 K under a wide range of denaturant conditions. In all cases, the folding and unfolding traces by fluorescence fit well into single exponential functions. The chevron plots appear to be linear in the complete range of denaturant concentrations studied, indicating the lack of detectable intermediates, according to a two-state model (Figure IV-2.4.10). The rate constants for folding (K_f) and unfolding (K_u) and their dependence on the denaturant concentration (m_f and m_u) are shown in Table IV-2.4.6. The kinetic data of FF domains are in good agreement with the equilibrium stabilities reported above. The

FF C57S mutant shows a folding rate very similar to that of FF WT, displaying, however, a faster unfolding rate. FF sulfonated and FF C57D domains fold ≈3 times slower than FF WT and both domains unfold very fast, compared to the other analysed FF domains. The ϕ_{F-U} values of 0.13, 0.19 and 0.09 calculated for the sulfonated FF, the FF C57D and FF C57S mutants, respectively, suggest that the cysteine57 residue is located in a region that does not participate in the folding nucleus. Interestingly, the drastic increase of the sulfonated FF and FF C57D unfolding rates confirms the specific effect of cysteine modification on the native state conformational stability.

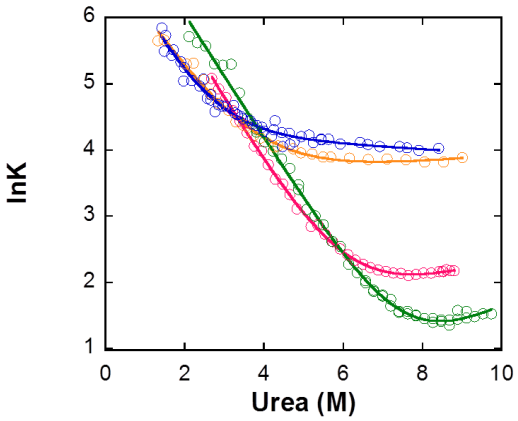

Figure IV-2.4.10| **Folding and unfolding kinetics of FF domains.** The kinetics of folding and unfolding were followed by tryptophan intrinsic fluorescence performing stopped-flow experiments at 298 K. URN1 FF domain wild type (green), sulfonated FF protein (blue), FF C57S (pink) and FFC57D (orange) mutant are shown. The rate constants were measured under conditions of apparent two-state folding.

Table IV-2.4.6| **Folding kinetic parameters for FF mutants**

	K_f (s^{-1})	K_u (s^{-1})	m_f (Kcal/mol·M)	m_u (Kcal/mol·M)	[a] m_{U-F} (Kcal/mol·M)	[b] $[\,]_{50\%}$(M)	[c] ΔG_{U-F} (Kcal/mol)	[d] $\Delta\Delta G_{U-F}$ (Kcal/mol)
FF WT	2819 ± 198	0.34± 0.12	0.94 ± 0.02	0.26 ± 0.04	0.71 ± 0.06	7.48	5.34	–
FF WT OX	1215 ± 198	73.60 ± 9.8	1.16 ± 0.1	0.03 ± 0.02	0.71 ± 0.12	2.33	1.66	- 3.68
FF C57D	954 ± 97	31.9 ± 0.6	0.88 ± 0.05	0.04 ± 0.02	0.55 ± 0.08	3.66	2.01	-3.33
FF C57S	2297 ± 122	0.99 ± 0.01	2.34 ± 0.36	0.14 ± 0.02	0.67 ± 0.03	6.06	4.08	-1.26

[a] Dependence of the Gibbs energy of unfolding with urea
[b] The urea concentration required to unfold 50% of the protein molecules
[c] Gibbs energy of unfolding with urea determined from the kinetic parameters
[d] The difference in free energy between the mutant protein and WT protein

Amyloid fibril formation by URN1 FF domains under physiological conditions

The aggregation kinetics of FF domains were analysed at pH 5.7 and 310 K following the changes over time in the fluorescence of the amyloid staining dye Thioflavine T (Th-T) (Figure IV-2.4.11). FF WT OX and the FF C57D aggregation rates are very fast displaying, Th-T fluorescence immediately after mixing. In particular, FF R18A OX reachess very fast the stationary phase of the aggregation process. A similar behaviour is exhibited by the FF WT OX and FF C57D domains which also lack a lag phase , despite requiring longer times to attain saturation of the aggregation reaction. As expected, no Th-T binding is observed for the wild type domain, FF R18A and FF C57S mutants. Therefore, in agreement with their conformational properties and stability, only the oxidized FF domains and the sulfonated mimic FF mutant experiment aggregation under native conditions illustrating how the post-translational oxidation of a single residue can promote dramatic conformation changes that ultimately might result in the formation of protein aggregates.

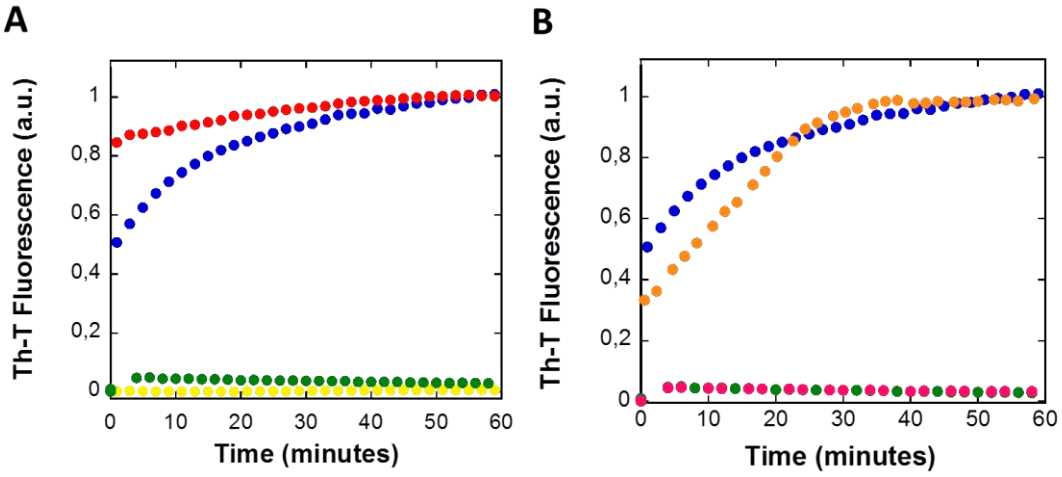

Figure IV-2.4.11| **Aggregation kinetics of FF domains under physiological conditions.** Aggregation kinetics of oxidized species (A) compared with aggregation kinetics of (B) FF cysteine mutants are represented. Aggregation processes were followed by the change in Th-T fluorescence emission at 475 nm. FF protein wild type (green), oxidized FF protein wild type (blue), R18A mutant (yellow), oxidized R18A mutant (red), FF C57S (pink) and FF C57D (orange) mutants are shown.

Thereafter, all FF domains were incubated at pH 5.7 and 310 K under agitation during one week in order to characterize the structural, tinctorial and morphological properties of the FF aggregates (Figure IV-2.4.12). Mature aggregates of both oxidized FF domains show intense Th-T binding, FF WT OX aggregates displaying higher emission intensity than those of FF R18A OX. After 7 days of incubation, the wild type domain does not show any Th-T binding, while the incubated unstable FF R18A solution shows dye binding, although displaying lower fluorescence intensity than that of the oxidized FF domains (Figure IV-2.4.12A).

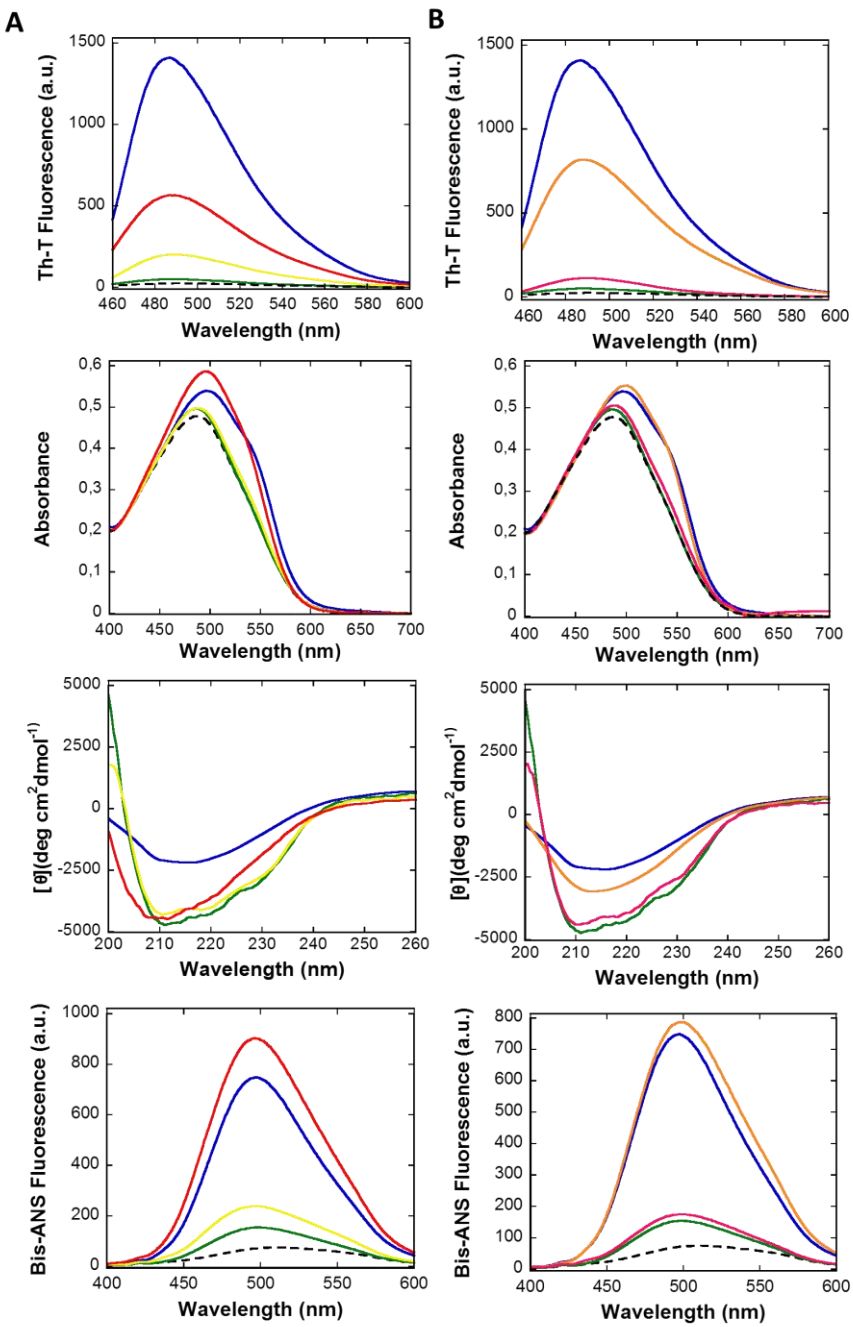

Figure IV-2.4.12| **Structural and tinctorial properties of URN1-FF aggregates at pH 5.7.** (A) Oxidized protein aggregates and (B) aggregates formed for FF mutants were analysed measuring Th-T fluorescence, Congo Red union, far-UV CD signal and recording the binding to Bis-ANS dye (from top to bottom). Free dyes are indicated with a dotted line. In all the cases, protein aggregates from FF protein wild type, oxidized FF protein wild type, R18A mutant, oxidized R18A mutant, FF C57S and FF C57D mutants are represented with a green, blue, yellow, red, pink and orange line, respectively.

As in the case of FF WT, the FF C57S mutant does not bind the amyloid dye. Conversely, FF C57D aggregates bind to Th-T (Figure IV-2.4.12B). In agreement with the Th-T binding data, the characteristic Congo Red (CR) signal shift at ≈ 540 nm upon

amyloid binding is observed for oxidized FF samples, while no CR binding is detected for non-oxidized FF WT and FF R18A solutions. The FF C57D binds CR dye promoting the same spectral changes that FF WT OX. As expected, FFC57S does not show CR binding, as in the case of the non-oxidized FF WT domain. The secondary protein structure in incubated solution was monitored by far-UV CD spectroscopy. The presence of β-sheet conformations was confirmed for FF WT OX as indicated by the typical minimum at ≈218 nm. FF R18A OX also exhibits a CD spectrum compatible with the presence of s β-sheet structures. The absence of structural changes for the non-oxidized FF proteins is confirmed by the observation of α-helical native spectra.

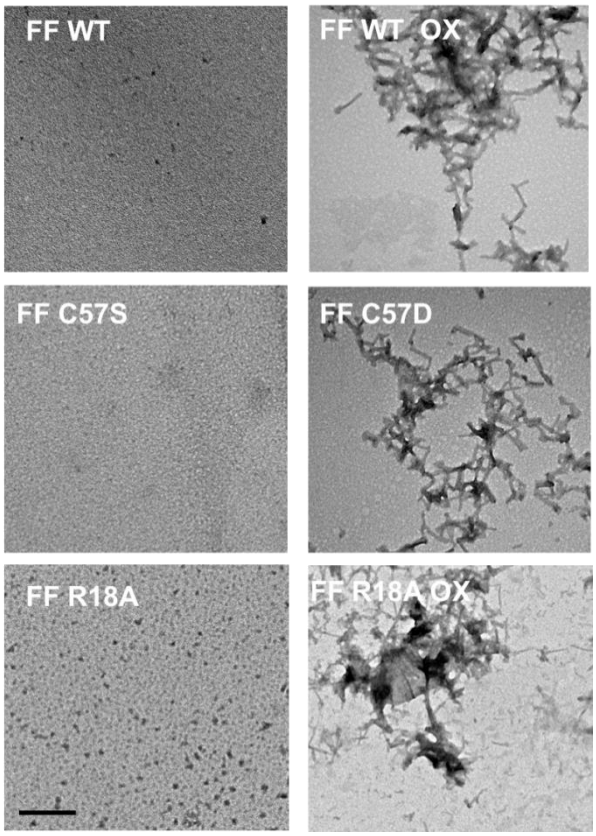

Figure IV-2.4.13| **Morphological properties of URN1-FF aggregates.** Representative micrographs of negatively stained URN1-FF aggregates incubated at 310 K for one week visualized by TEM The scale bar represents 500 nm.

The FF C57S also retains the native secondary structure whereas FF C57D spectrum is dominated by β-sheet component, resembling the FF WT OX spectrum. The bis-ANS binding to exposed hydrophobic regions is much higher in oxidized samples than in non-oxidized FF domains. Again, in good agreement with their structural properties, FF C57D displays a bis-ANS binding very similar to that of FF WT OX, while the FF C57S resembles the wild type domain. TEM observation was used to assess the morphological properties of the FF aggregates (Figure IV-2.4.13). According to conformational and tinctorial features described above, both FF WT OX and FF C57D form short amyloid-like fibrils, while FF R18A OX displays fibrillar structures combined

with abundant amorphous aggregates in agreement with its high aggregation kinetics. On the other hand, scarce smaller size aggregates are detected for R18A, while no recognizable aggregates are observed for FF WT and FF C57S.

In order to test whether the different structural states acquired by the mutants impact their cytotoxicity, incubated FF solutions were added to neuroblastoma cells cultures from the SH-SY5Y cell line and analysed by the EZ4U toxicity assay (Figure IV-2.4.14). The toxicity of protein aggregates is associated to their conformational properties. In this way, it was reported for various, and unrelated, proteins that binding to ANS-like dyes is strictly correlated with the toxicity of amyloid species indicating the exposure of hydrophobic regions as a critical characteristic of pathogenic assemblies (Bolognesi et al., 2010).

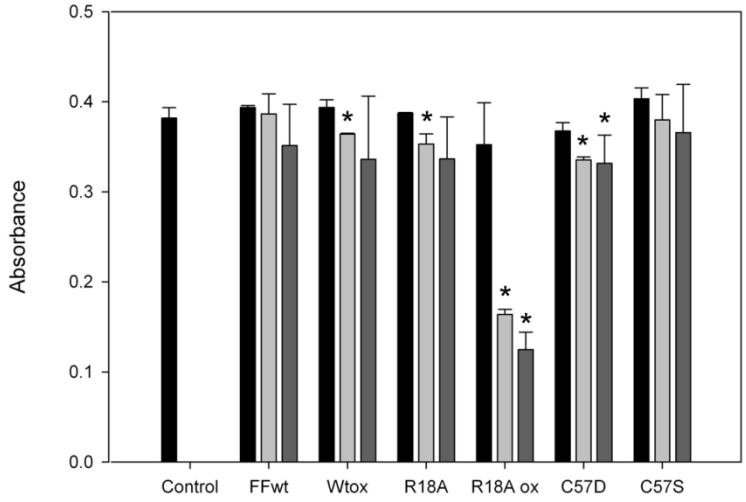

Figure IV-2.4.14| **Toxicity of extracellularly added FF aggregates on neuroblastoma (SH-SY5Y) cell cultures.** Cells were incubated for 72h with 10 (black), 20 (light grey) and 40 (dark grey) µM of FF aggregates after which cell viability was measured. Error bars indicate ± SE (n= 3). The asterisk indicates a $p < 0.05$.

Among all examined FF solutions, only the fibrils formed by FF R18A OX exhibited considerable cell toxicity, in agreement with their higher exposure of hydrophobic regions compared to other FF fibrils.

Taken together, these data demonstrated that, under physiological conditions, the FF sulfonation modification promotes the formation of amyloid-like aggregates in both analysed FF domains and, remarkably, the fibrils formed by the most destabilized domain (FF R18AOX) result to be toxic to neuronal cells, underlining the importance of this post translational modification for *in vivo* proteostasis.

Prediction of sulfonation effect in free solvent exposed cysteines

The data reported above indicate that the attachment of a sulfonic group to the cysteine residue in the FF domain causes conformational and structural changes that lead to the destabilization and aggregation of this small globular protein. Interestingly,

this effect can be emulated by the cysteine replacement with an aspartic amino acid, which acts as a sulfonated cysteine mimic, while the substitution of cysteine with a serine residue, as expected, does not recapitulate the sulfonation effect confirming the specificity of sulfonation-induced changes. In attempt to understand if the particular sulfonation destabilizing effect observed for the FF protein model may be a generic effectin natural proteins, a large set of different and unrelated proteins was analysed with the FoldX tool (Schymkowitz et al., 2005a). In this analysis, it was utilized a cysteine database containing a total number of 969 non-homologous proteins with 4136 cysteine-containing segments, 1446 of which are in the disulfide-bonded state and 2690 in the non-disulfide-bonded state (Martelli et al., 2002). All cysteines predicted to be solvent exposed were selected among the non-disulfide-bonded cysteines and were mutated into aspartic and serine residues with the FoldX modelling tool in order to mimic virtually the sulfonation effect.

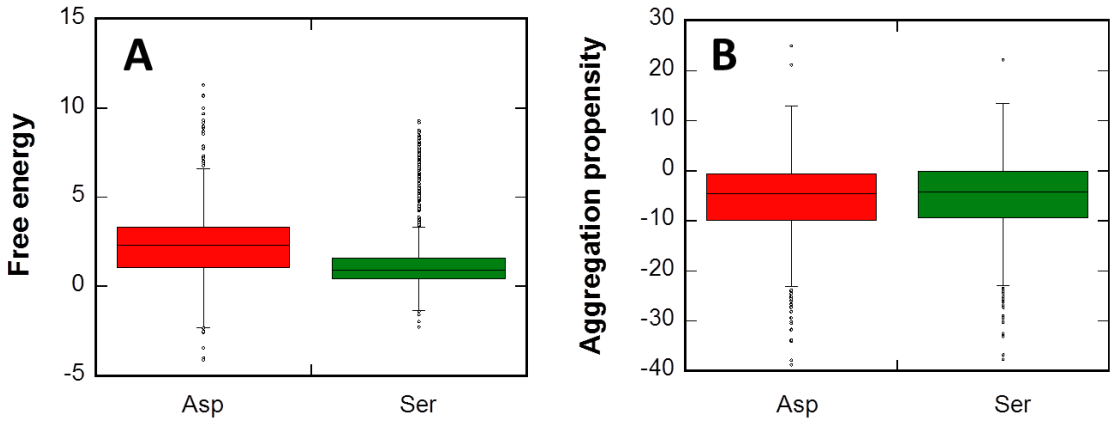

Figure IV-2.4.15| **Stability and aggregation prediction in a large set of proteins with solvent exposed cysteine residues.** (A) Solvent exposed cysteines from cys database (Martelli et al., 2002) were mutated in Aspartic and Serine residue by FoldX tool (Schymkowitz et al., 2005a) to predict the overall change in stability. (B) AGGRESCAN algorithm (Conchillo-Solé et al., 2007) was used to calculate the aggregation propensity of all mutants previously analyzed by FoldX program.

As discussed previously, FoldX is a well-validated algorithm used for predicting the impact of point mutations on protein stability by calculating the energetic change in native thermodynamic stability ($\Delta\Delta G_{total}= \Delta G_{mut} - \Delta G_{wt}$) (Schymkowitz et al., 2005a; Rakoczy et al., 2011). The analysis of all examined mutants reveals a clearly destabilizing effect of Asp mutation corresponding to a ΔΔG median value of 2.28 ± 0.05 kcal/mol (Figure IV-2.4.15A). On the other side, cysteine to serine mutations are predicted to impact much less stability as indicated by the ΔΔG median value of 0.92 ± 0.04. These results confirm that the sulfonation destabilizing effect is not restricted to the FF protein model but it could occur in a wide number of unlinked and physiologically relevant proteins, such as the human β-interferon (IFN-β) involved in the immune system regulation (Kasper & Reder, 2014) and the human apoptosis

regulator Bcl-xL (Hockenbery et al., 1990) showing a predicted destabilization for the Asp mutation corresponding to ΔΔG values of 4.57 and 6.38 kcal/mol, respectively.
In the previous section, it was demonstrated that the introduction of a sulfonic group to cys57 and, in a similar manner, the mutation of cys57 to aspartic residue can promote the FF aggregation with the subsequent formation of amyloid-like fibrils. Sequence-based aggregation prediction methods were used to assess if the sulfonic group and aspartic residue may act creating new or stronger "hot spots" that trigger the aggregation in the oxidized FF domain. Therefore, AGGRESCAN (Conchillo-Solé et al., 2007) and Zyggregator (Tartaglia & Vendruscolo, 2008) algorithms were selected to predict the aggregation propensity of FF C57D and C57S mutants compared to that of the wild type domain. As shown in Table IV-2.4.7, no significant changes in the FF C57S aggregation propensity are predicted for both methods, while, contrarily to experimental data, a reduction of the aggregation propensity is observed in the case of the FF C57D mutant suggesting that the FF amyloid formation is not a sequence, but a structure dependent event. Accordingly, the AGGRESCAN analysis of all mutants generated by FoldX shows similar aggregation propensities for both Asp and Ser variants confirming that, as demonstrated for the FF domain, the sulfonation induced aggregation is not mediated by novel "hot spots" interactions (Figure IV-2.4.15B).

Table IV-2.4.7|**Aggregation propensity of FF domains predicted by two different sequence-based methods.**

	AGGRESCAN	Zyggregator
FF WT	-18.20	-3.80
FF C57S	-19.70	-3.88
FF C57D	-22.40	-4.21

Discussion

Biological aging and accumulation of oxidized proteins are involved in a wide number of diseases, such as amyotrophic lateral sclerosis, Alzheimer's disease, respiratory distress syndrome, muscular dystrophy, cataractogenesis, rheumatoid arthritis, progeria, and Werner's syndrome (Berlett & Stadtman, 1997). Oxidative process seriously affects protein structure and could lead to the alteration in the secondary and tertiary structure of proteins with dissociation of subunits, unfolding, exposure of hydrophobic residues, aggregation, and backbone fragmentation (Dean et al., 1997). It has been proposed that oxidatively modified proteins are thermodynamically unstable and they might readily form aggregates that would act as intermediates in the formation of amyloid fibrils (Squier, 2001; Dunlop et al., 2009). However, little experimental evidence exists for this oxidation promoted conversion and whether this occurs to global or local modification of the proteins is unknown. In the present work, the hydrogen peroxide-induced cysteine oxidation of the URN1 FF domain was studied

to investigate the impact of the irreversible sulfonation modification upon the conformation and aggregation propensity of this small protein model. The data reported above demonstrate how the attachment of a sulfonic group at the cysteine in position 57 leads to strong secondary and tertiary structure perturbations with a significant decrease of thermal and chemical stability. The specificity of the sulfonation destabilizing effect was confirmed by the study of a sulfonic mimic, the FF C57D mutant which, contrarily to the FF C57S mutant, is able to fully recapitulate the sulfonated FF behaviour displaying conformational and stability features highly similar to those of the oxidized species indicating that the destabilization effect observed in the FF domain does not proceed from the lack of active cysteine residue but it is the consequence of the specific introduction of a sulfonic group in that position. The use of MD simulations allowed us to study in atomistic detail the effect of the aspartic mutation on the structural stability of the FF domain. In good agreement with the observed decrease in thermal and chemical stability, MD results indicate that the C57D mutation induces a significant destabilization, which is reflected in an increase in the radius of gyration (Rg) of the molecule. These effects result in a diminution in the helical content, affecting the helix1 and the helix 3, but not the helix2. This change in helicity impacts the solvent accessibility of the two Trp residues of the protein, in good agreement with our previous spectroscopic data. Importantly, we have shown before that the helix1 comprises the most aggregation-prone region in the FF domain (Castillo et al., 2013) and therefore it is not surprising that an increase in its flexibility or a decrease in its folded population, as suggested by MD simulations, would result in the aggregation of this domain. Next, we were able to analyse for the first time the impact of Cys sulfonation upon the folding and unfolding rates of the FF domain, revealing that the conformational changes imposed by the sulfonic group do not affect significantly the folding rate. Accordingly, the ϕ values calculated for the modified FF protein and its two variants are very close to 0, indicating that the C-terminal region does not present a native-like structure in the transition state for folding, as described for the human HYPA/FBP11 FF domain (Jemth et al., 2005). The unfolding rate, however, is dramatically accelerated by the introduction of the sulfonic group or by the presence of the aspartic residue, highlighting again the sulfonation impact on the maintenance of the FF native structure. As described for other oxidatevely modified proteins, this specific protein modification acts by significantly reducing the kinetic stability of the FF domain (Jayaraman et al., 2008). The destabilization of the native state is a usual requirement for amyloid fibril formation in globular proteins (Chiti et al., 1999). Increasing evidence indicates that there may be a link between oxidative stress induced by reactive oxygen species (ROS) and accumulation of aberrant or misfolded proteins to form aggregates (Bossy-Wetzel et al, 2004). In particular, recent works have described that protein cysteine sulfonation may be implied in the formation of amyloid deposits, such in the case of the mammalian Cu,Zn-superoxide dismutase (SOD1), parkin protein or transthyretin (TTR) (Choi et al., 2005; Meng, et al.,

2011; Nakanishi et al.,2010). In agreement with these observations, in the present work the sulfonation-induced destabilization of the two analysed FF domains leads to the formation of amyloid-like fibrils with a different morphology and cell toxicity due probably to intrinsic differences in their physicochemical properties (Olofsson et al., 2002; Mossuto et al., 2010; Bolognesi et al., 2010).

Remarkably, it was also demonstrated here that the sulfonation impact upon protein stability, as analysed for the FF domain, might also occur in a large number of unrelated proteins containing solvent exposed cysteines and, additionally, that sulfonation-mediated aggregation is not caused by an increase in sequence aggregation propensity but it rather responds to the conformational and structural changes promoted by cysteine sulfonation as indicated for other oxidatevely modified proteins (Milzani et al., 2000; Lasch et al., 2001; Zhou et al., 2014). The FF domain is a highly soluble, remarkably stable and fast folding protein. Despite this properties confront aggregation, the sulfonation of a single Cys residue suffices to promote its aggregation into amyloid fibrils. Since most of the proteins in the cell did not exhibit these beneficial properties it is likely that many of them would readily aggregate upon sulfonation of their exposed Cys residues under oxidative stress. Overall, this *in vitro* study offers new insights into the specific mechanism by which irreversible oxidative protein modifications occurring during aging or stress conditions might damage the cell.

Methods

URN1 FF cloning, expression and purification

FF domain of yeast URN1 protein corresponding to 212-266 residues was cloned into a pETM-30 vector as an N-terminal fusion protein with a His tag, GST protein and a TEV protease cleavage site as previously described (Bonet et al., 2008). Competent *E. coli* BL21 (DE3) cells were transformed with FF containing plasmid and incubated in Luria Broth medium overnight at 310 K. After a 1/100 (v/v) dilution, cells were induced at 0.6 optical density with 1 mM IPTG overnight at 298K. To protein fusion purification a His-tag column was used and GST protein was removed incubating overnight the purified product with TEV protease. The further step of gel filtration HiLoad^{TM} Superdex^{TM} 75 prepgrade (GE healthcare Life Sciences) was used to separate the GST protein and a final dialysis against water was employed to buffer exchange followed by the sample lyophilization. The purity of the samples was checked by SDS- PAGE and MALDI-TOF mass spectroscopy. Protein concentration was determined by UV absorption using an ε value of 1.948 mg^{-1} mL cm^{-1}. URN1 FF mutants were engineered by PCR site-directed mutagenesis and they were produced and purified as described above. Unless otherwise noted, all the experiments were carried out in 50 mM of sodium acetate buffer at pH 5.7.

FoldX modeling

URN1 FF mutant models were built using the BuildModel option of the FoldX algorithm (Schymkowitz et al., 2005a) (version 3b6) from the original WT-URN1 FF structure as deposited in the PDB under code 2JUC. BuildModel command duplicates internally the wild type Protein Data Bank coordinates and it mutates the selected position in the variant ensuring that the movement of the neighbouring side chains results the same in wild type and mutant structure. Five mutant structures are generated as output that were used to calculate the effect of the mutation in termini of ΔΔG values representing the difference between the energy of the wild type protein and that of the mutant protein ($\Delta\Delta G = \Delta G_{mut} - \Delta G_{wt}$). The ΔΔG values are provided in kilocalories per mole for all modelled structures. Variant stability should be significantly affected by ΔΔG values greater than 1.6 kcal/mol since these values correspond to twice the intrinsic standard deviation of FoldX (Guerois et al., 2002).

Sample oxidation

The oxidation of URN1 FF was obtained by incubating the FF sample (1 mg/mL) in presence of 6% w/w H_2O_2 solution for 24 h at 298 K. An extensive water dialysis at 277 K was used to remove non-reacted peroxide from the protein solution before the final lyophilization step. The oxidation state of the sample was checked by SDS-PAGE gel and mass spectroscopy measurements. The identification of specific oxidation site is determined by tryptic digestion of the sample where derived fragments were analyzed by MALDI–TOF technique. Mass spectrometry analyses were carried out in the Proteomics facility from UAB.

Circular dichroism, intrinsic fluorescence and bis ANS binding

Lyophilized samples were dissolved at 20 µM in 50 mM sodium acetate at pH 5.7 and protein solutions were filtered through a 0.22 µm filter and immediately analyzed.. Far-UV CD spectra were recorded from 260 to 200 nm, at 0.2 nm intervals, 1 nm bandwidth and a scan speed of 100 nm/min in a Jasco-710 spectropolarimeter (Jasco Corporation, Japan) thermostated at 298 K. Twenty accumulations were averaged for each spectrum. Tryptophan intrinsic fluorescence was analyzed at 298 K in a Jasco FP8200 spectrofluorometer (Jasco Corporation, Japan).Three averaged spectra were accumulated using an excitation wavelength of 280 nm and recording the emission from 300 to 400 nm; slit widths were 5 nm for both excitation and emission. To study the binding to 4,4'-dianilino-1,1'- binaphthyl-5,5'-disulfonic acid (bis-ANS), protein solutions were prepared at 20 µM in presence of 10 µM of bis-ANS and analyzed immediately. Three spectra were recorded in a Jasco FP8200 spectrofluorometer (Jasco Corporation, Japan) using. an excitation wavelength of 370 nm and emission between 400 and 600 nm, with slit widths of 5 nm.

Thermal and chemical denaturation

Protein samples were dissolved at 20 µM to study thermal stabilities by far-UV CD and intrinsic fluorescence intensity. The molar ellipticity at 222 nm was registered in a Jasco-710 spectropolarimeter (Jasco Corporation, Japan) in a range of 273 and 363 K, each 0.1 K with 2 min of temperature equilibrium between measures. Intrinsic tryptophan fluorescence emission was recorded at 360 nm after excitation at 280 nm, using slit widths of 5 nm for both wavelengths in a Jasco FP8200 spectrofluorometer (Jasco Corporation, Japan). The fluorescence emission was monitored each 0.5 K with 2 min of temperature equilibrium between measures.

For chemical denaturation, samples were prepared at 20 µM in presence of different concentrations of denaturant agent (0-9 M of urea). The reaction was left to equilibrate for 20h at room temperature and analyzed by CD and fluorescence intrinsic intensity at 222 nm and 280/360 nm (ex/em), respectively. All experimental data proceeding by denaturation assays were fitted to a two-state transition curve where the signals of the folded and unfolded states are dependent on the temperature/denaturant concentration using the nonlinear least squares algorithm provided with Kaleidagraph (Abelbeck Software).

NMR spectroscopy

Lyophilized URN1 FF proteins were prepared at 50 µM in 20 mM sodium phosphate buffer and 130 mM NaCl, at pH 5.7, using a 9:1 H_2O/D_2O ratio. One-dimensional NMR spectra were accumulated at specific temperatures on a Bruker AVANCE 600-MHz spectrometer using solvent suppression WATERGATE techniques. The collected spectra were processed and analyzed using the TopSpinv2.0 software packages from Bruker Biospin.

Folding and unfolding kinetics parameters

Kinetics of folding and unfolding reactions were followed at 298 K in a Bio-Logic SFM-3 stopped-flow instrument using excitation at 280 nm and a 320 nm fluorescence cut-off filter. An initial protein solution at 20 µM was prepared and appropriate volumes of the same buffer containing 9.5 M urea were added to promote the unfolding reaction. For the refolding reaction, selected volumes of free urea buffer were added to the initial protein solution at 20 µM containing 9.5 M urea. Kinetic constants were fitted to the equation describing the folding of a two-state protein, using the Kaleidagraph version 4.0 (Synergy Software). Kinetic and free energy values were determined using the following equations:

$$\Delta G_{U\text{-}F} = -RT \ln (k_u / k_f); \quad m_{U\text{-}F} = RT (m_f + m_u)$$

where *kf* and *ku* are the rates of folding and unfolding, respectively, and the m_f and m_u values represent the slopes of the respective folding and unfolding zones.

Unfolding free energy differences ($\Delta\Delta G_{F\text{-}U}$) between the wild-type protein and each mutant (*) are calculated as:

$$\Delta\Delta G_{F\text{-}U}\,(\Delta G_{F\text{-}U}) - (\Delta G_{F\text{-}U})^* = \Delta\Delta G_{\ddagger\text{-}U} - \Delta\Delta G_{\ddagger\text{-}F}$$

$$\Delta\Delta G_{\ddagger\text{-}U} = RT \ln (k_f / k_f^*) \qquad \Delta\Delta G_{\ddagger\text{-}F} = RT \ln (k_u / k_u^*)$$

$$\phi_F = \Delta\Delta G_{\ddagger\text{-}U} \,/\, \Delta\Delta G_{F\text{-}U}$$

where the parameter ϕ_F represents the fraction of the mutated residue's interactions in the transition state.

Molecular dynamics (MD) simulations

The experimental structure of the WT FF domain (pdb ID 2JUC) was used as starting structures for our WT simulations. The C57S and C57D mutations were modelled using Pymol (www.pymol.org). Simulations were performed with Gromacs 4.6.5 with the PLUMED 1.3 plugin (www.gromacs.org) for the metadynamics calculations and with Gromacs alone for the conventional MD simulations. The proteins were solvated in a dodecahedral box (length: 1.0 nm) of tip3p water molecules, and we used CHARMM22* force field. The van der Waals interactions were smoothly shifted to zero between 0.8 and 1.0 nm, and the long range electrostatic interactions were calculated by the particle mesh Ewald algorithm with a 0.12 nm mesh spacing combined with a switch function for the direct space between 0.8 and 1.0 nm. The system evolves in the canonical ensemble, coupled with a velocity-rescale thermostat and a time step of 2 fs. Conventional MD simulations and metadynamics simulations were 1 microsecond and 400 ns long, respectively. For metadynamics, we used MetaD in the well-tempered ensemble to enhance sampling. In such simulations, sampling of the free energy surface is enhanced by adding a history-dependent potential to a set of collective variables. The simulations were subjected to a bias along four collective variables which accounts for radius of gyration and helical content of helix1, 2 and 3.

Aggregation Kinetics

Lyophilized URN1-FF samples were prepared at 100 µM in sodium acetate buffer at pH 5.7 in presence of 25 µM of Thioflavine T (Th-T). After equilibrating during 5 minutes at 310 K, Th-T fluorescence intensity was monitored during 60 minutes in a Jasco FP8200 spectrofluorometer (Jasco Corporation, Japan) using the Th-T excitation and emission wavelength at 440 and 475 nm, respectively. Slit widths of 5 nm were used for excitation and emission.

Structural analysis and amyloid dyes binding of protein aggregates

Lyophilized protein samples were prepared at 50 µM in sodium acetate buffer at pH 5.7 and incubated at 310 K under agitation during one week.

Aggregated samples were diluted to 10 µM in sodium acetate buffer at pH 5.7 in presence of 25 µM of Th-T and analyzed in a Jasco FP8200 spectrofluorometer (Jasco Corporation, Japan). Three Th-T spectra were accumulated exciting at 440 nm and acquiring fluorescence emission between 460 and 600 nm with excitation and emission slit widths of 5 nm. To study Congo Red (CR) binding, URN1-FF aggregates were diluted to 10 µM in presence of 20 µM of CR and optical absorption spectra were recorded from 400 to 700 nm at 298 K. Each trace was the average of 3 accumulated spectra in a Cary-400 UV/Vis spectrophotometer (Varian Inc.). Spectra of only protein and CR solution were acquired to remove protein scattering and dye contribution.

Aggregated URN1-FF solutions, diluted at 10 µM, were also analysed by far-UV CD and bis-ANS dye binding as described previously for the soluble samples.

Transmission Electron Microscopy (TEM)

Aggregated samples were diluted tenfold with water and 10 µL were placed on carbon-coated copper grids during 5 min. The grids were then washed with distilled water and stained with 2% (w/v) uranyl acetate for 1 min. HITACHI H-7000 transmission electron microscope operating at an accelerating voltage of 75 kV was used to fibrils analysis.

Cell viability assays

Lyophilized FF protein was prepared at 50 µM in sterile sodium acetate buffer at pH 5.7 and incubated under agitation during one week at 310 K. The toxicity of extracellularly added FF aggregates was examined on neuroblastoma (SH-SY5Y) cell cultures. SH-SY5Y cells were seeded in 96-well plates in F-12 medium supplemented with 10% FBS at 30% confluence and incubated for 72h in the presence of 10, 20 and 40 µM FF aggregates resuspended in sterile PBS in triplicates. The corresponding controls were performed by incubating cells with PBS buffer. Treated cells were incubated with EZ4U reagent (Biomedica), following the manufacturer instructions, for 1h at 37ºC and the absorbance at 450 nm and 620 nm as a reference were read.

PART V – General discussion

In the last decades, the increasing number of human diseases associated with amyloid deposition and, in parallel, the formation of insoluble aggregates (IBs) during the recombinant expression of proteins with therapeutic interest has strongly promoted the study of protein aggregation in both the biomedical and biotechnological fields. In addition, the recent evidences of functional amyloid structures in different organisms ranging from bacteria to mammalians has led to the idea that aggregation propensity could be considered as a generic ability of most polypeptide chains. This concept was also supported by the identification of a recurrent amyloid-like structure in IBs. These findings opened a door for the research of efficient strategies against protein aggregation by analysing physiologically functional amyloids, by modeling intracellular protein aggregation in simple but biologically relevant conditions , by dissecting the extrinsic and intrinsic determinants of protein amyloid formation under controlled conditions *in vitro* and, finally, by assisting these analysis with the power of bioinformatics.

Accordingly, in this thesis a battery of biophysical and computational approaches coupled with the use of *E.coli* as simple *in vivo* experimental model has allowed us to address the mechanisms and molecular determinants of functional and non-functional aggregation in different protein models.

The analysis of the *S.aureus* biofilm formation has confirmed the phenol soluble modulins (PSMs) ability to promote the assembly of an amyloid scaffold, as previously reported, revealing, however, that not all the PSMs participate in this process. Interestingly, we have demonstrated that the well characterized amyloid-like features exhibited by PSM1 and PSM4 do not affect cell fitness. By contrast the highly soluble PSM3 possesses a cytostatic activity, in agreement with its previously described lytic activity. Therefore despite their highly similar amino acidic content, sequence and common amphipathic α-helical structure, it seems that the distinct functions shown for these peptides might be directed by a subtle balance of hydrophobic/hydrophilic forces in the context of their native α helical structures.

The study of the IBs formed by the prion-like HNRPDL in bacteria shows their amyloid-like nature as confirmed by their amyloid dyes binding capability, their intermolecular β-sheet conformations and their fibrillary structure. Here we demonstrated that the presence of a short amyloidogenic stretch in PrLDs can determine its amyloid propensity as predicted with pWALTZ algorithm, supporting the so called amyloid-stretch hypothesis. The inherent aggregation of human PrLD-containing proteins and their association to disease, strongly suggest that these domains are conserved because they are needed for functional purposes constituting yet a new example where the determinants involved in functional interactions and those for accounting for the formation of toxic amyloid assemblies overlap significantly.

Using the all α-helical globular FF domain we have characterized the TFE-induced changes in its aggregation process revealing that micro-environments which affect the competition between native-like and amyloidogenic contacts might influence not only

the transition between the soluble and fibrillar states, but also the conformational properties of the aggregated assemblies, and thus potentially their cytotoxicity. In this context, it is feasible that chemical compounds that selectively stabilize native α-helices in proteins will increase the energy barrier of fibrillation and therefore can be used to fight amyloid formation at a molecular level.

The introduction of a novel disulfide bond in the small globular FF protein has allowed us to demonstrate that the large stabilization provided by the covalent bond does not protect against the aggregation phenomenon, unless it protects the specific aggregation prone regions. However, our data show that when the aggregation process is triggered by this specific protein stretch, the kinetics, structure, morphology and likely the toxicity of the final amyloid aggregates are strongly influenced by the structural/topological characteristics of the aggregation prone state of the protein.

This thesis also provides new insights about the effect of specific irreversible post-translational sulfonation on globular proteins conformational properties, showing the drastic impact exerted by this modification upon the stability and aggregation propensity of our protein model. The data reported here indicate that the specific sulfonation of the only Cys residue in the FF domain compromises seriously its structural properties causing a large decrease in kinetic stability and promoting the formation of eventually toxic amyloid fibrils under physiological conditions. These results highlight the considerable impact that sulfonation modifications of exposed protein cysteines might have on global cell proteostasis under oxidative stress conditions.

PART VI – Concluding remarks

Functional aggregation

1. Dissecting the contribution of *S. aureus* αPSMs to biofilm amyloid structure and toxicity

- We have demonstrated that, contrary to what was previously assumed, not all αPSMs contribute to the amyloid structure in the *S.aureus* biofilm, consistent with the selectivity required for functional aggregation process.

- Only αPM1 and αPM4 peptides participate in the amyloid fibrillogenesis, rendering similar final fibrillar structures but displaying different aggregation kinetics, which suggests that they might contribute differently to the initiation of amyloid assembly in the extracellular space.

- αPM1 and αPM4 preformed fibrils do not suffice to promote the aggregation of the other PSMs, while both peptides seed their homologous soluble counterparts. The cross-seeding effect is observed only for αPM1 and not for αPM4 showing that the aggregation of αPM1 is driven by a region shared with αPM4 sequence and that, despite the high sequential similarity, αPM4 aggregation may be driven by alternative aggregation-prone regions.

- We have also demonstrated that the natural compound EGCg can be used as potential inhibitor of *S. aureus* biofilm formation and propagation due to its inhibitory and disaggregational activity towards PSMs fibrillization.

- The intracellular αPSMs expression in *E. coli* does not affect the cell growth supporting the idea that functional amyloids are not inherently toxic for the cell since they are designed to be used as structural or protective tools. The bacterial fitness is perturbed only when the soluble αPSM3 is expressed, in agreement with its toxic extracellular role. Therefore this peptide accounts for all the reported PSM-associated bacterial virulence.

- Analysing in detail the hydrophobicity of the common PSM α-helical structure we have proposed that the resulting PSM ability to form amyloid structure or to exert a toxic effect may depend on a thin balance of hydrophobic/ hydrophilic forces in the context of PSM amphipathic structures, showing thus an example of the tight regulation employed by Nature to switch between functional amyloid and toxic conformations.

Non-functional aggregation

2. The prion-like RNA-processing protein HNRPDL forms inherently toxic amyloid-like inclusion bodies in bacteria

- We have demonstrated that the presence of a short amyloidogenic stretch in prion-like domains directs to a large extent its amyloid ability, as predicted by the pWALTZ algorithm.

- The *in vivo* characterization of HNRPDL in a bacterial model system confirms the presence of amyloid structures in its intracellular aggregates as previously demonstrated for the yeast prion Ure2p and the Alzheimer's peptide Aβ42.

- The *in vitro* analysis of purified HNRPDL IBs shows that they fulfil all the amyloid hallmarks since the aggregates bind to amyloid dyes, are enriched in intermolecular β-sheet conformation and contain inner fibril-like structure indicating that the amorphous aggregates previously described for the TDP-43 are not the generic conformation achieved by human PrLD-containing proteins.

- The neuro-toxicity exhibited for HNRPDL IBs confirms the neurotoxicity exerted by the aggregates formed by human prion-like proteins.

- Overall, our data suggest that at least some of the disorders caused by the aggregation of human prion-like proteins would rely on the formation of classical amyloid assemblies. Moreover, they also illustrate the power of microbial cell factories to model amyloid aggregation.

3. Trifluoroethanol modulates amyloid formation by the all α-helical URN1 FF domain

- The FF domain, under acidic conditions and in the presence of TFE, maintains the α-helical secondary structure showing, however, only rudimentary tertiary contacts and highly accessible hydrophobic side-chains.

- The presence of TFE at moderate concentrations delays or prevents FF domain amyloid fibril formation at low pH, which contrasts with the pro-aggregational effect shown for TFE in the same concentration range for a wide range of proteins. Therefore, these data indicate that the effect of this cosolvent is not generic and depends on the conformational properties of the target protein.

- Studying the FF aggregation kinetics we have shown that this cosolvent acts at the early stages of the aggregation reaction interfering with the structural transition from

the initially monomeric and soluble form to aggregation-prone species, probably, by the reinforcement of the backbone hydrogen bonds sustaining the helical structure competing with the intermolecular hydrogen bonds that promote amyloid formation.

- Seeding and cross-seeding analysis show that TFE not only perturbs the formation of the initial nucleus but also changes the conformational and morphological properties of the FF fibrils.

- The remarkable different bis-ANS binding of the fibrils formed in the absence and presence of TFE suggests that the polarity of the fibrillar microenvironment might be an important determinant of the toxicity of the resulting aggregates.

- Taken together, these data propose that chemical compounds that selectively stabilize native α-helices in proteins will increase the energy barrier of fibrillation and therefore they might represent good candidates to fight amyloid formation at the molecular level.

4. Polypeptide chain cross-linking by a novel disulfide bond affects the stability, folding and amyloid fibril formation capability of the all-α FF domain

- The introduction of a disulfide bond in the FF domain results in one of the highest stabilization obtained for a small domain with a single point mutation.

- The crosslinking of the FF protein stabilizes the domain against both thermal and chemical denaturation. The data are in agreement with the theoretical models for the configurational entropy of polymeric chains where the major effect of the crosslink on the stability is caused by a decrease in the configurational entropy of the unfolded state.

- The loss of configurational freedom in the unfoded and crosslinked FF domain results in an enormous increase in its folding rate, which approaches the folding speed limit. The decrease in the unfolding rate indicates that the opening of the crosslinked regions may represent a rate-limiting step in the unfolding process. The lower ϕ values obtained for the reduced form suggests that in the absence of the structural constraint the Cys containing regions are not structured in the transition state, as previously demonstrated for the human FF domain.

- The compact and over-stabilized crosslinked FF domain does not aggregate under physiological conditions. Conversely, at acidic pH it experiments amyloid fibrillation, which indicates that a high global stabilization in a protein does not suffices to prevent

amyloid formation, unless it protects the specific regions that nucleate the aggregation.

- We have shown that the structural/topological characteristics of the aggregation prone state play a critical role in the aggregation rate, structure, morphology and likely the toxicity of the final amyloid state.

5. Cysteine specific post-translational sulfonation reduces FF domain stability and promotes amyloid fibrils formation

- We have demonstrated that the hydrogen peroxide-induced oxidation of the FF domain results in the specific introduction of a sulfonic group in the unique cysteine residue of this small protein model.

- FF domain cysteine sulfonation promotes important structural and conformational changes, resulting in a significant decrease in thermal and chemical stability.

- Using a FF mutant as sulfonated mimic, we have shown that the destabilization effect observed in the FF domain is not caused by the lack of the original cysteine residue but instead it is totally attributable to the specific introduction of the sulfonic group in that position.

- The sulfonation induced-conformational changes were also demonstrated by metadynamic MD simulations where the opening of the globular structure and a decrease in the helical content is clearly observed for the C57D mutant.

- The introduction of the sulfonic group does not affect significantly the folding rate and ϕ analysis indicate that the C-terminal region does not present a native-like structure in the transition state for folding as previously reported for the human FF domain. However, the unfolding rate is dramatically affected suggesting that this specific protein modification acts by significantly reducing the kinetic stability of the FF domain.

- The sulfonation-induced kinetic destabilization of the two analyzed FF domains promotes the formation of amyloid aggregates under physiological conditions which show distinct morphology and cytotoxicity, probably owing to differences in their physicochemical properties.

- Interestingly, we have demonstrated by a computational study that the sulfonation effect on protein stability might also occur in a large number of unrelated proteins containing solvent exposed cysteines.

- Using the AGGRESCAN algorithm we shown that sulfonation-mediated aggregation does not depend on the increase in sequence aggregation propensity but rather on the conformational and structural changes promoted by the thiol group modification.

- Overall, this study offers important insights at the residue level for understanding the impact of the irreversible oxidative protein modifications occurring during aging or stress conditions on protein conformation and amyloid propensity.

PART VII - References

Ahmed, A. B., & Kajava, A. V. (**2013**). Breaking the amyloidogenicity code: Methods to predict amyloids from amino acid sequence. *FEBS Lett*, *587*, 1089–1095.

Alberti, S., Halfmann, R., King, O., Kapila, A., & Lindquist, S. (**2009**). A Systematic Survey Identifies Prions and Illuminates Sequence Features of Prionogenic Proteins. *Cell*, *137*, 146–158.

Allen, M., Friedler, A., Schon, O., Bycroft, M. (**2002**). The structure of an FF domain from human hypa/fbp11. *J Mol Biol, 323*, 411–416.

Andersen, C. B., Yagi, H., Manno, M., Martorana, V., Ban, T., Christiansen, G., … Rischel, C. (**2009**). Branching in amyloid fibril growth. *Biophys J*, *96*, 1529–1536.

Anderson, V.L. & Webb, W.W. (**2012**). A desolvation model for trifluoroethanol-induced aggregation of enhanced green fluorescent protein. *Biophys J*, 102, 897–906.

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

Arolas, J.L., D'Silva, L., Popowicz, G.M., Aviles, F.X., Holak, T.A. & Ventura S. (**2005**). NMR structural characterization and computational predictions of the major intermediate in oxidative folding of leech carboxypeptidase inhibitor. *Structure, 13*, 1193–1202.

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

Babu, M. M., van der Lee, R., de Groot, N. S., & Gsponer, J. (**2011**). Intrinsically disordered proteins: Regulation and disease. *Curr Opin Struct Biol*, *21*, 432–440.

Baldwin, R.L. & Rose, G.D. (**2013**). Molten globules, entropy-driven conformational change and protein folding. *Curr Opin Struct Biol.,23*,4-10.

Baldwin, A. J., Knowles, T. P. J., Tartaglia, G. G., Fitzpatrick, A. W., Devlin, G. L., Shammas, S. L., … Dobson, C. M. (**2011**). Metastability of native proteins and the phenomenon of amyloid formation. *J.Am.Chem.Soc., 133*, 14160–14163.

Baldwin, R. L. (**1996**). On-pathway versus off-pathway folding intermediates. *Folding & Design*, *1*, R1–R8.

Baldwin, R. L. (**2008**). The search for folding intermediates and the mechanism of protein folding. *Ann Rev Biophys, 37*, 1–21.

Barnhart, M. M., & Chapman, M. R. (**2006**). Curli Biogenesis and Function. *Annu Rev Microbiol*, *60*, 131–147.

Bedford, M.T.; Leder, P. (**1999**). The FF domain: A novel motif that often accompanies WW domains. *Trends Biochem Sci, 24*, 264–265.

Beerten, J., Schymkowitz, J., & Rousseau, F. (**2012**). Aggregation prone regions and gatekeeping residues in protein sequences. *Curr Top Med Chem, 12*, 2470–8.

Belli, M., Ramazzotti, M., & Chiti, F. (**2011**). Prediction of amyloid aggregation in vivo. *EMBO Rep*, *12*, 657–663.

Bemporad, F.; Calloni, G.; Campioni, S.; Plakoutsi, G.; Taddei, N.; Chiti, F. (**2006**). Sequence and structural determinants of amyloid fibril formation. *Acc. Chem. Res., 39*, 620–627.

Berke, S.J.S. & Paulson, H.L., (**2003**). Protein aggregation and the ubiquitin proteasome pathway: gaining the UPPer hand on neurodegeneration. *Curr Opin Genet Dev*, 13, 253–261.

Berlett, B. S. & Stadtman, E. R. (**1997**). Protein oxidation in aging, disease, and oxidative stress. *J Biol Chem, 272*, 20313–20316.

Bernstein, F.C., Koetzle, T.F., Williams, G.J., Meyer, E.F. Jr, Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., Tasumi, M. (**1977**). The Protein Data Bank: a computer-based archival file for macromolecular structures. *J Mol Biol*, *112*, 535–542.

Betz, S. F. (**1993**). Disulfide bonds and the stability of globular proteins. *Protein Sci*, *2*, 1551–1558.

Betz, S.F. & Pielak, G.J. (**1992**). Introduction of a disulfide bond into cytochrome c stabilizes a compact denatured state. *Biochemestry, 31*, 12337–12344.

Bieler, S., Estrada, L., Lagos, R., Baeza, M., Castilla, J., & Soto, C. (**2005**). Amyloid formation modulates the biological activity of a bacterial protein. *J Biol Chem*, *280*, 26880–26885.

Bieschke, J., Russ, J., Friedrich, R. P., Ehrnhoefer, D. E., Wobst, H., Neugebauer, K., & Wanker, E. E. (**2010**). EGCG remodels mature alpha-synuclein and amyloid-beta fibrils and reduces cellular toxicity. *Proc Natl Acad Sci USA, 107*, 7710–7715.

Blanco, A. R., Sudano-roccaro, A., Spoto, C., Nostro, A., Rusciano, D., & Spoto, G. C. (**2005**). Epigallocatechin Gallate Inhibits Biofilm Formation by Ocular Staphylococcal Isolates. *Antimicrob Agents Chemother, 49*, 4339–4343.

Blokhuis, A.M., Groen, E.J., Koppers, M., van den Berg, L.H., Pasterkamp, R.J. (**2013**). Protein aggregation in amyotrophic lateral sclerosis. *Acta Neuropathol, 125*, 777–794.

Boles, B. R., & Horswill, A. R. (**2011**). Staphylococcal biofilm disassembly. *Trends Microbiol*, *19*, 449–455.

Bolognesi, B., Kumita, J.R., Barros, T.P., Esbjorner, E.K., Luheshi, L.M., Crowther, D.C., Wilson MR, Dobson CM, Favrin G, and Yerbury JJ. (**2010**). ANS binding reveals common features of cytotoxic amyloid species. *ACS Chem Biol, 5*, 735–740.

Bonet, R., Ramirez-Espain, X. & Macias, M.J. (**2008**). Solution structure of the yeast urn1 splicing factor FF domain: Comparative analysis of charge distributions in FF domain structures-ffs and surps, two domains with a similar fold. *Proteins*, *73*, 1001–1009.

Booth, D.R., Sunde, M., Bellotti, V., Robinson, C.V., Hutchinson, W.L., Fraser, P.E., Hawkins, P.N., Dobson, C.M., Radford, S.E., Blake, C.C. & Pepys, M.B. (**1997**). Instability, unfolding and aggregation of human lysozyme variants underlying amyloid fibrillogenesis. *Nature, 385*, 787–793.

Bossy-Wetzel, E., Schwarzenbacher, R., & Lipton, S. a. (**2004**). Molecular pathways to neurodegeneration. *Nature Med, 10* , S2–S9.

Bousset, L., Pieri, L., Ruiz-Arlandis, G., Gath, J., Jensen, P. H., Habenstein, B., … Melki, R. (**2013**). Structural and functional characterization of two alpha-synuclein strains. *Nat Commun*, *4*, 2575.

Brender, J.R., Nanga, R.P., Popovych, N., Soong, R., Macdonald, P.M., Ramamoorthy, A. (**2011**). The amyloidogenic sevi precursor, pap248–286, is highly unfolded in solution despite an underlying helical tendency. *Biochim. Biophys. Acta, 1808*, 1161–1169.

Brieger, K., Schiavone, S., Miller, J., & Krause, K. (**2012**). Reactive oxygen species: from health to disease. *Swiss Medical Weekly*, *142*, 1–14.

Brockwell, D. J. & Radford, S. E. (**2007**). Intermediates: ubiquitous species on folding energy landscapes? *Curr Opin Struct Biol*, *17*, 30–37.

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

Bryan, A. W., Menke, M., Cowen, L. J., Lindquist, S. L., & Berger, B. (**2009**). Betascan: Probable β-amyloids identified by pairwise probabilistic analysis. *PLoS Comp Biol*, *5*.

Bryngelson, J. D. (**1995**). Funnels, Pathways and the Energy Landscape of Protein Folding: A Synthesis. *Proteins*, *21*, 167-195

Caflisch, A. (**2006**). Computational models for the prediction of polypeptide aggregation propensity. *Curr Opin Chem Biol*, *10*, 437–444.

Calamai, M., Chiti, F. & Dobson, C.M. (**2005**). Amyloid fibril formation can proceed from different conformations of a partially unfolded protein. *Biophys. J., 89*, 4201–4210.

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

Canet, D., Sunde, M., Last, A. M., Miranker, A., Spencer, A., Robinson, C. V., & Dobson, C. M. (**1999**). Mechanistic studies of the folding of human lysozyme and the origin of amyloidogenic behavior in its disease-related variants. *Biochemestry*, *38*, 6419–6427.

Canet-Avilés, R. M., Wilson, M. a, Miller, D. W., Ahmad, R., McLendon, C., Bandyopadhyay, S., … Cookson, M. R. (**2004**). The Parkinson's disease protein DJ-1 is neuroprotective due to cysteine-sulfinic acid-driven mitochondrial localization. *Proc Natl Acad Sci USA*, *101*, 9103–9108.

Capitini, C., Conti, S., Perni, M., Guidi, F., Cascella, R., De Poli, A. et al. (**2014**). TDP-43 inclusion bodies formed in bacteria are structurally amorphous, non-amyloid and inherently toxic to neuroblastoma cells. *PLoS One, 9*, e86720.

Carrió, M. M., Corchero, J. L., & Villaverde, a. (**1998**). Dynamics of in vivo protein aggregation: Building inclusion bodies in recombinant bacteria. *FEMS Microbiology Letters*, *169*, 9–15.

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

Castillo, V., Chiti, F., & Ventura, S. (**2013**). The N-terminal helix controls the transition between the soluble and amyloid states of an FF domain. *PloS One, 8*.

Castillo, V., Espargaro, A., Gordo, V., Vendrell, J., & Ventura, S. (**2010**). Deciphering the role of the thermodynamic and kinetic sta- bilities of SH3 domains on their aggregation inside bacteria. *Proteomics, 10*, 4172–4185.

Chambers, H.F.. (**2001**).The changing epidemiology of Staphylococcus aureus? *Emerg. Infect. Dis. 7* ,178–182.

Chang, J. (**2011**). Diverse Pathways of Oxidative Folding of Disulfide Proteins: Underlying Causes and Folding Models. *Biochemestry*, *50*, 3414–3431.

Chatani, E., Yagi, H., Naiki, H., Goto, Y. (**2012**). Polymorphism of beta2-microglobulin amyloid fibrils manifested by ultrasonication-enhanced fibril formation in trifluoroethanol. *J. Biol. Chem, 287*, 22827–22837.

Chatterjee, S. S., Joo, H.S., Duong, A. C.,Dieringer, T. D.,Tan, V.Y., Song, Y., Fischer, E.R.,Cheung, G. Y. C., Li, M.& Otto, M.(**2013**).Essential Staphylococcus aureus toxin export system. *Nat Med, 19*, 364–367.

Cheung, G. Y. C., Duong, A. C., & Otto, M. (**2012**). Direct and synergistic hemolysis caused by Staphylococcus phenol-soluble modulins: implications for diagnosis and pathogenesis. *Microbes Infect., 14*, 380–386.

Cheung, G. Y. C., Joo, H. S., Chatterjee, S. S., & Otto, M. (**2014a**). Phenol-soluble modulins - critical determinants of staphylococcal virulence. *FEMS Microbiol Rev*, *38*, 698–719.

Cheung, G. Y. C., Kretschmer, D., Queck, S. Y., Joo, H. S., Wang, R., Duong, A. C., … Otto, M. (**2014b**). Insight into structure-function relationship in phenol-soluble modulins using an alanine screen of the phenol-soluble modulin (PSM) α3 peptide. *FASEB J*, *28*, 153–161.

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

Chiti, F., & Dobson, C. M. (**2006**). Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem*, *75*, 333–66.

Chiti, F., Calamai, M., Taddei, N., Stefani, M., Ramponi, G., & Dobson, C. M. (**2002a**). Studies of the aggregation of mutant proteins in vitro provide insights into the genetics of amyloid diseases., *Proc Natl Acad Sci USA, 99*, 16419–16426.

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*, 805–808.

Chiti, F., Taddei, N., Baroni, F., Capanni, C., Stefani, M., Ramponi, G., & Dobson, C. M. (**2002b**). Kinetic partitioning of protein folding and aggregation. *Nat Struct Biol*, *9*, 137–143.

Chiti, F., Webster, P., Taddei, N., Clark, a, Stefani, M., Ramponi, G., & Dobson, C. M. (**1999**). Designing conditions for in vitro formation of amyloid protofilaments and fibrils. *Proc Natl Acad Sci USA, 96,* 3590–3594.

Choi, J., Rees, H. D., Weintraub, S. T., Levey, A. I., Chin, L. S., & Li, L. (**2005**). Oxidative modifications and aggregation of Cu,Zn-superoxide dismutase associated with alzheimer and Parkinson diseases. *J Biol Chem, 280*, 11648–11655.

Colletier, J., Laganowsky, A., Landau, M., Zhao, M., & Soriaga, A. B. (**2011**). Molecular basis for amyloid-β polymorphism*., Proc Natl Acad Sci USA, 108*, 16938–16943.

Conchillo-Solé, O., de Groot, N. S., Avilés, 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 Bioinform*, *8*, 65.

Couthouis, J., Hart, M.P., Erion, R., King, O.D., Diaz, Z., Nakaya, T. et al. (**2012**). Evaluating the role of the FUS/TLS-related gene EWSR1 in amyotrophic lateral sclerosis. *Hum Mol Genet, 21*,2899–2911.

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

Creighton, T.E. (**1992)**. Protein Folding. W.H. *Freeman and Company*.

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

Crick, F. (**1970**). Central Dogma of Molecular, *227*, 6–8.

Da Cruz, S. & Cleveland, D.W. (**2011**). Understanding the role of TDP-43 and FUS/TLS in ALS and beyond. *Curr Opin Neurobiol 21*, 904–919.

Daggett, V., & Fersht, A. R. (**2003**). Is there a unifying mechanism for protein folding? *Trends Biochem Sci*, *28*, 18–25.

Dalle-Donne, I., Rossi, R., Giustarini, D., Milzani, A., & Colombo, R. (**2003**). Protein carbonyl groups as biomarkers of oxidative stress. *Clin Chim Acta*, *329*, 23–38.

Darby, N. & Creighton, T.E. (**1995**). Disulfide bonds in protein folding and stability. *Meth Mol Biol,40*, 219–252.

Darby, N., & Creighton, T. E. (**1997**). Probing protein folding and stability using disulfide bonds. *Molec Biotech*, *7*, 57–77.

Das, S., Tanwar, J., Hameed, S., Fatima, Z., & Manesar, G. **(2014**). Antimicrobial potential of epigallocatechin-3- gallate (EGCG): a green tea polyphenol. *J Biochem Pharm Res*, *2*, 167–174.

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.

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

de Groot, N.S., Castillo, V., Grana-Montes, R., Ventura, S. (**2012**). AGGRESCAN: method, application, and perspectives for drug design. *Meth Mol Biol, 819*, 199–220.

Dean, R.T., Fu, S., Stocker, R. & Davies, M.J.(**1997**). Biochemistry and pathology of radical-mediated protein oxidation. *Biochem J, 324* ,1–18.

Dill, K. a., & MacCallum, J. L. (**2012**). The Protein-Folding Problem, 50 Years On. *Science*, *338*, 1042–1046.

Dill, K. A., Ozkan, S. B., Shell, M. S., & Weikl, T. R. (**2008**). The Protein Folding Problem. *Annu Rev Biophys*, 37, 289–316.

Dill, K.A. (**1985**). Theory for the folding and stability of globular proteins. *Biochemestry*, *24*, 1501-1509.

Dobson, C.M. (**2004**) Principles of protein folding, misfolding and aggregation. *Semin Cell Dev Biol*, 15, 3–16

Dobson, C.M. (**2003**) Protein folding and misfolding. *Nature,* 426, 884–890

Dombkowski, A. A., Sultana, K. Z., & Craig, D. B. (**2014**). Protein disulfide engineering. *FEBS Letters, 588*, 206–212.

Dreyfuss, G., Matunis, M.J., Pinol-Roma, S., Burd, C.G. (**1993**). hnRNP proteins and the biogenesis of mRNA. *Annu Rev Biochem, 62*,289–321.

Drozdetskiy, A., Cole, C., Procter, J., & Barton, G. J. (**2015**). JPred4: a protein secondary structure prediction server. *Nucleic Acids Res*, 1–6.

DuBay, K. F., Pawar, A. P., Chiti, F., Zurdo, J., Dobson, C. M., & Vendruscolo, M. (**2004**). Prediction of the absolute aggregation rates of amyloidogenic polypeptide chains. *J Mol Biol*, *341*, 1317–1326.

Dueholm, M. S., Albertsen, M., Otzen, D., & Nielsen, P. H. (**2012**). Curli Functional Amyloid Systems Are Phylogenetically Widespread and Display Large Diversity in Operon and Protein Structure. *PLoS One*, *7*, 1–10.

Dumoulin, M., Kumita, J. R., & Dobson, C. M. (**2006**). Normal and aberrant biological self-assembly: Insights from studies of human lyspzyme and its amyloidogenic variants. *Acc Chem Res*, *39*, 603–610.

Dunlop, R. A., Brunk, U. T. & Rodgers, K. J. (**2009**).Oxidized proteins: mechanisms of removal and consequences of accumulation. *IUBMB Life, 61*, 522–527.

Eane, E.D. & Glenner, G.G. (**1968**). X-ray diffraction studies on amyloid filaments. *J Histochem Cytochem, 16*, 673–677.

Eaton, W. A., Munoz, V., Hagen, S. J., Jas, G. S., Lapidus, L. J., Henry, E. R., & Hofrichter, J. (**2000**). Fast kinetics and mechanisms in protein folding. *Annu Rev Biophys Biomol Struct, 29*, 327–359.

Ehrnhoefer, D.E. et al., (**2008**). EGCG redirects amyloidogenic polypeptides into unstructured, off-pathway oligomers. *Nat Struct Mol Biol*, 15, 558–566.

Eichner, T. & Radford, S.E. (**2011**). Diversity of Assembly Mechanisms of a Generic Amyloid Fold. *Mol. Cell. 43*, 8–18.

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

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

Espargaro, A., Sabate, R., Ventura, S. (**2012a**). Thioflavin-S staining coupled to flow cytometry. A screening tool to detect in vivo protein aggregation. *Mol Bio Syst, 8*, 2839–2844.

Espargaro, A., Villar-Pique, A., Sabate, R., Ventura, S. (**2012b**). Yeast prions form infectious amyloid inclusion bodies in bacteria. *Microb Cell Fact, 11*, 89.

Espinosa Angarica, V., Angulo, A., Giner, A., Losilla, G., Ventura, S., Sancho, J. (**2014**). PrionScan: an online database of predicted prion domains in complete proteomes. *BMC Genom, 15*, 102.

Fahnert, B., Lilie, H., & Neubauer, P. (**2004**). Inclusion bodies: formation and utilisation. *Adv Biochem Eng/Biotechnol, 89*, 93–142.

Falsone, A. & Falsone, S.F. (**2015**). Legal but lethal: functional protein aggregation at the verge of toxicity. *Front Cell Neurosci, 9*, 45.

Fandrich, M.; Forge, V.; Buder, K.; Kittler, M.; Dobson, C.M.; Diekmann, S. (**2003**).Myoglobin forms amyloid fibrils by association of unfolded polypeptide segments. *Proc. Natl. Acad. Sci. USA, 100*, 15463–15468.

Fass, D. (**2012**). Disulfide bonding in protein biophysics, *Annu Rev Biophys, 41*, 63–79.

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

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.

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

Fersht, A. R. (**1997**). Nucleation mechanisms in protein folding. *Curr Opin Struct Biol*, *7*, 3–9.

Fink, A. L. (**1998**). Protein aggregation: Folding aggregates, inclusion bodies and amyloid. *Folding and Design*, *3*, 9–23.

Foss, T.R., Wiseman, R.L. & Kelly, J.W. (**2005**). The pathway by which the tetrameric protein transthyretin dissociates. *Biochemestry, 44*, 15525–15533.

Fowler, D. M., Koulov, A. V., Balch, W. E., & Kelly, J. W. (**2007**). Functional amyloid - from bacteria to humans. *Trends Biochem Sci*, *32*, 217–224.

Freedman, R. B. (**1989**). Protein disulfide isomerase: multiple roles in the modification of nascent secretory proteins. *Cell*, *57*, 1069–1072.

Frousios, K. K., Iconomidou, V. a, Karletidi, C.-M., & Hamodrakas, S. J. (**2009**). Amyloidogenic determinants are usually not buried. *BMC Struct Biol*, *9*, 44.

Galzitskaya, O.V., Garbuzynskiy, S.O., Lobanov, M.Y. (**2006**). FoldUnfold: web server for the prediction of disordered regions in protein chain. *Bioinformatics, 22*, 2948–2949.

Garbuzynskiy, S. O., Lobanov, M. Y., & Galzitskaya, O. V. (**2009**). FoldAmyloid: A method of prediction of amyloidogenic regions from protein sequence. *Bioinformatics*, *26*, 326–332.

García-Fruitós, E. (**2010**). Inclusion bodies: a new concept. *Microb Cell Fact*, *9*, 80.

García-Fruitós, E., González-Montalbán, N., Morell, M., Vera, A., Ferraz, R. M., Arís, A., … Villaverde, A. (**2005**). Aggregation as bacterial inclusion bodies does not imply inactivation of enzymes and fluorescent proteins. *Microb Cell Fact*, *4*, 27.

Gasch, A., Wiesner, S., Martin-Malpartida, P., Ramirez-Espain, X., Ruiz, L., Macias, M.J. (**2006**). The structure of prp40 ff1 domain and its interaction with the crn-tpr1 motif of clf1 gives a new insight into the binding mode of FF domains. *J Biol Chem*, *281*, 356–364

Geser, F., Martinez-Lage, M., Kwong, L.K., Lee, V.M., Trojanowski, J.Q. (**2009**). Amyotrophic lateral sclerosis, frontotemporal dementia and beyond: the TDP-43 diseases. *J Neurol, 256*,1205–1214.

Giangaspero, A., Sandri, L., & Tossi, A. (**2001**). Amphipathic, α-helical antimicrobial peptides. *Eur J Biochem, 268*, 5589–5600.

Gidalevitz, T., Prahlad, V., & Morimoto, R. I. (**2011**). The stress of protein misfolding: From single cells to multicellular organisms. *Cold Spring Harb Persp Biol*, *3*, 1–18.

Goldenberg, D.P. & Creighton, T.E. (**1984**). Folding pathway of a circular form of bovine pancreatic trypsin inhibitor. *J Mol Biol, 179*, 527–545.

Golding, I., & Cox, E. C. (**2006**). Physical nature of bacterial cytoplasm. *Phys Rev Lett*, *96*, 14–17.

Goormaghtigh, E. Cabiaux, V., Ruysschaert, J.M. (**1990**). Secondary structure and dosage of soluble and membrane proteins by attenuated total reflection Fourier-transform infrared spectroscopy on hydrated films. *Eur J Biochem 193,* 409–420.

Gosal, W.S.; Clark, A.H.; Ross-Murphy, S.B. (**2004**). Fibrillar beta-lactoglobulin gels: Part 1. Fibril formation and structure. *Biomacromolecules, 5*, 2408–2419.

Graña-Montes, R., de Groot, N. S., Castillo, V., Sancho, J., Velazquez-Campoy, A., & Ventura, S. (**2012**). Contribution of disulfide bonds to stability, folding, and amyloid fibril formation: the PI3-SH3 domain case. *Antiox Red Signal, 16*, 1–15.

Grudzielanek, S.; Jansen, R.; Winter, R. Solvational tuning of the unfolding, aggregation and amyloidogenesis of insulin. (**2005**). *J Mol Biol, 351*, 879–894.

Guerois, R., Nielsen, J.E., Serrano, L. (**2002**). Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J Mol Biol, 320*, 369–387.

Halfmann, R., Jarosz, D. F., Jones, S. K., Chang, A., Lancaster, A. K., & Lindquist, S. (**2012**). Prions are a common mechanism for phenotypic inheritance in wild yeasts. *Nature*, *482*, 363–368.

Harrison, S. C. & Durbin, R. (**1985**). Is there a single pathway for the folding of a polypeptide chain?, *Proc Natl Acad Sci USA*, *82*, 4028–4030.

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

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

Herva, M. E., & Spillantini, M. G. (**2014**). Parkinson's disease as a member of Prion-like disorders. *Virus Res*, *207,* 38–46

Hijirida, D. H., Do, K. G., Michal, C., Wong, S., Zax, D., & Jelinski, L. W. (**1996**). ^{13}C NMR of Nephila clavipes major ampullate silk gland. *Biophys J*, *71*, 3442–3447.

Hockenbery D., Nunez G., Milliman C., Schreiber R.D., Korsmeyer S.J. (**1990**). Bcl-2 is an inner mitochondrial membrane protein that blocks programmed cell death. *Nature, 348*,334-336.

Hogg, P. J. **(2003**). Disulfide bonds as switches for protein function. *Trends Biochem Sci*, *28*, 210–214.

Holde, P., Sweat Faye, & Levine M. (1961). On the binding of Congo Red by Amyloid. *J Histochem Cytochem, 10*, 355–364.

Hossain, K. S. M. T., Bailey-Kellogg, C., Friedman, A. M., Bradley, M. J., Baker, N., & Ramakrishnan, N. (**2011**). Using physicochemical properties of amino acids to induce graphical models of residue couplings. *Proceedings of the Tenth International Workshop on Data Mining in Bioinformatics - BIOKDD '11*, 1–10.

Hou, L., Kang, I., Marchant, R. E., & Zagorski, M. G. (**2002**). Methionine 35 oxidation reduces fibril assembly of the amyloid aβ-(1-42) peptide of Alzheimer's disease. *J Biol Chem, 277*, 40173–40176.

Howie, A.J. & Brewer, D.B. (**2009**). Optical properties of amyloid stained by Congo red: History and mechanisms. *Micron, 40*, 285–301

Invernizzi, G., Papaleo, E., Sabate, R., Ventura, S. (**2012**). Protein aggregation: mechanisms and functional consequences. *Int J Biochem Cell Biol*, 44,1541–1554

Ivanova, M. I., Sawaya, M. R., Gingery, M., Attinger, A., & Eisenberg, D. (**2004**). An amyloid-forming segment of beta2-microglobulin suggests a molecular model for the fibril., *Proc Natl Acad Sci USA, 101*, 10584–10589.

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

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

Jahn, T.R. & Radford, S.E. (**2005**). The yin and yang of protein folding. *FEBS J*., *272*, 5962–5970.

Jain, N., Bhattacharya, M., & Mukhopadhyay, S. (**2011**). Chain collapse of an amyloidogenic intrinsically disordered protein. *Biophys J*, *101*, 1720–1729.

Jayaraman, S., Gantz, D.,L. & Gursky, O.(**2008**).Effects of protein oxidation on the structure and stability of model discoidal high-density lipoproteins. *Biochemestry, 47,* 3875–3882.

Jemth P., Day R., Gianni S., Khan F., Allen M., Daggett V., Fersht, A. R.(**2005**). The Structure of the Major Transition State for Folding of an FF Domain from Experiment and Simulation. *J Mol Biol, 350*, 363–378.

Jha, A.; Narayan, S.; Udgaonkar, J.B.; Krishnamoorthy, G. (**2012**). Solvent-induced tuning of internal structure in a protein amyloid protofibril. *Biophys. J., 103*, 797–806.

Johnston, J. a., Ward, C. L., & Kopito, R. R. (**1998**). Aggresomes: A cellular response to misfolded proteins. *J Cell Biol*, *143*, 1883–1898.

Jung Tobias, Höhn Annika, & Grune, T. (**2014**). The proteasome and the degradation of oxidized proteins: Part II - protein oxidation and proteasomal degradation. *Red Biol*, *2*, 99–104.

Kaito, C., Saito, Y., Ikuo, M., Omae, Y., Mao, H., Nagano, G., … Sekimizu, K. (**2013**). Mobile Genetic Element SCCmec-encoded psm-mec RNA Suppresses Translation of agrA and Attenuates MRSA Virulence. *PLoS Pathog*, *9*.

Kaito, C., Saito, Y., Nagano, G., Ikuo, M., Omae, Y., Hanada, Y., … Sekimizu, K. (**2011**). Transcription and translation products of the cytolysin gene psm-mec on the mobile genetic element SCCmec regulate Staphylococcus aureus virulence. *PLoS Pathog*, *7*.

Karplus, M., Weaver, D.L. (**1976**). Protein-folding dynamics. *Nature*, *260*, 404-406.

Karplus, M., & Weaver, D. L. (**1994**). Protein folding dynamics: the diffusion-collision model and experimental data. *Protein Sci*, *3*, 650–668.

Kasper, L. H., & Reder, A. T. (**2014**). Immunomodulatory activity of interferon-beta. *Ann Clin Trans Neurol, 1*, 622–631.

Kenney, J. M., Knight, D., Wise, M. J., & Vollrath, F. (**2002**). Amyloidogenic nature of spider silk. *Eur J Biochem*, *269*, 4159–4163.

Kikis, E. A., Gidalevitz, T., & Morimoto, R. I. (**2010**). Protein homeostasis in models of aging and age-related conformational disease. *Adv Exp Med Biol*, *694*, 138–159.

Kim, P.S. & Baldwin, R.L. (**1982**). Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. *Annu Rev Biochem*, *51*, 459-489.

Kim, H. J., Kim, N. C., Wang, Y.-D., Scarborough, E. a, Moore, J., Diaz, Z., … Taylor, J. P. (**2013**). Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. *Nature*, *495*, 467–73.

Kim, H.-J., Ha, S., Lee, H. Y., & Lee, K.-J. (**2014**). ROSics: Chemistry and proteomics of cysteine modifications in redox biology. *Mass Spectrom Rev*, 1-25

Kim, W., & Hecht, M. H. (**2006**). Generic hydrophobic residues are sufficient to promote aggregation of the Alzheimer's Abeta42 peptide. *Proc Natl Acad Sci USA, 103*, 15824–15829.

Kim, W., & Hecht, M. H. (**2008**). Mutations Enhance the Aggregation Propensity of the Alzheimer's Aβ Peptide. *J Mol Biol*, *377*, 565–574.

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

Klar, J., Sobol, M., Melberg, A., Mabert, K., Ameur, A., Johansson, A.C. et al. (**2013**). Welander distal myopathy caused by an ancient founder mutation in TIA1 associated with perturbed splicing. *Hum Mutat, 34*, 572–577.

Knowles, T. P. J., Vendruscolo, M., & Dobson, C. M. (**2014**). The amyloid state and its association with protein misfolding diseases. *Nat Rev Mol Cell Biol*, *15*, 384–96. .

Knowles, T. P. J., Waudby, C. a, Devlin, G. L., Cohen, S. I. A, Aguzzi, A., Vendruscolo, M., … Dobson, C. M. (**2009**). An analytical solution to the kinetics of breakable filament assembly. *Science*, *326*, 1533–1537.

Korzhnev, D.M., Religa, T.L., Banachewicz, W., Fersht, A.R., Kay, L.E. (**2010**). A transient and low-populated protein-folding intermediate at atomic resolution. *Science, 329*, 1312–1316.

Korzhnev, D.M., Religa, T.L., Kay, L.E. (**2012**). Transiently populated intermediate functions as a branching point of the FF domain folding pathway. *Proc. Natl. Acad. Sci. USA*, *109*, 17777–17782.

Krebs, M. R. H., Morozova-Roche, L. a, Daniel, K., Robinson, C. V, & Dobson, C. M. (**2004**). Observation of sequence specificity in the seeding of protein amyloid fibrils. *Protein Sci*, *13*, 1933–1938.

Krecic, A. M., & Swanson, M. S. (**1999**).hnRNP complexes:composition,structure,and function. *Curr Opin Cell Biol*, *11*, 363–371.

Kuhlman, B., & Baker, D. (**2000**). Native protein sequences are close to optimal for their structures. *Proc Natl Acad Sci USA*, *97*, 10383–10388.

Kuwajima, K. (**1996**). The molten globule state of alpha-lactalbumin. *FASEB J, 10*, 102–109.

Laabei, M., Jamieson, W. D., Yang, Y., van den Elsen, J., & Jenkins, a. T. a. (**2014**). Investigating the lytic activity and structural properties of Staphylococcus aureus phenol soluble modulin (PSM) peptide toxins. *BBA Biomembranes*, *1838*, 3153–3161.

Lancaster, A.K., Nutter-Upham, A., Lindquist, S., King, O.D. (**2014**). PLAAC: a web and command-line application to identify proteins with prion-like amino acid composition. *Bioinformatics, 30*, 2501–2502.

Lasch, P., Petras, T., Ullrich, O., Backmann, J., Naumann, D., & Grune, T. (**2001**). Hydrogen Peroxide-induced Structural Alterations of RNase A. *J Biol Chem, 276,* 9492–9502.

Le, K. Y., Dastgheyb, S., Ho, T. V., & Otto, M. (**2014**). Molecular determinants of staphylococcal biofilm dispersal and structuring. *Front Cell Inf Microbiol*, *167*, 1–7.

Leonard, S. E., & Carroll, K. S. (**2011**). Chemical "omics" approaches for understanding protein cysteine oxidation in biology. *Curr Opin Chem Biol*, *15*, 88–102.

LeVine, H. (**1993**). Thioflavine T interaction with synthetic Alzheimer's disease beta-amyloid peptides: detection of amyloid aggregation in solution. *Protein Sci*, *2*, 404–410.

Levine, R. L., Mosoni, L., Berlett, B. S., & Stadtman, E. R. (**1996**). Methionine residues as endogenous antioxidants in proteins. *Proc Natl Acad Sci USA*, *93*, 15036–15040.

Levinthal, C. (**1968**). Are there pathways for protein folding? *J Med Phys*, *65*, 44–45.

Li L. & Lindquist S. (**2000**). Creating a protein-based element of inheritance. *Science, 287*,661–664

Li, Y., Ogunnaike, B. a., & Roberts, C. J. (**2010**). Multi-variate approach to global protein aggregation behavior and kinetics: Effects of pH, NaCl, and temperature for ??-chymotrypsinogen A. *J Pharm Sci*, *99*, 645–662.

Li, Y.R., King, O.D., Shorter, J., Gitler, A.D. (**2013**). Stress granules as crucibles of ALS pathogenesis. *J Cell Biol, 201*, 361–372.

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.

Lindorff-Larsen, K., Rogen, P., Paci, E., Vendruscolo, M., and Dobson, C. M. (**2005**). Protein folding and the organization of the protein topology universe. *Trends Biochem Sci, 30*, 13–19

Liu, H.-L., Wu, Y.-C., Zhao, J.-H., Liu, Y.-F., Huang, C.-H., Fang, H.-W., & Ho, Y. (**2007**). Insights into the conformational changes of several human lysozyme variants associated with hereditary systemic amyloidosis. *Biotech Progr*, *23*, 246–54.

Lo Conte, M., & Carroll, K. S. (**2013**). The redox biochemistry of protein sulfenylation and sulfinylation. *J Biol Chem*, *288*, 26480–8.

Lodish, H., Berk, A., Zipursky, S.L., et al. (**2000**). Molecular Cell Biology. New York: W. H. Freeman

Lomakin, A., Chung, D. S., Benedek, G. B., Kirschner, D. A., & Teplow, D. B. (**1996**). On the nucleation and growth of amyloid beta-protein fibrils: detection of nuclei and quantitation of rate constants. *Proc Natl Acad Sci USA*, *93*, 1125–1129.

Luo, P. & Baldwin, R.L. (**1997**). Mechanism of helix induction by trifluoroethanol: A framework for extrapolating the helix-forming properties of peptides from trifluoroethanol/water mixtures back to water. *Biochemestry, 36*, 8413–8421.

Makin, O. S. & Serpell, L. C. **(2005**). Structures for amyloid fibrils. *FEBS J*, *272*, 5950–5961.

Malinchik, S. B., Inouye, H., Szumowski, K. E., & Kirschner, D. a. (**1998**). Structural analysis of Alzheimer's beta(1-40) amyloid: protofilament assembly of tubular fibrils. *Biophys J*, *74*, 537–545.

Malinovska, L., Kroschwald, S., Alberti, S. (**2013**). Protein disorder, prion propensities, and self-organizing macromolecular collectives. *Biochim Biophys Acta, 1834*, 918–931.

Mannini, B., Mulvihill, E., Sgromo, C., Cascella, R., Khodarahmi, R., Ramazzotti, M. et al (**2014**). Toxicity of protein oligomers is rationalized by a function combining size and surface hydrophobicity. *ACS Chem Biol, 9*, 2309–2317.

Marino, S. M., & Gladyshev, V. N. (**2012**). Analysis and functional prediction of reactive cysteine residues., *J Biol Chem , 287*, 4419–4425.

Maris, C., Dominguez, C., & Allain, F. H. T. (**2005**). The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J*, *272*, 2118–2131.

Martelli, P. L., Fariselli, P., Malaguti, L., & Casadio, R. (**2002**). Prediction of the disulfide bonding state of cysteines in proteins with hidden neural networks*. Protein Eng, 15*, 951–953.

Matsumura, M. & Matthews, B.W. (**1991**). Stabilization of functional proteins by introduction of multiple disulfide bonds. *Meth Enzymol, 202*,336–356.

Mattoo, R. U. H., & Goloubinoff, P. (**2014**). Molecular chaperones are nanomachines that catalytically unfold misfolded and alternatively folded proteins. *Cell Mol Life Sci*, *71*, 3311–3325.

Maurer-Stroh, S., Debulpaep, M., Kuemmerer, N., Lopez de la Paz, M., Martins, I. C., Reumers, J., … Rousseau, F. (**2010**). Exploring the sequence determinants of amyloid structure using position-specific scoring matrices. *Nature Meth*, *7*, 237–242.

Mehlin, C., Headley, C. M., & Klebanoff, S. J. (**1999**). An inflammatory polypeptide complex from Staphylococcus epidermidis: isolation and characterization. *J Exp Med*, *189*, 907–918.

Meng, F., Yao, D., Shi, Y., Kabakoff, J., Wu, W., Reicher, J., … Gu, Z. (**2011**). Oxidation of the cysteine-rich regions of parkin perturbs its E3 ligase activity and contributes to protein aggregation. *Mol Neurodegen*, *6*, 34.

Merz, P.A.; Wisniewski, H.M.; Somerville, R.A.; Bobin, S.A.; Masters, C.L.; Iqbal, K. (**1983**). Ultrastructural morphology of amyloid fibrilsfrom neuritic and amyloid plaques. *Acta Neuropathol.,60*, 113–124.

Middelberg, a. P. J. (**2002**). Preparative protein refolding. *Trends Biotechnol*, *20*, 437–443.

Milligan, S. a, Burke, P., Coleman, D. T., Bigelow, R. L., Steffan, J. J., Carroll, J. L., … Cardelli, J. a. (**2009**). The green tea polyphenol EGCG potentiates the antiproliferative activity of c-Met and epidermal growth factor receptor inhibitors in non-small cell lung cancer cells. *Clin Cancer Res*, *15*, 4885–4894.

Milzani, a, Rossi, R., Di Simplicio, P., Giustarini, D., Colombo, R., & DalleDonne, I. (**2000**). The oxidation produced by hydrogen peroxide on Ca-ATP-G-actin. *Protein Sci, 9*, 1774–1782.

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

Morell, M., Bravo, R., Espargaro A, Sisquella X, Aviles FX, Fernandez-Busquets X et al (**2008**). Inclusion bodies: specificity in their aggregation process and amyloid-like structure. *Biochim Biophys Acta, 1783*, 1815–1825.

Morris, K. & Serpell L. (**2010**). From natural to designer self-assembling biopolymers, the structural characterisation of fibrous proteins & peptides using fibre diffraction. *Chem Soc Rev, 39,* 3445-3453.

Mossuto, M. F., Bolognesi, B., Guixer, B., Dhulesia, A., Agostini, F., Kumita, J. R., … Salvatella, X. (**2011**). Disulfide bonds reduce the toxicity of the amyloid fibrils formed by an extracellular protein. *Angew Chem Int Ed, 50*, 7048–7051.

Mossuto, M. F., Dhulesia, A., Devlin, G., Frare, E., Kumita, J. R., de Laureto, P. P., … Salvatella, X. (**2010**). The Non-Core Regions of Human Lysozyme Amyloid Fibrils Influence Cytotoxicity. *J Mol Biol, 402*, 783–796.

Munishkina, L.A.; Phelan, C.; Uversky, V.N.; Fink, A.L. (**2003**). Conformational behavior and aggregation of alpha-synuclein in organic solvents: Modeling the effects of membranes. *Biochemestry, 42*, 2720–2730.

Myers, J.K.; Pace, C.N.; Scholtz, J.M. (**1998**). Trifluoroethanol effects on helix propensity and electrostatic interactions in the helical peptide from ribonuclease t1. *Protein Sci., 7*, 383–388.

Nagle, D. G., Ferreiraa, D., & Zhou, Y.-D. (**2006**). Epigallocatechin-3-gallate (EGCG): Chemical and biomedical perspectives. *Phytochem*, *67*, 1849–1855.

Nakanishi, T., Yoshioka, M., Moriuchi, K., Yamamoto, D., Tsuji,M., Takubo, T. (**2010**). S-sulfonation of transthyretin is an important trigger step in the formation of transthyretin-related amyloid fibril. *Biochim Biophys Acta, 1804*,1449-56

Narayan, M., Welker, E., Wedemeyer, W. J., & Scheraga, H. a. (**2000**). Oxidative folding of proteins. *Acc Chem Res*, *33*, 805–812.

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

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

Neumann, M., Bentmann, E., Dormann, D., Jawaid, A., DeJesus-Hernandez, M., Ansorge, O. et al (**2011**). FET proteins TAF15 and EWS are selective markers that distinguish FTLD with FUS pathology from amyotrophic lateral sclerosis with FUS mutations. *Brain, 134*, 2595–2609.

Nickson, A. a., Wensley, B. G., & Clarke, J. (**2013**). Take home lessons from studies of related proteins. *Curr Opin Struct Biol*, *23*(1), 66–74.

Notredame, C., Higgins, D. G., & Heringa, J. (**2000**). T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol, 302*, 205–217.

Nowotny, K., Jung, T., Grune, T., & Höhn, A. (**2014**). Accumulation of modified proteins and aggregate formation in aging. *Exp Gerontol*, *57*, 122–31.

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

Oli, M. W., Otoo, H. N., Crowley, P. J., Heim, K. P., Nascimento, M. M., Ramsook, C. B., … Brady, L. J. (**2012**). Functional amyloid formation by Streptococcus mutans. *Microbiol, 158*, 2903–2916.

Olofsson, A., Ostman, J. & Lundgren, E. (**2002**). Amyloid: morphology and toxicity. *Clin. Chem. Lab. Med.,40*, 1266-70.

Olzmann, J. a, Li, L., & Chin, L. S. (**2008**). Aggresome formation and neurodegenerative diseases: therapeutic implications. *Curr Med Chem*, *15*(1), 47–60.

Onuchic, J. N., & Wolynes, P. G. (**2004**). Theory of protein folding. *Curr Opin Struct Biol*, *14*, 70-75.

Onuchic, J. N., Luthey-Schulten, Z., & Wolynes, P. G. (**1997**). THEORY OF PROTEIN FOLDING: The Energy Landscape Perspective. *Annu Rev Phys Chem*, *48*, 545–600.

Onuchic, J. N., Socci, N. D., Luthey-Schulten, Z., & Wolynes, P. G. (**1996**). Protein folding funnels: the nature of the transition state ensemble. *Folding & Design*, 1, 441–450.

Otto, M. (**2008**). Staphylococcal biofilms. *Curr Top Microbiol Immunol*, *322*, 207–228.

Otto, M. (**2013**). Staphylococcal infections: mechanisms of biofilm maturation and detachment as critical determinants of pathogenicity. *Annu Rev Med, 64*, 175–188.

Otto, M. (**2014**). Phenol-soluble modulins. *Int J Med Microbiol*, *340*, 164–169.

Otto, M. (**2010**). Basis of virulence in community-associated methicillin-resistant Staphylococcus aureus, *Annu Rev Microbiol, 64,* 143–162.

Otzen, D. (**2010a**). Functional amyloid: turning swords into plowshares. *Prion*, *4*, 256–264.

Otzen, D. E., Kristensen, O., & Oliveberg, M. (**2000**). Designed protein tetramer zipped together with a hydrophobic Alzheimer homology: a structural clue to amyloid assembly. *Proc Natl Acad Sci USA, 97*, 9907–9912.

Otzen, D.E. (**2010b**). Amyloid formation in surfactants and alcohols: Membrane mimetics or structural switchers? *Curr Protein Pept Sci, 11*, 355–371.

Pace, C. N., Grimsley G.R., Thomson J.A., & Barnett B.J. (**1988**). Conformational Stability and Activity of Ribonuclease T1 with Zero, One, and Two Intact Disulfide Bonds. *J Biol Chem, 263*, 11820–11825.

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.

Papaleo, E., Sutto, L., Gervasio, F. L., & Lindorff-Larsen, K. (**2014**). Conformational Changes and Free Energies in a Proline Isomerase. *J Chem Theory Comput*, *10*, 4169–4174.

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

Paulsen, C. E., Truong, T. H., Garcia, F. J., Homann, A., Gupta, V., Leonard, S. E., & Carroll, K. S. (**2011**). Peroxide-dependent sulfenylation of the EGFR catalytic site enhances kinase activity. *Nat Chem Biol*, *8*, 57–64.

Pawar, A. P., DuBay, K. F., Zurdo, J., Chiti, F., Vendruscolo, M., & Dobson, C. M. (**2005**). Prediction of "aggregation-prone" and "aggregation- susceptible" regions in proteins associated with neurodegenerative diseases. *J Mol Biol*, *350*, 379–392.

Pechmann, S., Levy, E. D., Tartaglia, G. G., & Vendruscolo, M. (**2009**). Physicochemical principles that regulate the competition between functional and dysfunctional association of proteins., *Proc Natl Acad Sci USA, 106*, 10159–10164.

Peggion, C., Sorgato, M.C., Bertoli, A. (**2014**). Prions and prion-like pathogens in neurodegenerative disorders. *Pathogens, 3*, 149–163.

Pepys, M. B., Hawkins, P. N., Booth, D. R., Vigushin, D. M., Tennent, G. a, Soutar, a K., … Terry, C. J. (**1993**). Human lysozyme gene mutations cause hereditary systemic amyloidosis. *Nature*, *362*, 553–557.

Periasamy, S., Chatterjee, S. S., Cheung, G. Y. C., & Otto, M. (**2012a**). Phenol-soluble modulins in staphylococci: what are they originally for? *Commun Integr Biol*, *5*, 275–277.

Periasamy, S., Joo, H.-S., Duong, a. C., Bach, T.-H. L., Tan, V. Y., Chatterjee, S. S., … Otto, M. (**2012b**). How Staphylococcus aureus biofilms develop their characteristic structure. *Proc Natl Acad Sci USA, 109*, 1281–1286.

Peschel A, Otto M (**2013**). Phenol-soluble modulins and staphylococcal infection. *Nat Rev Microbiol*, *11*, 667–673.

Petsko, G.A. & Ringe, D. (**2004**). Protein structure and function, *New Science Press*

Pfefferkorn, C. M., McGlinchey, R. P., & Lee, J. C. (**2010**). Effects of pH on aggregation kinetics of the repeat domain of a functional amyloid, Pmel17. *Proc Natl Acad Sci USA, 107*, 21447–21452.

Pham, C. L., Kwan, a H., & Sunde, M. (**2014**). Functional amyloid: widespread in Nature, diverse in purpose. *Essays Biochem*, *56*, 207–219.

Pinkus, J.L., Amato, A.A., Taylor, J.P., Greenberg, S.A. (**2014**). Abnormal distribution of heterogeneous nuclear ribonucleoproteins in sporadic inclusion body myositis. *Neuromuscul Disord, 24*, 611–616.

Pouplana, S., Espargaro, A., Galdeano, C., Viayna, E., Sola, I., Ventura, S. et al (**2013**). Thioflavin-S staining of bacterial inclusion bodies for the fast, simple, and inexpensive screening of amyloid aggregation inhibitors. *Curr Med Chem 21,* 1152–1159.

Prilusky, J., Felder, C.E., Zeev-Ben-Mordehai, T., Rydberg, E.H., Man, O., Beck-mann, J.S. (**2005**). FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics, 21*, 3435–3438.

Privalov, P. L. (**1979**). Stability of proteins: small globular proteins. *Adv Prot Chem*, *33*, 167-241

Prusiner, S. B. (**1998**). Prions. *Proc Natl Acad Sci USA*, *95*, 13363–13383.

Quintas, a. (**2013**). What drives an amyloid protein precursor from an amyloidogenic to a native-like aggregation pathway ? *OA Biochem*, *1*, 1–8.

Quintas, A., Saraiva, M. J. M., & Britot, R. M. M. (**1999**). The tetrameric protein transthyretin dissociates to a non-native monomer in solution. A novel model for amyloidogenesis. *J Biol Chem*, *274*, 32943–32949.

Rahal, A., Kumar, A., Singh, V., Yadav, B., Tiwari, R., Chakraborty, S., & Dhama, K. (**2014**). Oxidative stress, prooxidants, and antioxidants: The interplay. *BioMed Res Int*,1-19

Rajan, R. S., Illing, M. E., Bence, N. F., & Kopito, R. R. (**2001**). Specificity in intracellular protein aggregation and inclusion body formation. *Proc Natl Acad Sci USA*, *98*, 13060–13065.

Rakoczy, E.P., Kiel, C., McKeone, R., Stricher, F., Serrano, L. (**2011**) Analysis of disease-linked rhodopsin mutations based on structure, function, and protein stability calculations. *J Mol Biol, 405*, 584–606.

Ramamoorthy, A. (**2013**). Bacterial curli protein promotes the conversion of pap248–286 into the amyloid sevi: Cross-seeding of dissimilar amyloid sequences. *Peer J., 1*, e5.

Rasigade, J. P., Trouillet-Assant, S., Ferry, T., Diep, B. A., Sapin, A., Lhoste, Y., … Laurent, F. **(2013**). PSMs of Hypervirulent Staphylococcus aureus Act as Intracellular Toxins That Kill Infected Osteoblasts. *PLoS One*, *8*.

Rattenholl, A., Lilie, H., Grossmann, A., Stern, A., Schwarz, E., & Rudolph, R. (**2001**). The pro-sequence facilitates folding of human nerve growth factor from Escherichia coli inclusion bodies. *Europ J Biochem*, *268*, 3296–3303.

Rezai-Zadeh, K., Shytle, D., Sun, N., Mori, T., Hou, H., Jeanniton, D., … Tan, J. (**2005**). Green tea epigallocatechin-3-gallate (EGCG) modulates amyloid precursor protein cleavage and reduces cerebral amyloidosis in Alzheimer transgenic mice. *The Journal of Neuroscience : J Neurosci, 25*, 8807–8814.

Rhee, S. G., Jeong, W., Chang, T.-S., & Woo, H. a. (**2007**). Sulfiredoxin, the cysteine sulfinic acid reductase specific to 2-Cys peroxiredoxin: its discovery, mechanism of action, and biological significance. *Kidney Int Suppl*, *106*, S3–S8.

Romero, P., Obradovic, Z., Li, X., Garner, E. C., Brown, C. J., & Dunker, a. K. (**2001**). Sequence complexity of disordered protein. *Proteins*, *42*, 38–48.

Roos, G., Foloppe, N., & Messens, J. (**2013**). Understanding the pKa of Redox Cysteines: The Key Role of Hydrogen Bonding. *Antiox Red Signal*, *18*, 94–127.

Rousseau, F., Serrano, L., Schymkowitz, J. W. (**2006**) How evolutionary pressure against protein aggregation shaped chaperone specificity*. J Mol Biol* , *355*, 1037-47.

Rubin, R. J., Harrington, C. a., Poon, a., Dietrich, K., Greene, J. a., & Moiduddin, a. (**1999**). The economic impact of Staphylococcus aureus infection in New York City hospitals. *Emerg Infect Dis 5*, 9–17.

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

Sabate, R., de Groot, N.S., Ventura, S. (**2010**). Protein folding and aggregation in bacteria. *Cell Mol Life Sci, 67*, 2695–2715.

Sabate, R., Espargaro, A., Grana-Montes, R., Reverter, D., Ventura, S. (**2012a**). Native structure protects SUMO proteins from aggregation into amyloid fibrils. *Biomacromolecules, 13*, 1916–1926.

Sabate, R., Rousseau, F., Schymkowitz, J., & Ventura, S. (**2015**). What Makes a Protein Sequence a Prion? *PLoS Comput Biol*, *11*.

Sabate, R., Villar-Pique, A., Espargaro, A., Ventura, S. (**2011**). Temperature dependence of the aggregation kinetics of Sup35 and Ure2p yeast prions. *Biomacromolecules 13*, 474–483.

Sabate, R., Villar-Pique, A., Espargaro, A., Ventura, S. (**2012b**). Temperature dependence of the aggregation kinetics of sup35 and ure2p yeast prions. *Biomacromolecules, 1*3, 474–483.

Sakono, M. & Zako, T. (**2010**). Amyloid oligomers: Formation and toxicity of Aβ oligomers. *FEBS J*, *277*, 1348–1358.

Sánchez de Groot, N., Pallarés, I., Avilés, F. X., Vendrell, J., & Ventura, S. (**2005**). Prediction of “hot spots” of aggregation in disease-linked polypeptides. *BMC Struct Biol*, *5*, 18.

Sanchez de Groot, N., Torrent, M., Villar-Piqué, A., Lang, B., Ventura, S., Gsponer, J., & Babu, M. M. (**2012**). Evolutionary selection for protein aggregation. *Biochem Soc Trans*, *40*, 1032–7.

Sarroukh, R., Goormaghtigh, E., Ruysschaert, J. M., & Raussens, V. (**2013**). ATR-FTIR: A "rejuvenated" tool to investigate amyloid proteins. *BBA- Biomembranes, 1828*, 2328–2338.

Sawaya, M. R., Sambashivan, S., Nelson, R., Ivanova, M. I., Sievers, S. a, Apostol, M. I., … Eisenberg, D. (**2007**). Atomic structures of amyloid cross-beta spines reveal varied steric zippers. *Nature*, *447*, 453–457.

Sawyer, E. B., Claessen, D., Gras, S. L., & Perrett, S. (**2012**). Exploiting amyloid: how and why bacteria use cross-β fibrils. *Biochem Soc Trans*, *40*, 728–734.

Schwartz, K., Syed, A. K., Stephenson, R. E., Rickard, A. H., & Boles, B. R. (**2012**). Functional amyloids composed of phenol soluble modulins stabilize Staphylococcus aureus biofilms. *PLoS Pathog*, *8*, 1-11.

Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F., & Serrano, L. (**2005a**). The FoldX web server: An online force field. *Nucleic Acids Res*, *33*, 382–388.

Seeliger, J.; Estel, K.; Erwin, N.; Winter, R. (**2013**).Cosolvent effects on the fibrillation reaction of human iapp. *Phys. Chem Chem Phys,* 15, 8902–8907.

Sena, L. a., & Chandel, N. S. (**2012**). Physiological roles of mitochondrial reactive oxygen species. *Mol Cell*, *48*, 158–166.

Sevier, C. S., & Kaiser, C. a. (**2002**). Formation and transfer of disulphide bonds in living cells. *Nat Rev Mol Cell Biol*, *3*, 836–847.

Shacter, E. (**2000**). Quantification and significance of protein oxidation in biological samples. *Drug Metab Rev, 32*, 307–26.

Shirahama, T., & Cohen, A. S. (**1967**). High-resolution electron microscopic analysis of the amyloid fibril. *J Cell Biol, 33*, 679–708.

Shivu, B., Seshadri, S., Li, J., Oberg, K.A., Uversky, V.N., Fink, A.L. (**2013**). Distinct beta-sheet structure in protein aggregates determined by ATR-FTIR spectroscopy. *Biochemestry, 52*, 5176–5183.

Sies, H. (**1993**). Strategies of antioxidant defense. *Eur J Biochem / FEBS*, *215*, 213–219.

Sipe, J. D., & Cohen, a S. (**2000**). Review: history of the amyloid fibril. *J Struct Biol*, *130*, 88–98.

Šmerák, P., Šestáková, H., Polívková, Z., Štětina, R., Langová, M., Bárta, I., … Bártová, J. (**2006**). Antimutagenic effect of epigallocatechin gallate and its effect on the immune response in mice. *Czech J Food Sci*, *24*, 180–192.

Smith, C. D., Carney, J. M., Starke-Reed, P. E., Oliver, C. N., Stadtman, E. R., Floyd, R. a, & Markesbery, W. R. (**1991**). Excess brain protein oxidation and enzyme dysfunction in normal aging and in Alzheimer disease. *Proc Natl Acad Sci USA, 88*, 10540–10543.

Smith, D. P., & Radford, S. E. (**2001**). Role of the single disulphide bond of beta(2)-microglobulin in amyloidosis in vitro. *Protein Sci, 10,* 1775–1784.

Sørensen, H. P., & Mortensen, K. K. (**2005**). Soluble expression of recombinant proteins in the cytoplasm of Escherichia coli. *Microb Cell Fact, 4*, 1.

Soto, C. (**2013**). Protein Misfolding in Neurodegenerative Diseases : The Key Pending Questions, 19–22.

Soto, C., Saborio, G. P., & Anderes, L. (**2002**). Cyclic amplification of protein misfolding: application to prion-related disorders and beyond. *Trends Neurosci*, *25*, 390–394.

Squier, T.C . (**2001**). Oxidative stress and protein aggregation during biological aging *Exp Gerontol, 36*, 1539–1550.

Srisailam, S.; Kumar, T.K.; Rajalingam, D.; Kathir, K.M.; Sheu, H.S.; Jan, F.J.; Chao, P.C.; Yu, C. (**2003**). Amyloid-like fibril formation in an all beta-barrel protein. Partially structured intermediate state(s) is a precursor for fibril formation. *J Biol Chem, 278,* 17701–17709.

Stadtman, E. R. & Levine, R. L. (**2000**). Protein oxidation*. Ann N Y Acad Sci, 899*, 191–208.

Stadtman, E.R., Van Remmen, H., Richardson, A., Wehr, N.B., Levine, R.L. (**2005**). Methionine oxidation and aging. *Biochim Biophys Acta,* 1703–135.

Stefani, M. (**2004**). Protein misfolding and aggregation: New examples in medicine and biology of the dark side of the protein world. *BBA - Mol Basis Dis*, *1739*, 5–25.

Sun, H., Gao, J., Ferrington, D.A., Biesiada, H., Williams, T.D., et al. (**1999**). Repair of oxidized calmodulin by methionine sulfoxide reductase restores ability to activate the plasma membrane Ca-ATPase. *Biochemestry, 38*, 105–112.

Swaminathan, R., Ravi, V. K., Kumar, S., Kumar, M. V. S., & Chandra, N. (**2011**). Lysozyme: A model protein for amyloid research. *Adv Prot Chem Struct Biol*, *84*, 63-111.

Taddei, N.; Capanni, C.; Chiti, F.; Stefani,M.; Dobson, C.M.; Ramponi, G. (**2001**). Folding and aggregation are selectively influenced by the conformational preferences of the alpha-helices of muscle acylphosphatase. *J Biol Chem, 276*, 37149–37154.

Takagi, H., Takahashi, T., Momose, H., Inouye, M., Maeda, Y., Matsuzawa, H., & Ohta, T. (**1990**). Enhancement of the thermostability of subtilisin E by introduction of a disulfide bond engineered on the basis of structural comparison with a thermophilic serine protease. *J Biol Chem*, *265*(12), 6874–6878.

Tartaglia, G. G. & Vendruscolo, M. (**2008**). The Zyggregator method for predicting protein aggregation propensities. *Chem Soc Rev, 37*, 1395–1401.

Tartaglia, G. G., Cavalli, A., & Vendruscolo, M. (**2007**). Prediction of Local Structural Stabilities of Proteins from Their Amino Acid Sequences. *Structure, 15*, 139–143.

Teng, P. K., & Eisenberg, D. (**2009**). Short protein segments can drive a non-fibrillizing protein into the amyloid state*. Protein Eng Des Sel*, *22*, 531–536.

Thomas, M.P., Chartrand, K., Reynolds, A., Vitvitsky, V., Banerjee, R., Gendelman, H.E. (**2007**) Ion channel blockade attenuates aggregated alpha synuclein induction of microglial reactive oxygen species: relevance for the pathogenesis of Parkinson's disease. *J Neurochem, 100*, 503–19.

Thompson, M. J., Sievers, S. a, Karanicolas, J., Ivanova, M. I., Baker, D., & Eisenberg, D. (**2006**). The 3D profile method for identifying fibril-forming segments of proteins. *Proc Natl Acad Sci USA*, *103*, 4074–4078.

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

Toombs, J.A., Petri, M., Paul, K.R., Kan, G.Y., Ben-Hur, A., Ross, E.D. (**2012**). De novo design of synthetic prion domains. *Proc Natl Acad Sci USA ,109*, 6519–6524.

Trivedi, M. V, Laurence, J. S., & Siahaan, T. J. (**2009**). The role of thiols and disulfides on protein stability. *Current Protein & Peptide Science*, *10*, 614–625.

Trovato, A., Chiti, F., Maritan, A., & Seno, F. (**2006**). Insight into the structure of amyloid fibrils from the analysis of globular proteins. *PLoS Comput Biol*, *2*, 1608–1618.

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

Tsolis, A. C., Papandreou, N. C., Iconomidou, V. a., & Hamodrakas, S. J. (**2013**). A Consensus Method for the Prediction of "Aggregation-Prone" Peptides in Globular Proteins. *PLoS One*, *8*, 1–6.

Turrens, J. F. (**2003**). Mitochondrial formation of reactive oxygen species. *J Physiol*, *552*, 335–344.

Tycko, R. (**2011**). Solid state NMR studies of amyloid fibril structure. *Ann Rev Phys Chem*, *62*, 279–299.

Tyther, R., Ahmeda, A., Johns, E., McDonagh, B., & Sheehan, D. (**2010**). Proteomic profiling of perturbed protein sulfenation in renal medulla of the spontaneously hypertensive rat. *J Prot Res*, *9*, 2678–2687.

Uversky, V. N. (**2008**). Amyloidogenesis of natively unfolded proteins. *Curr Alzheimer Res*, *5*, 260–287.

Uversky, V. N., & Dunker, a. K. (**2010**). Understanding protein non-folding. *BBA – Proteins Proteomics*, *1804*, 1231–1264.

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

Uversky, V. N., Lee, H. J., Li, J., Fink, A. L., & Lee, S. J. (**2001**). Stabilization of Partially Folded Conformation during ??-Synuclein Oligomerization in Both Purified and Cytosolic Preparations. *J Biol Chem*, *276*, 43495–43498.

Uversky, V. N., Oldfield, C. J., Midic, U., Xie, H., Xue, B., Vucetic, S., … Dunker, a K. (**2009**). Unfoldomics of human diseases: linking protein intrinsic disorder with diseases. *BMC Genomics*, *10.*

Ventura, S. (**2005**). Sequence determinants of protein aggregation: tools to increase protein solubility. *Microb Cell Fact*, *4*, 11.

Ventura, S. & Villaverde, A. (**2006**). Protein quality in bacterial inclusion bodies. *Trends Biotechnol*, *24*, 179–185.

Ventura, S., Zurdo, J., Narayanan, S., Parreño, M., Mangues, R., Reif, B., … Serrano, L. (**2004**). Short amino acid stretches can mediate amyloid formation in globular proteins: the Src homology 3 (SH3) case. *Proc Natl Acad Sci USA ,101*, 7258–7263.

Vieira, N.M., Naslavsky, M.S., Licinio, L., Kok, F., Schlesinger, D., Vainzof, M. et al (**2014**). A defect in the RNA-processing protein HNRPDL causes limb-girdle muscular dystrophy 1G (LGMD1G). *Hum Mol Genet, 23*, 4103–4110

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. *J Mol Biol, 421*, 270.

Villaverde, A., García-Fruitós, E., Rinas, U., Seras-Franzoso, J., Kosoy, A., Corchero, J. L., & Vazquez, E. (**2012**). Packaging protein drugs as bacterial inclusion bodies for therapeutic applications. *Microb Cell Fact, 11*, 76.

Villegas, V., Zurdo, J., Filimonov, V. V, Avilés, F. X., Dobson, C. M., & Serrano, L. (**2000**). Protein engineering as a strategy to avoid formation of amyloid fibrils. *Protein Sci*, *9*, 1700–1708.

Vivekanandan, S., Brender, J.R., Lee, S.Y., Ramamoorthy, A. (**2011**). A partially folded structure of amyloid-beta (1–40) in an aqueous environment. *Biochem Biophy Res Commun, 411*, 312–316.

Wang, I., Hennig, J., Jagtap, P.K., Sonntag, M., Valcarcel, J., Sattler, M. (**2014**). Structure, dynamics and RNA binding of the multi-domain splicing factor TIA-1. *Nucleic Acids Res, 42*, 5949–5966.

Wang, L., Maji, S. K., Sawaya, M. R., Eisenberg, D., & Riek, R. (**2008**). Bacterial inclusion bodies contain amyloid-like structure. *PLoS Biol*, *6*, 1791–1801.

Wang, R., Braughton, K. R., Kretschmer, D., Bach, T.-H. L., Queck, S. Y., Li, M., … Otto, M. (**2007**). Identification of novel cytolytic peptides as key virulence determinants for community-associated MRSA. *Nature Med*, *13*, 1510–1514.

Ward, J. J., Sodhi, J. S., McGuffin, L. J., Buxton, B. F., & Jones, D. T. (**2004**). Prediction and Functional Analysis of Native Disorder in Proteins from the Three Kingdoms of Life. *J Mol Biol*, *337*, 635–645.

Watt, B., Van Niel, G., Raposo, G., & Marks, M. S. (**2013**). PMEL: A pigment cell-specific model for functional amyloid formation. *Pigment Cell Melanoma Res*, *26*, 300–315.

Wedemeyer, W. J., Welker, E., Narayan, M., & Scheraga, H. a. (**2000**). Disulfide bonds and protein folding. *Biochemistry*, *39*, 4207–4216.

Wetlaufer, D. B. (**1973**). Nucleation, rapid folding, and globular intrachain regions in proteins. *Proc Natl Acad Sci USA*, *70*, 697–701.

Widlansky, M. E., Hamburg, N. M., Anter, E., Holbrook, M., Kahn, D. F., Elliott, J. G., … Vita, J. a. (**2007**). Acute EGCG supplementation reverses endothelial dysfunction in patients with coronary artery disease. *J Am Coll Nutr*, *26*, 95–102.

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

Wolozin, B. (**2012**). Regulated protein aggregation: stress granules and neurodegeneration. *Mol Neurodegener, 7*, 56.

Wong, Y. Q., Binger, K. J., Howlett, G. J., & Griffin, M. D. W. (**2010**). Methionine oxidation induces amyloid fibril formation by full-length apolipoprotein A-I. *Proc Natl Acad Sci USA, 107*, 1977–1982.

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

Wright, P. E., & Dyson, H. J. (**2015**). Intrinsically Disordered Proteins in Cellular Signaling and Regulation. *Nat Rev Mol Cell Biol*, *16*, 18–29.

Wurth C, Guimard NK & Hecht MH (**2002**). Mutations that reduce aggregation of the Alzheimer's Abeta42 peptide: an unbiased search for the sequence determinants of Abeta amyloidogenesis. *J Mol Biol, 319*, 1279–1290.

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

Yarnell, A. (**2009**). Cysteine Oxidation: New chemical tools are poised to help scientists explore the roles oxidized cysteine residues might play in biology. *Chem Eng News*, *87*, 38-40.

Yazaki, M., Farrell, S. a., & Benson, M. D. (**2003**). A novel lysozyme mutation Phe57Ile associated with hereditary renal amyloidosis. *Kidney Int*, *63*, 1652–1657.

Yoshimura, Y., Lin, Y., Yagi, H., Lee, Y.-H., Kitayama, H., Sakurai, K., … Goto, Y. (**2012**). Distinguishing crystal-like amyloid fibrils and glass-like amorphous aggregates from their kinetics of formation. *Proc Natl Acad Sci USA*, *109*, 14446–14451.

Zambrano, R., Conchillo-Sole, O., Iglesias, V., Illa, R., Rousseau, F., Schymkowitz, J., … Ventura, S. (**2015b**). PrionW: a server to identify proteins containing glutamine/asparagine rich prion-like domains and their amyloid cores. *Nucleic Acids Res*, 1–7.

Zambrano, R., Jamroz, M., Szczasiuk, A., Pujols, J., Kmiecik, S., & Ventura, S. (**2015a**). AGGRESCAN3D (A3D): server for prediction of aggregation properties of protein structures. *Nucleic Acids Res*, *43*, 306–313.

Zavodszky, M., Chen, C. W., Huang, J. K., Zolkiewski, M., Wen, L., & Krishnamoorthi, R. (**2001**). Disulfide bond effects on protein stability: designed variants of Cucurbita maxima trypsin inhibitor-V. *Protein Sci*, *10*, 149–160.

Zhang, T., Faraggi, E., Li, Z., & Zhou, Y. (**2013**). Intrinsically Semi-disordered State and Its Role in Induced Folding and Protein Aggregation. *Cell Biochem Biophys*, *67*, 1193–1205.

Zhou, H.X. (**2003**). Effect of backbone cyclization on protein folding stability: chain entropies of both the unfolded and the folded states are restricted. *J Mol Biol, 332*,257–264.

Zhou, X., Mester, C., Stemmer, P.M. & Reid, G.E. (**2014**).Oxidation-Induced Conformational Changes in Calcineurin Determined by Covalent Labeling and Tandem Mass Spectrometry. *Biochemistry, 53*,6754–6765.

Zibaee, S., Makin, O. S., Goedert, M., & Serpell, L. C. (**2007**). A simple algorithm locates b-strands in the amyloid fibril core of a-synuclein , Ab, and tau using the amino acid sequence alone. *Protein Sci*, *16*, 906–918.

Zorov, D. B., Bannikova, S. Y., Belousov, V. V., Vyssokikh, M. Y., Zorova, L. D., Isaev, N. K., … Plotnikov, E. Y. (**2005**). Reactive oxygen and nitrogen species: Friends or foes? *Biochemistry (Moscow)*, *70*, 215–221.

Zurdo, J.I. ,Guijarro & C.M. Dobson. (**2001**). Preparation and characterization of purified amyloid fibrils*, J Am Chem Soc, 123*, 8141–8142.

Zwanzig, R., Szabo, A., & Bagchi, B. (**1992**). Levinthal's paradox, *Proc Natl Acad Sci USA, 89,*20–22.

PART VIII - List of scientific works

This thesis is based on the following published works:

Marinelli, P., Castillo, V., & Ventura, S. (**2013**). **Trifluoroethanol modulates amyloid formation by the all α-helical URN1 FF domain.** *International Journal of Molecular Sciences, 14*, 17830-44.

Navarro, S., Marinelli, P., Diaz-Caballero, M., & Ventura, S. (**2015**). **The prion-like RNA-processing protein HNRPDL forms inherently toxic amyloid-like inclusion bodies in bacteria.** *Microbial Cell Factories, 14*, 102.

Unpublished works included in the thesis:

Marinelli, P. & Ventura, S. **Dissecting the contribution of *S. aureus* αPSMs to biofilm amyloid structure and toxicity.**

Marinelli, P. et al. **Polypeptide chain cross-linking by a novel disulfide bond affects the stability, folding and amyloid fibril formation capability of the all-α FF domain.**

Marinelli, P. et al. **Cysteine specific post-translational sulfonation reduces FF domain stability and promotes amyloid fibrils formation.**

Published works not included in the thesis:

Graña-Montes*****, R., Marinelli*****, P., Reverter, D., & Ventura, S. (**2014**). **N-Terminal Protein Tails Act as Aggregation Protective Entropic Bristles: The SUMO Case.** *Biomacromolecules, 15*, 4.

***** These authors contributed equally