Universitat Autònoma de Barcelona

Prion inspired nanomaterials and their biomedical applications

Weiqiang Wang

Universitat Autònoma de Barcelona

Institut de Biotecnologia i Biomedicina

and

Departament de Bioquímica i Biologia Molecular

Prion inspired nanomaterials and their biomedical applications

Weiqiang Wang

Bellaterra, August 2020

Prion inspired nanomaterials and their biomedical applications

Doctoral thesis submitted by Weiqiang Wang for degree of Ph.D in Biochemistry, Molecular Biology and Biomedicine

Universitat Autònoma de Barcelona

Institut de Biotecnologia i Biomedicina

and

Departament de Bioquímica i Biologia Molecular

Weiqiang Wang

Prof. Salvador Ventura Zamora

Dra. Susanna Navarro Cantero

Bellaterra, August 2020

Summary in English

Amyloids display a highly ordered fibrillar structure. Many of these assemblies appear associated with human disease. However, the controllable, stable, tunable, and robust nature of amyloid fibrils can be exploited to build up remarkable nanomaterials with a wide range of applications.

Functional prions constitute a particular class of amyloids. These transmissible proteins exhibit a modular architecture, with a disordered prion domain responsible for the assembly and one or more globular domains that account for the activity. Importantly, the original globular protein can be replaced with any protein of interest, without compromising the fibrillation potential. These genetic fusions form fibrils in which the globular domain remains folded, rendering functional nanostructures. However, in many cases, steric hindrance restricts the activity of these fibrils. This limitation can be solved by dissecting prion domains into shorter sequences that keep their self-assembling properties while allowing better access to the protein in the fibrillar state.

In this PhD thesis, we exploited the "soft amyloid core (SAC)" of the Sup35p yeast prion as a modular self-assembling unit, which recapitulates the aggregation propensity of the complete prion domain. We fused the SAC to different globular proteins of interest differing in conformation and sizes, building up a general and straightforward genetic approach to generate nanofibrils endowed with desired functionalities. Computational modeling allowed us to gain insights into the relationship between the size of the globular domains and the length of the linker that connects them to the SAC, providing the basis for the design of nanomaterials with different mesoscopic properties, either nanofibrils or nanoparticles. On this basis, we designed and produced, for the first time, highly active, non-toxic, spherical amyloid nanoparticles of defined size and engineered bifunctional nanostructures with application in targeted drug delivery. The lessons learned in these exercises resulted in the construction of a bispecific antibody-like nanofibril, showing potential in immunotherapy. In summary, the prion-like functional nanomaterials described here take profit of the genetic fusion approach to render a novel set of structures with application in biomedicine and biotechnology.

Resum en català

Els amiloides presenten una estructura fibril·lar molt ordenada. Molts d'aquests conjunts de proteïnes apareixen associats a malalties humanes. No obstant això, es pot aprofitar la naturalesa controlable, estable, ajustable i robusta de les fibres amiloides per crear nanomaterials amb una àmplia gamma d'aplicacions.

Els prions funcionals constitueixen una classe particular d'amiloides. Aquestes proteïnes transmissibles presenten una arquitectura modular, amb un domini prió desordenat responsable del assemblatge i d'un o més dominis globulars que proporcionen l'activitat. És important destacar que la proteïna globular original es pot substituir per qualsevol proteïna d'interès, sense comprometre el potencial de fibril·lació. Aquestes fusions genètiques formen fibres en les quals el domini global roman plegat, formant nanoestructures funcionals. Tot i això, en molts casos, els impediments estèrics poden restringir l'activitat d'aquestes fibres. Aquesta limitació es pot solucionar disseccionant els dominis priònics en seqüències més curtes que mantenen les seves propietats d'auto-assemblatge alhora que permeten un millor accés a la proteïna en estat fibril·lar.

En aquesta tesi doctoral, vam aprofitar el "soft amyloid core" (SAC) del prió de llevat Sup35p com una unitat de muntatge modular, que recapitula la propensió a l'agregació del domini priònic complet. Vam fusionar el SAC amb diferents proteïnes globulars d'interès que difereixen en la conformació i la mida, creant un mètode genètic general i senzill per generar nanofibres dotades de les funcionalitats desitjades. El modelatge computacional ens va permetre conèixer la relació entre la mida dels dominis globulars i la longitud del enllaç que els connecta al SAC, proporcionant les bases per al disseny de nanomaterials amb diferents propietats mesoscòpiques, ja siguin nanofibres o nanopartícules. Sobre aquesta base, hem dissenyat i produït, per primera vegada, nanopartícules amiloides esfèriques altament actives, no tòxiques, de mida definida, i s'han produït nanoestructures bifuncionals amb aplicació en el subministrament específic de fàrmacs. Les lliçons apreses en aquests exercicis van

donar lloc a la construcció d'una nanofibrilla similar a un anticòs biespecífic amb potencial per la immunoteràpia.

En resum, els nanomaterials funcionals de tipus priònic descrits aquí aprofiten l'enfocament de la fusió genètica per crear un nou conjunt d'estructures amb aplicacions en biomedicina i biotecnologia.

Content

List of abbreviations

ALS	Amyotrophic lateral sclerosis
ANOVA	Analysis of variance
AP	Alkaline phosphatase
AR	Aggregation-prone region
ATR-FTIR	Attenuated total reflectance Fourier transform infrared spectroscopy
BAP	Biotin acceptor peptide
bFPNw	Bifunctional protein nanowires
BirA	Biotin holoenzyme synthetase
BsAbs	Bispecific antibodies
BTB	Bromothymol blue
CA	Carbonic anhydrase
CD	Circular dichroism
CG	Coarse-grained
CR	Congo red
DHF	Dihydrofolate
DHFR	Dihydrofolate reductase
DLS	Dynamic light scattering
DNA	Deoxyribonucleic acid
EGFR	Epidermal growth factor receptor
ELISA	Enzyme linked immunosorbent assay
EM	Electron microscopy
EPR	Enhanced permeability and Retention
ER	Endoplasmic reticulum
GFP	Green fluorescent protein
GST	Glutathione S-transferase
HA	Hemagglutinin
HRP	Horseradish peroxidase
HSPs	Heat shock proteins
IgG	Immunoglobulin G
IMAC	Immobilized metal ion affinity chromatography
mAb	Monoclonal antibody
MP	Methyl parathion
MPH	Methyl-parathion hydrolase
mRNA	Messenger RNA
MTX	Methotrexate
NADP	Nicotinamide adenine dinucleotide phosphate
NMR	Nuclear magnetic resonance
PAI-1	Plasminogen activator inhibitor 1
PDB	Protein data bank
PK	Proteinase K
PrD	Prion domain
PrLDs	Prion-like domains
PrP	Prion protein
Rd	Rubredoxin
RNA	Ribonucleic acid
SACs	Soft amyloid cores
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEM	Scanning electron microscopy
SpA	*Staphylococcus aureus* protein A
SPG	Protein G from *Streptococcus*
STEM	Scanning transmission electron microscope
SUMO	Small ubiquitin-like modifier
TEM	Transmission electron microscope
THF	Tetrahydrofolate
Th-T	Thioflavin-T
TMD	Targeted molecular dynamics
tRNA	Transfer RNA
UV	Ultraviolet
ZZ	Z-domain dimer

Acknowledgements

First of all, I would like to appreciate my supervisor Salvador Ventura. I sincerely thank you for the great contributions to my publications, thesis and PhD training. I am very pleased that I got lots of theoretical and experimental knowledges, ranging from protein folding, aggregation to self-assembled nanomaterials, under your patient guide.

Then, I would like to thank my co-supervisor Susanna Navarro, for the kind help on my thesis but also many experimental technics involved in my PhD thesis. I would also thank Dr. Andrey and Dr. Rafayel for the great contributions to the second chapter.

I am also grateful for the colleagues from PPMC-EP: Anita, Manuel, Marta, Cri, Cristina, Marcos, Molood, Natalia, Jara, Helena, Bin and Ying. Especially, Anita helped me a lot for initiating my PhD project. I appreciated that all of you helped me integrate into a new environment at the beginning of my PhD studying and helped me for some experimental tricks. I am very pleased for the time we spend together in the laboratory, especially for the Fondue parties.

I also would like to thank the colleagues from Pros: Irantzu, Chari, Valentin, Jordi, Samuel, Francisca and Jaime. Thanks for the time we spend in the group meeting and the exchange of the ideas and knowledges. I also had a deep impression for the activities we joined together, such as anniversary party of UAB, food and beer seminars and also the barbecue party in Gironella.

I also want to thank the technicians from SCAC in IBB: Olga, Fran, Manuela, Arais, Roger and Carles, and the technicians from Microscopy service: Marti, Nuria, and Helena, not only for the technical support but also for many convenience during the Covid-19 pandemic.

For the GTS big family: Mon, Mari, Maria Angles, Vero, Manu, Sergi, Jorge, Sandra, Clara, Albert, Iris, Victor, Elena, Jingjing, Jun, Lou and Dong, I would appreciate you that you help Tingting a lot not only in the lab but also in daily life. I

also appreciate for the time we spend together, which will be the most precious treasures in my life. Thanks to your kind and friendly organizations, Tingting and I have the opportunities to visit many beautiful cities, such as Miranda, San Sebastian, Victoria, Logrono, Zaragoza, the delta of Ebro and even the small village Ribaforada. Many experiences like eating pinchos, seeing the Secuoya, skiing for the first and only time, seeing the Flamengo (the drama and the bird), etc. are very impressive and memorable even when I am recalling them at this moment. These experiences not only make our life fantastic but also make us see the authentic Spanish culture.

I would also thank some Chinese friends I met and visit here: Jie Ji, Lu Lu, Hao Li, Hailin Wang, Shuang Wu, Rui You, Yue Zhang, Teni, Zhihong Ye, Pengmei Yu, Yu Zhang, Yong Zuo, Hanlin Xu, Depeng Chen, Rui He, and Wenyi Zheng, Li Qiu and Wei Cao from Munich, for the time we spend together, ranging from sports, parties to travels.

At last but importantly, I would like to appreciate my family: my parents, my parents in law, my brother, my little nephew and niece and my two grandmas. Thank you for your kind support without any hesitation as always. Especially, I sincerely appreciate my wife Tingting Xiao for your accompany, support and encourage in the past four years. Without your accompany, I would not finish my PhD studying.

Moreover, I would thank this beautiful, romantic and passionate city Barcelona, and the small and quiet town Cerdanyola del valles, in which I have been living for four years. The culture, environment and customs deeply impress me, and forever.

In addition, I am grateful to the China Scholarship Council (CSC) for the financial support.

Weiqiang Wang

August, 2020

Introduction

1 Proteins

Proteins are biomacromolecules, consisting of one or more polypeptide chains in which 20 types of amino acids are covalently joined together by amide or peptide bonds.[1] They are the executors of a wide range of physiological functions within organisms, functioning in catalysis, DNA replication, cell signaling, cytoskeleton organization, immunity, or cargos transportation.[2] The three-dimensional structure and function of proteins are fundamentally determined by their primary sequences of amino acids (**Figure 1A**). Polypeptides form local folded structures thanks to interactions of the atoms in their backbones. The most common types of secondary structures are the α-helix and the β-pleated-sheet (**Figure 1B**). Both structures are held by hydrogen bonds, which form between amino and carbonyl groups of amino acids in sequential or spatial vicinity. The accommodation of these secondary structure elements in the space renders the tertiary structure of a protein, that is generally maintained by non-covalent interactions between amino acids side chains, like hydrogen bonds, ionic bridges, hydrophobic contacts, or van der Waals interactions (**Figure 1C**). Most proteins consist of a single polypeptide; however, some proteins are made up of multiple polypeptide chains, also known as subunits, and form complex quaternary structures bound by non-covalent interactions (**Figure 1D**). The subunits can be identical or different.[3] The quaternary structure facilitated the apparition of allosteric regulation, a fascinating way to control protein activity.

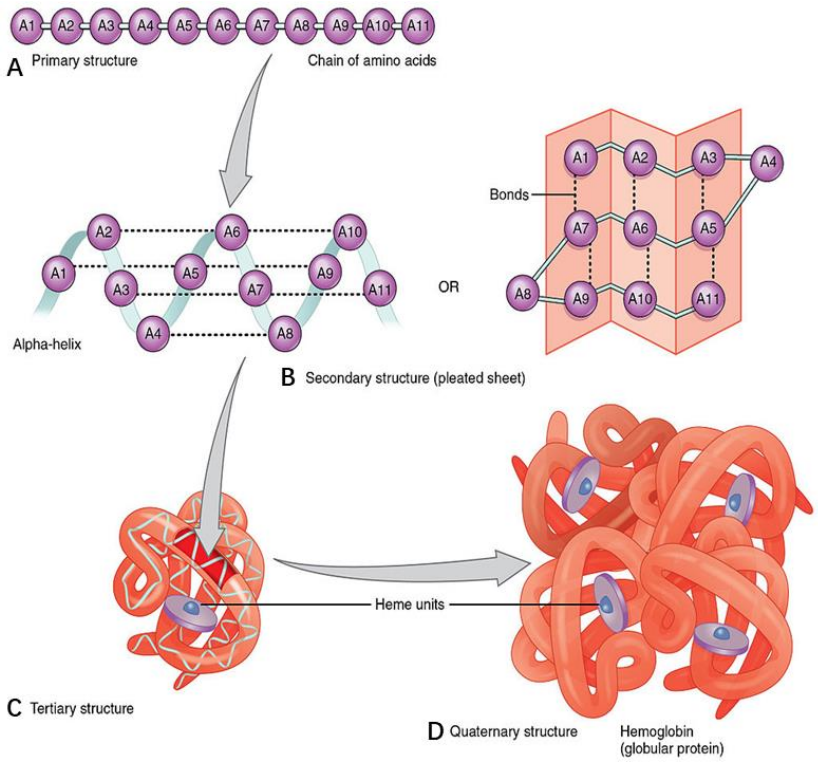

Figure 1. Structure of proteins. (A) The primary structure corresponds to the sequences of amino acids forming the polypeptides chains, **(B)** polypeptides fold into defined secondary structures: α-helix (left) and β-sheet (right), **(C)** the secondary structure further folds into the three-dimensional structure: tertiary structure, **(D)** quaternary structure occurs as a result of the interaction between two or more tertiary subunits. Adapted with permission.[3]

2 Protein folding

In the 1950s, it was first proposed by Francis Crick[4,5] that the "central dogma of molecular biology", also stated as "DNA makes RNA and RNA makes protein", explains the flow of genetic information in organisms. Specifically, the genetic information encoded within DNA goes into RNA with transcription. The biosynthesis of protein polypeptides occurs in the ribosome, where amino acids are connected in an order specified by mRNA, using tRNAs to carry amino acids and read the information of mRNA following the rules determined by "genetic code".[6,7,8] Finally, the polypeptide folds into the intended functional structure, spontaneously or under the assistance of molecular chaperones. Alternatively, unfolded or misfolded proteins might establish

non-native interactions and aggregate either inside or outside the cell, a phenomenon involved in a wide range of conformational diseases, including Alzheimer's, Parkinson's, and Huntington's diseases or Amyotrophic lateral sclerosis (ALS), among others.[9] The process of protein folding is still one of the less characterized aspects of the "central dogma of molecular biology" and deciphering how the information encoded in a string of amino acids renders a defined structure[10] following precise steps in the folding pathway[11] is still the subject of intense research.

2.1. Thermodynamic and kinetics of protein folding

In the past 50 years, many investigations have addressed how the folding process specifies the exact position that each atom occupies in the functional form of a protein, mostly through *in vitro* experiments and computational simulations. However, many questions remain still open.

In 1931, the theory of protein denaturation was first put forward by Hsien Wu,[12,13] which stated that protein denaturation is an unfolding process, instead of resulting from hydrolysis of the peptide bond[14] or dehydration of the protein.[15] He proposed that the compact and crystalline structure of natural proteins is formed by regular repeated patterns of folded chains stabilized by a large number of secondary valence linkages, which are easily destroyed by physical and chemical forces such as heating, high pressure, acids and alkalis, polar reagents and salts. In this regard, denaturation was seen as the process of disassociation of these labile connections. Then the regular arrangement in the rigid structure of the protein molecule converted into an irregular and diffuse configuration with a flexible and open-chain, in which the polar surface was altered, and the non-polar interior was exposed. Furthermore, it was proposed that the denaturation process might be reversible, referred to as the possibility of a denatured polypeptide chain to revert to the native state, a phenomenon first observed for myoglobin.[16] Later on, the reversible denaturation of serum albumin and other proteins was also confirmed, since the native and renatured state of these proteins shared similar physical properties, including their solubility, crystallization, and spectroscopic

features.[17,18] This posed denaturation as a thermodynamically-determined reversible process. Consequently, the refolding process of protein was considered spontaneous and directed by the composition and order of amino acid sequences, which account for establishing specific bonds in the space. In the 1970s, the groundbreaking ribonuclease experiment carried out by Anfinsen provided the definitive evidence on the nature of the determinants that control the folding of polypeptide chains into the three-dimensional structure and how this relationship emerges from the action of natural selection.[19,20,21] Full denaturation of ribonuclease in the presence of urea and a reductant, led to the breaking up of non-covalent secondary contacts, but also four covalent pairings of disulfide bonds, which resulted in the conversion of the stable native structure into a disordered open chain, with the consequent loss of catalytic activity. The denatured polypeptide chains spontaneously refolded into the native and catalytically active conformation after the sequential removal of urea and the reducing agent. Remarkably, the stochastic possibility of forming the four native disulfide bonds was only one of 105 possible pairings, which argued that cooperative effectors should assist the formation of the native structure, something that was confirmed after obtaining fully active proteins from chemically synthesized insulin[22,23] and ribonuclease.[24] Thus it appeared that a network of weak intermolecular secondary interactions was sufficient to collapse the polypeptide chain in a stable form that allows the formation of the native disulfide bonds. Based on the ribonuclease experiments, together with other supporting studies, Anfinsen proposed the "thermodynamic hypothesis" of protein folding and described it in his 1972 Nobel lecture as follows:[25]

This hypothesis states that the three-dimensional structure of a native protein in its normal physiological milieu (solvent, pH, ionic strength, presence of other components such as metal ions or prosthetic groups, temperature, and other) is the one in which the Gibbs free energy of the whole system is lowest; that is, that "the native conformation is determined by the totality of interatomic interactions and hence by the amino acid sequence, in a given environment". In summary, the primary amino acid sequences determine the native protein conformations. Since then, many physicochemical

methods were developed to study the role of thermodynamic parameters (enthalpy and entropy) and other factors like solvation and hydrophobicity exposure in the denaturation and refolding reactions of proteins displaying different three-dimensional structures, allowing to demonstrate the generality of the so-called "Anfinsen's principle".

In the 1980s, the site-directed mutagenesis technique was implemented as a critical strategy to test individual residues' roles in the folding process.[26] Although many experimental pieces of evidence confirmed the reversible folding reaction of many small proteins according to the "thermodynamic hypothesis", the irreversibility of the folding of many large protein molecules was still challenging the concept of thermodynamic stability. At the same time, it was known that many catalytically active proteases were cleaved out of a larger polypeptide precursor. The two most extensively studied cases are subtilisin[27] and α-lytic protease.[28] Both of them contain an N-terminal pro region, which is necessary for folding and directly related to the rate of on-pathway folding reactions, but not to stabilize the folded form of enzymes.[29] In fact, the unfolded α-lytic protease in the absence of pro region refolded into an intermediate state with a partially organized tertiary structure when the denaturant was removed. On the contrary, the addition of the pro region could prompt the rapid conversion of the intermediate state to the native state. These findings were also reported for subtilisin.[30] This evidence suggests that there might exist competing pathways leading to kinetically accessible intermediate states with multiple energy minima, from which the native state is the lowest energy conformation. Therefore, if the energy barriers between the intermediate and native states are high enough, the folding will strongly depend on the starting point of the reaction. Accordingly, in the proteases mentioned above, the pro region played a crucial role in reducing the free energy barrier between the intermediate and native states, via intermolecular interaction. Another example of the existence of multiple energy minima during folding is the serpin family of protease inhibitors. The denatured plasminogen activator inhibitor 1 (PAI-1) initially folded into an active state, but then slowly converted to an inactive form over several hours during the refolding reaction.[31]

It was clear that the inhibitory form of PAI-1 was stable but not the lowest energy state, with low energy barriers of conversion between active and inactive states. Similar behavior was also found for the influenza virus hemagglutinin (HA), which initially folded into a metastable native conformation (HA-N) at neutral pH, but further converted to a shifted conformation with lower free energy at lower pH (HA-L), triggering the fusion of the virus and the membrane during endocytosis and promoting the entry of the virus into cells.[32] All these results appeared to indicate that the folding of proteins might involve multiple pathways with different accessible intermediate states and several downhills routes towards the native state, depending on the starting conditions. The native state might not be the global energy minima, and only a tiny fraction of the vast conformational space is accessible for a given polypeptide chain during folding. Therefore, the folding reactions of many proteins appear to be controlled kinetically rather than thermodynamically, as proposed by David Baker and David A. Agard in 1990s.[33]

2.2. The "new view" of protein folding

New biophysical techniques, including circular dichroism and NMR spectroscopy combined with stopped or quenched flow methodologies, were incorporated in the analysis of the complex process of protein folding over time, capturing real-time information of metastable intermediate states.[34] Simultaneously, the application of computational simulation, as a theoretical approach, played an essential role in understanding the pathways and fundamental mechanisms of protein folding.[35] Additionally, the concept of free energy surfaces or landscapes allows the visualization and description of the process of protein folding rationally. Martin Karplus and co-workers exploited this knowledge to propose a lattice model to depict both simple and complex folding reactions by using the Monte Carlo method, linking for the first time theory and experimentation (**Figure 2**).[36] The resulting models, were consistent with the simplified one-dimensional free energy surface introduced by David Baker in 1994 to describe the behavior of protein folding under thermodynamic or kinetic control.[33]

In particular, a string of beads arranged on a lattice represented a polypeptide chain, in which one bead at one position interacted with each other by pairwise contact energies as effective energies representing residue-residue interactions. The effective energies guide the beads attracting each other to a final lowest energy conformation similar to the highly organized structure of a native protein. As shown in **Figure 2A**, the free-energy surface for the folding of a 27-mer with a 3×3×3 cubic lattice, representing a small protein, was calculated as a function of two reaction progress variables: total number of non-covalent contacts between residues (C) and the number of native contacts (Q_0). Firstly, the unfolded chains rapidly folded into a broad minimum on the free energy surface, where the species had ~60% of the total number of contacts, of which only ~25% were native-like. So, the possibly accessible conformations were reduced from ~10^{16} to ~10^{10}. Then the partially folded conformations further arrived at a rate-limiting stage, which involved a random search of ~10^3 possible conformations with 80%~90% of native contacts that leads rapidly to the native state. Therefore, the lattice model demonstrated that, effectively, only a small fraction of the total possible conformations were accessible to polypeptide chain during folding, which indeed appeared to be the solution to "Levinthal Paradox".[37] For many small protein molecules, the free energy surface of folding could be described as a sharp "funnel-like" space, which appeared to be a fast and effectively two-state reaction. Instead of a series of mandatory steps, the folding of a polypeptide chain involves a stochastic search among many accessible conformations. In such a stochastic search, the more stable native-like contacts from different regions within the partially folded state force the folding undergoing to the lowest energy structure, in contrast to the less stable non-native interactions.

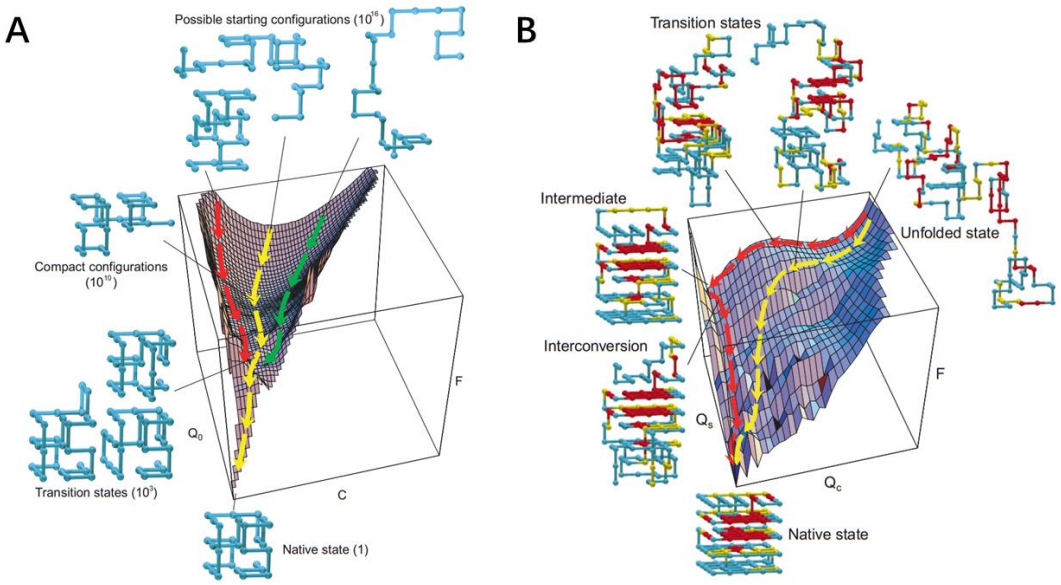

Figure 2. Schematic energy landscape of protein folding. **(A)** Free-energy (F) surface of a 27-mer as a function of the number of native contacts (Q_0) and the total number of (native and non-native) contacts (C) obtained by sampling the accessible configuration space with Monte Carlo simulation. The calculated free energy surface served as a "funnel" for a stochastic search of the most stable conformation from many accessible conformations to a final unique structure with different tracks. **(B)** Calculated free-energy surface for the folding of a 125-mer lattice model, plotted as a function of the number of 'core' contacts (denoted Qc; residues are shown in red), the number of 'surface' contacts (denoted Qs; residues are shown in yellow), and the residues not involved in these variables are blue. The yellow trajectory represents an observed 'fast track' in which partial formation of the core leads directly to the native state, and the red trajectory represents a 'slow track' in which non-core contacts are formed before the core is complete leading to the chain to a misfolded intermediate that corresponds to a local free-energy minimum. Adapted with permission.[36]

In contrast to small proteins, large proteins would give rise to higher barriers to reorganizing partially folded conformations and introducing more complexities in the folding landscape. In a lattice polymer, a 125-mer with a 5×5×5 cubic structure was used to simulate the folding behavior of a large protein (**Figure 2B**). The free-energy surface was plotted as a function of two progress variables, corresponding to the number of "core" contacts involved in fast folding (Q_c, residues shown in red) and the

number of other native contacts (Q_s, shown in yellow), respectively. As the 27-mer, the 125-mers rapidly started folding into a semi-compact transition state with ~60% of the core contacts, which present a shallow entropic barrier. Then several pathways concurred to pass through this state. Typically, approximately 15% of molecules found additional core contacts and directly folded into the native state with a "fast track" (yellow path). Meanwhile, another ~40% of molecules were trapped into stable intermediates with two independent domains, resulting from a large number of non-native contacts out of the core (red path). It was required to overcome an energetic barrier to escape from this state and fold into a native state. The remaining 45% of molecules also folded into a partially misfolded state but escaped from the traps more rapidly. The simulation results were in agreement with the experimental observations of the folding process of hen lysozyme, which had multiple kinetic pathways involving the formation of different intermediates with different folding rates. The diversity of folding pathways primarily depends on the particular sequences involved in the secondary-structure formation and the spatial distribution of low-energy contacts in the native state.

Therefore, the folding process can be recapitulated as two stages: in the early stage, the preference for secondary structure, together with the free-energy preference for burying hydrophobic residues and exposing the hydrophilic residues force the polypeptides chain to fold into a native-like semi-compact structure but lacks unique tertiary interactions; then the highly cooperative interactions between side chains of all residues prompts the conversion of the intermediate into the native state. The energy differences between the native state and intermediates ensure that, under a defined condition, only in one pathway, the energy barriers are low enough to be productive for a given sequence. The folding rate depends on internal native contacts and, thus, on the stability or free energy. The "new view" of protein folding proposed by Karplus and co-workers[38] led to a unified mechanism of folding and the determinants of folding rates, which could describe the generic folding behavior of polypeptide molecules.

Overall, the folded native protein is in an equilibrium between attractive forces,

including electrostatics, hydrogen-bonding, van der Waals interactions, intrinsic propensities, and hydrophobic interactions, and opposing forces such as conformational entropy and electrostatic repulsion.[39] These first approximations to the folding code allowed to understand the role of natural selection and the impact of mutations in protein folding. The research on folding mechanisms also pushed the development of predictive algorithms, for the resolution of the three-dimensional structure of specific proteins, and even for the design of novel proteins with versatile functionalities based on genome sequencing.[36] Different computational algorithms exploit the quantitative theory of minimal frustration to predict the structure of proteins, with good overlap with the native structures determined empirically, and to design sequences that can efficiently fold to a unique structure.[40] Several pathologies, including Alzheimer's and Parkinson's diseases, are related to a failure of folding and the aggregation of partially folded intermediates.[41] The fundamental study of the folding process at a molecular level might provide new ways to treat these diseases (i.e., blocking the misfolding pathways).

3 Protein misfolding and aggregation

3.1. Protein folding and misfolding in cells

As Levinthal proposed,[37] the mechanism of protein folding can be described as a cumulative search for conformations close to the native state as the number of native-like interactions increases. Those regions with native-like interactions are retained in each search and consequently restrict the conformational space, resulting in the existence of a series of sequential partially folded intermediates in different pathways, that rapidly reach the overall three-dimensional native structure within microseconds.[42] However, despite its paramount relevance, our fundamental knowledge of protein folding mechanisms was only based on *in vitro* studies. It is clear that the folding of a nascent polypeptide released from the ribosome is much more sophisticated, especially when considering the complexity of intracellular circumstances, with a high local and

total concentration of macromolecules.[43] Many proteins in the cytoplasm begin to fold as a nascent chain and experiment a significant part of their folding events after release from the ribosome, while the secreted proteins fold in the endoplasmic reticulum (ER) followed by translocation to the membrane.[44] Therefore, the folding behavior of a nascent polypeptide chain *in vivo* depends on the environment in which the reaction takes place, but comply with the fundamental mechanisms of folding discussed above.[45] The free-energy surface of folding accessible to a given polypeptide chain is often rugged with kinetic energy barriers that must be surpassed (**Figure 3**).[46] *In vivo*, partially folded states are often kinetically trapped and accumulated as intermediate states. There exist two predominant intermediates during folding (**Figure 3**): a "molten globule" state, initially lacking native contacts but that rapidly collapsing into a native-like globular structure once they cooperatively accumulate; and a transient intermediate state stabilized and trapped by non-native interactions in a stable non-native topological structure, which is clearly distinct from the native state.[47] The latter species are usually considered as misfolding states. The initiation of misfolding reaction is a stochastic event, but its frequency seems to be connected with the aging of individuals.[48] In the misfolded state, hydrophobic residues and unstructured polypeptide regions, which are initially buried in the native state, are exposed and prone to non-specific contacts with other molecules in the cell. Consequently, aggregation might occur, driven by exposed hydrophobic regions in a concentration-dependent manner, leading to the formation of amorphous aggregates or amyloid fibrillar structures (**Figure 3**), which, in turn, are related to many neurodegenerative diseases.[49] In this scenario, a series of protection strategies to prevent misfolding have been evolved in living systems, such as molecular chaperones, folding catalysts, and a stringent protein "quality control" mechanism.[50] Besides, the escaped unfolded and misfolded proteins can be recognized and degraded by the ubiquitin-proteasome pathway.[51]

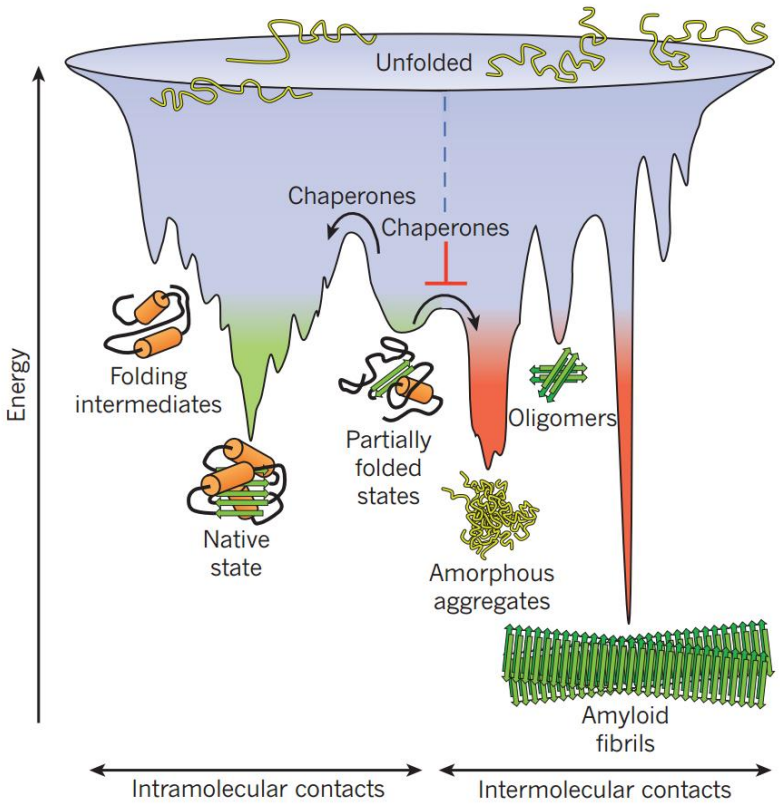

Figure 3. Competing reactions of protein folding and aggregation. Scheme of the funnel-shaped free-energy surface explored by proteins as they move towards the native state (green) by forming intramolecular contacts. Chaperones reduce the accumulation of kinetically trapped misfolded conformations that need to traverse free-energy barriers to reach a favorable downhill path. When several molecules fold simultaneously in the same compartment, the free-energy surface of folding may overlap with that of intermolecular contacts, resulting in the formation of amorphous aggregates, toxic oligomers, or ordered amyloid fibrils (red), which can be prevented by chaperones. Adapted with permission.[46]

Molecular chaperones are defined as a series of proteins that exist in all living systems, from bacteria to humans, which non-covalently bind nascent polypeptide chains or misfolded proteins to prevent their misfolding in different ways.[44] They are also known as stress proteins or heat shock proteins (HSPs) since they are significantly increased in cells under stress conditions. Molecular chaperones are classified according to their molecular weight, such as HSP40, HSP60, HSP70, and so on, which combine into cooperative pathways and integrated networks for proteostasis. Some

molecular chaperones non-specifically interact with exposed hydrophobic residues of nascent chains released from the ribosome, avoiding the intermolecular aggregation driven by those aggregation-prone regions. Others are involved in guiding the later stage of folding of misfolded or partially folded polypeptide chains and assist them in crossing over energy barriers towards the native state (**Figure 3**). It also has been demonstrated that some molecular chaperones rescue misfolded protein from aggregates and enable them to refold correctly.[52] The remainder of molecular chaperones are involved in protein trafficking and assist proteolytic degradation. Furthermore, there are several classes of folding catalysts that significantly accelerate the folding step, contributing to reduce the possibility of misfolding. The most well-known examples are peptidyl-prolyl isomerases and disulfide isomerases, corresponding to the *cis/trans* isomerization of proline, and the formation of disulfide bonds, two reactions that sharply limit the rate of folding.[53] Moreover, a complex series of glycosylation and deglycosylation processes prevent misfolded proteins from being secreted from the cell, constituting the quality control mechanism in the endoplasmic reticulum.[52]

3.2. Protein aggregation and amyloid formation

The presence of molecular chaperones is of paramount importance in preventing protein misfolding, as the natural functionality of a protein strongly relies on its native three-dimensional structure. For this reason, unsurprisingly, the failure of a protein to fold, or misfolding, might lead to a deficiency of the protein function and, consequently, might induce disorders. For instance, many neurodegenerative diseases, including Alzheimer's and Parkinson's diseases and the spongiform encephalopathies, are directly associated with the presence of protein aggregates in brain tissues. The initiation of aggregation is driven by solvent-exposed hydrophobic regions within a misfolded protein and is strongly dependent on the protein concentration, which can be altered by genetic dosage,[54] and polymorphisms in promotor sites of the involved genes.[55] In other cases, mutations and covalent modifications such as oxidative

modifications,[56] phosphorylation,[57] and SUMOylation[58] might significantly change the aggregation propensity of proteins related to diseases.[59]

Protein aggregation is a complex process, involving a series of intermediate species that lead to amorphous aggregates or fibrillar structures called amyloids. The formation of such aggregates is dependent on the presence of partially folded globular conformations and the exposure of hydrophobic regions or results from the fragmentation of proteins into unstructured segments through proteolysis.[60] Typically, the process of amyloid formation has been described as a nucleation-dependent polymerization reaction, generally undergoing a lag phase followed by a rapid growth, and a final equilibrium stage.[61] Particularly, in the first phase of the aggregation, a soluble oligomeric state is formed, which is thought to be toxic and responsible for many misfolding diseases. The specific toxicity mechanism of these poorly ordered oligomers is still unknown, but it has been suggested to display a strong correlation with their exposed hydrophobic regions.[62] During this phase, the partially folded proteins may undergo a rapid assembly into bead-like structures, as imaged by electron microscopy. The specific structural features of these intermediate species were first revealed by the evidence that monoclonal antibodies generated to interact with oligomeric species did not bind monomeric or fibrillar forms of polyglutamine, Aβ, α-synuclein and prion protein.[63] These soluble oligomers may further assemble into short, and thin fibrillar species, known as “protofibrils”. Finally, these prefibrillar species undergo a rearrangement to form the mature fibrils.[45]

Recently, cellular and structural biology studies have advanced our understanding on the physiological processes that experiment a nascent polypeptide chain, from its synthesis to its degradation in cells,[64] including now aggregative pathways, as represented in the scheme of **Figure 4**. The variety of conformational states accessible for a given protein under specific conditions depend on the thermodynamic stabilities of the different states and/or the kinetics of interconversion pathways.

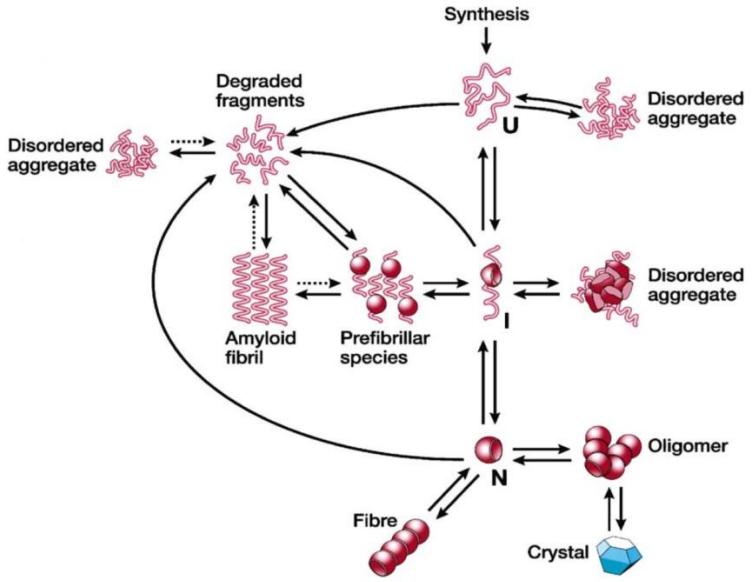

Figure 4. Schematic representation of accessible states for a nascent polypeptide chain released from a ribosome. The relative populations of the different states depend on the kinetics and thermodynamics of the various equilibria shown in the diagram. U, Unfolded; I, Intermediate; N, Native. Adapted with permission.[45]

4 Amyloid fibrils

4.1. Characteristics of amyloid fibrils

The presence of amyloid aggregates, derived from misfolded or unstructured proteins, is a hallmark of many neurodegenerative diseases (**Figure 5**).[65] Each amyloid disease usually involves the aggregation of a single specific protein, although different counterparts, including other proteins and carbohydrates, might also be incorporated into amyloid aggregates in the complex and crowded intracellular environment. The formation of amyloids has been characterized *in vitro* for a significant number of globular or unstructured proteins linked to defined amyloid diseases. Although the characteristics of the soluble forms of these proteins are varied, their amyloid forms share many common features.[66]

Amyloid fibrils are long, unbranched, sometimes twisted, filaments with a width of

several nanometers, and a length of up to 10 μm (**Figure 5**). A characteristic feature of such filaments, observed initially by X-ray diffraction, is the presence of cross-β core structure, which consists of in-register and parallel β-sheets running perpendicular to long fibrils axis (**Figure 5**).[67] An intermolecular hydrogen bond network involving the main chain stabilizes the core amyloid structure, and it is arranged perpendicular to the axis of fibrils. In contrast, the individual β-strands are orthogonal to this axis. This explains the amyloid fibril structural similarity for polypeptides with different amino acid sequences since the main chain is common to all of them. This was further consistent with the experimental evidence that a conformation-specific antibody generated to react with amyloid fibrils of Aβ peptide did neither bind its soluble monomeric form nor its amorphous globular protein aggregates, recognized the amyloid aggregates derived from other proteins, like polyglutamine.[68] The in-register cross-β structure enables the monomeric form of a protein to self-assemble into reproducible amyloid fibrils *in vitro* without the assistance of any other components. Amyloid fibrils show tinctorial properties in the presence of specific dye molecules such as Congo red and thioflavin-T (Th-T), which have been used as amyloid-specific dyes in diagnosis for many years. The mechanism of such binding has not been yet well understood, but it is thought to be correlated with the incorporation and insertion of these dye molecules into the grooves of intermolecular β-sheet structure.[69]

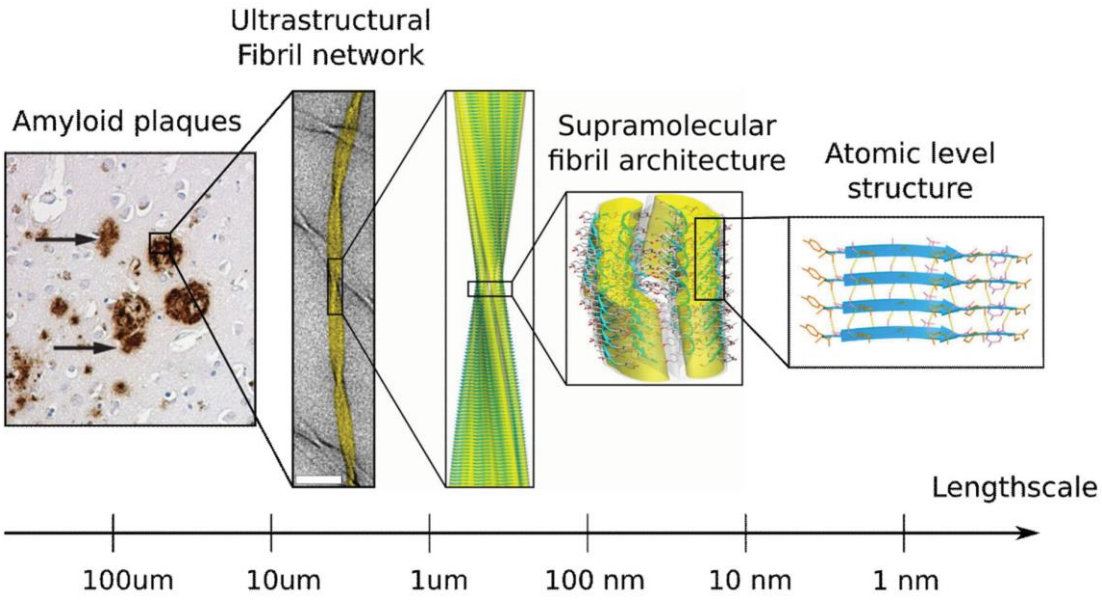

Figure 5. The length scales at different levels of amyloid architecture, ranging from the atomic level to amyloid plaques. Adapted with permission.[70]

For a long time, it was generally suggested that the ability to form amyloids was limited to those proteins that possess specific sequences motifs compatible with the amyloid core structure. However, recent studies on amyloid aggregates formed from many non-disease-related proteins support that the ability to form amyloid fibrils is a common, intrinsic, or generic feature of all polypeptide chains.[71,72] However, the aggregation propensity of different sequences can be dramatically different.[73] Some residues are less soluble than others in such a way that aggregation-prone residues concentration is required to initiate the aggregation, and their existence and potency depend on the polypeptide sequence. In this context, it became evident that the substitution of residues, even of single amino acids, might substantially change the nucleation speed or the rate of aggregation in a given polypeptide chain. Consequently, many *in silico* algorithms, including AGGRESCAN,[74] FoldAmyloid,[75] PASTA 2.0,[76], and Zyggregator[77], were developed to predict the aggregation propensities of polypeptide chains and to evaluate the effect of a site-specific mutation on their aggregation features. The accuracy of these algorithms allowed designing *de novo* peptides with desired aggregation properties.

4.2. Functional amyloid and amyloid-based nanomaterials

Amyloid proteins can self-assemble into ordered fibrillar structures, which were initially thought to be exclusively associated to human amyloidosis as well as to neurodegenerative diseases.[2,78,79] However, the discovery of functional amyloids in different organisms, from bacteria to humans, assisting different physiological functions in living cells,[80] has changed the amyloid/pathogenesis paradigm. For instance, functional amyloids are involved in curli-mediated biofilm formation in *E. coli*,[81] memory persistence in *Drosophila*,[82] hypersensitive response activation in plants,[83] melanin polymerization in mammalian cells,[84] and hormone storage in humans,[85] among others. In addition, the remarkably unique "cross-β" structure,[86] stabilized by numerous hydrogen bonds,[67] together with π-π and hydrophobic

interactions,[87] endows amyloid fibrils with high stability and high resistance to extreme physicochemical conditions, such as treatment with proteinases, chaotropic agents, and high temperatures.[88] These properties make amyloid fibrils ideal building blocks for designing novel protein/peptide self-assembled nanostructures. Indeed, amyloid fibrils have been incorporated as structural components in functional materials with different biomedical and biotechnological applications, including nanodevices,[89] biomembranes,[90] hydrogels,[91] biosensors,[92] and energy conversion materials.[93] However, the inherent conformational transition towards a β-sheet-rich structure during fibrillation necessarily implies the loss of the globular native fold and the subsequent inactivation of the protein. Therefore, designing fibrillar structures containing correctly folded and active proteins is still a challenging task. A way to bypass this limitation is by designing a hybrid structure in which the globular proteins are chemically attached to preformed amyloid fibrils.[94] However, this approach is restrained by a reduced protein conjugating chemistry, undesired cross-reactivity between the polypeptides, and the unavoidable inactivation of a fraction of the protein during conjugation.

5 Prions

Prions are proteins able to switch between two or more conformations, of which at least one corresponds to an amyloid fold.[95,96,97] So far, only a prion protein (PrP) has been identified in mammals; it is well-conserved across species and linked to different neurological pathologies.[98] The concept of infectious amyloids was initially proposed to lie behind different neurological diseases, including Creutzfeldt-Jacob disease in humans, sheep scrapie and bovine spongiform encephalopathy, all caused by the amyloid state of the prion protein (PrP).[99,100] Thus, prions were assumed to be pathogenic agents without any beneficial functional implication.[101] Roughly ten prion proteins have been characterized in the yeast *Saccharomyces cerevisiae*. They are usually referred to as functional prions since, in contrast to "toxic" mammalian prions, they do not seem to have a significant impact on viability, but instead, they play physiological roles, especially in the adaptation to changing extracellular

environments.[102,103] They usually consist of an intrinsically unstructured region of low complexity enriched in asparagine (N) and glutamine (Q) residues known as the prion domain (PrD) accompanied by one or more globular domains[104,105] The PrD is both necessary and sufficient for self-assembly, while the globular domain accounts for the protein functionality.[106] Yeast prions are thus modular and, in principle, the original globular domain can be replaced with any desired functional protein without impacting the fibrillation propensity significantly. Domains with properties resembling those of PrDs have been identified in the proteins of different organisms, including humans, and are named prion-like domains (PrLDs).

5.1. Functional prions in yeast

Initially, the phenotypes [PSI+] and [URE3] of yeast were identified as non-chromosomal genetic elements. Later on, Aigle and Lacroute reported that *ure2* nuclear gene mutant strains were unable to propagate the [URE3] element but manifested the same phenotype as [URE3] cells, namely de-repression of nitrogen catabolism genes;[107] however, did not realize that [URE3] could be a prion of Ure2p. It was Reed B. Wickner, who, based on these pieces of evidence and his own experiments, recognized that the relationship between *ure2* and [URE3] was not that expected for a chromosomal gene necessary for the propagation of a nucleic acid replicon, but instead that [URE3] was an inactive form of Ure2p able to inactivate the normal protein form, acting thus as a prion.[108] Then his group inferred that this could also be true for the [PSI+] phenotype,[108] and concluded that previous studies[109,110,111] supported [PSI+] being a prion of Sup35p. Moreover, they established a series of genetic criteria that identify yeast prions:[112] i) a reversible curability, ii) the phenotype appearance is induced by overexpression of the prion protein, iii) the phenotype resembles that of a prion protein gene recessive mutation.

These genetic criteria were useful for the discovery of a range of new prions including [PIN+],[113] [SWI+],[114] [MOT+][115] and [MOD+].[102] The formation of amyloid *in vitro* by Sup35p,[116] and Ure2p,[117] the protease resistance of Ure2p in [URE3] strains,[118] together

with the seeding properties of cellular extracts in [PSI+],[119] all converged to indicate that the prion form of these proteins was an amyloid. This assumption was further confirmed after observing that *in vitro* formed Sup35p amyloids transmit the [PSI+] trait.[120]

The inheritance of phenotypic traits due to prion formation involves a structural conversion, which in a majority of cases is driven by a Q/N rich PrD.[121] PrDs proved to be sufficient to induce and propagate the prion state, and they become integrated into the core of the amyloid structure, which is thought to be arranged as an in-register parallel β-sheet architecture,[86] according to nuclear magnetic resonance (NMR) and electron microscopy studies of the amyloids of Sup35p and Ure2p PrDs (**Figure 6**).[122,123] One or more globular domains appended to the PrDs can retain their native structure in the amyloid filaments (**Figure 6**), although the activity of these domains was found to be somehow reduced, likely because of steric impediments.[124] Such folded, in-register parallel β-sheet architecture perfectly explains the mechanism of protein templating in prion conformational conversion (**Figure 6**).[125] In particular, the favorable interactions between hydrophobic or hydrophilic side chains along the long axis of the filament maintains the in-register β-sheet structure. The same interactions force the new monomer to join the end of the filament to adopt the same conformation as the molecules already in the fibril, just as nucleic acids template their sequences.[101] An exception to this rule is the HET-s prion, from the filamentous fungi *Podospora anserine*, involved in heterokaryon incompatibility, protecting cells from fungal viruses.[126] The amyloid formed by the PrD of HET-s displays a β-helix structure, in which each monomer contributes two turns to the β-helix.[127] Recently, cryo-EM and computational studies suggest that PrP^{Sc} would adopt a similar β-solenoid structure.[128,129] In fungal prions, the self-propagation of the prion phenotype during cell division is generally accomplished by the action of the chaperone system, which fragments existing fibrils, generating seeds that initiate the polymerization reaction in daughter cells.[130,131]

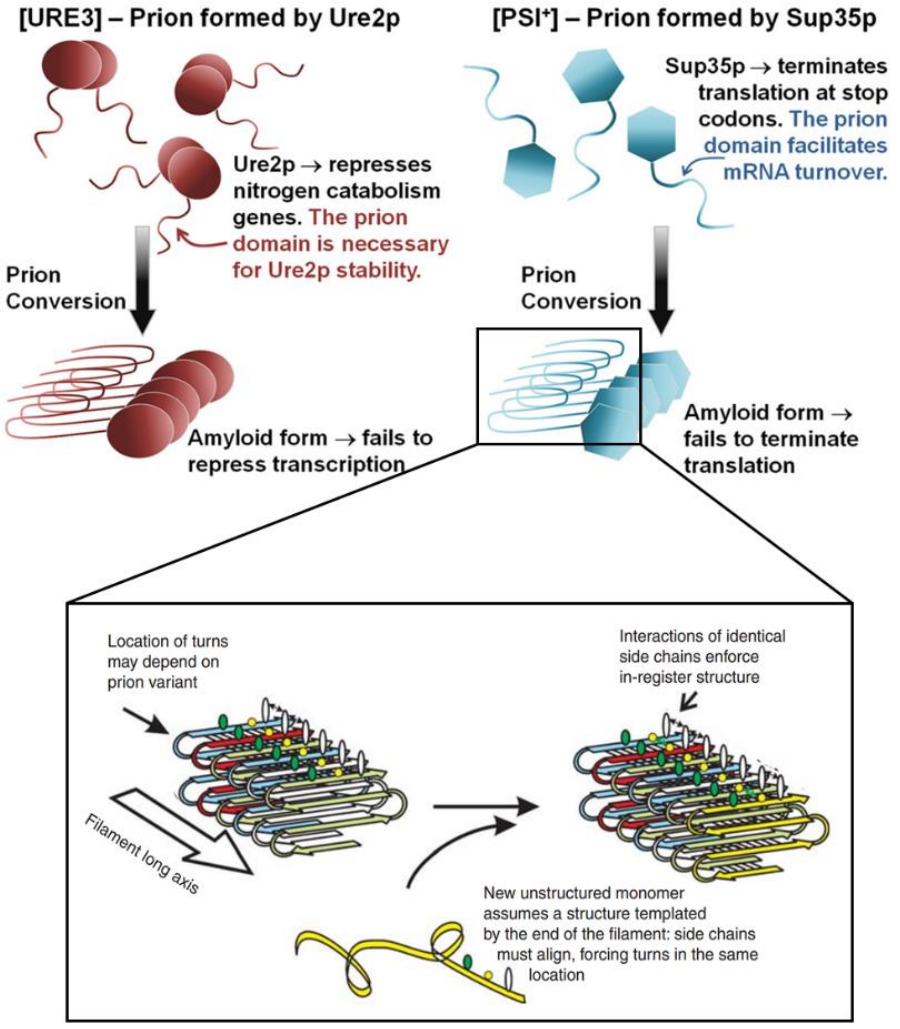

Figure 6. Typical prions [URE3] and [PSI+] of S. cerevisiae. These prions rely on self-propagating amyloids of Ure2p and Sup35p, respectively. The insert is the proposed mechanism of conformational templating by the prion domain. Energetically favorable interactions between identical side chains enforce the in-register architecture of these amyloids. A new monomer being added to the end of the filament must assume the same conformation as that of molecules previously incorporated into the filament. Adapted with permission[112]

The first and most studied of yeast prion proteins are Sup35p and Ure2p (**Figure 6**), involved in translation termination and nitrogen catabolism regulation, respectively.[108] Sup35p, a subunit of the translation termination factor, has a clear three-domain structure. The N-terminal residues 1-123 (Sup35N) are sufficient to induce and

propagate the [PSI+] prion phenotype and thus constitute the PrD. The C-terminal residues 254-685 (Sup35C), fold into a globular domain, responsible for the translation termination function.[132] The charged middle domain (Sup35M), corresponding to residues 124 to 253, is important but not required for [PSI+] appearance and maintenance.[133,134] Sup35N drives self-assembly of Sup35p into an amyloid, which results in a substantial reduction of its translation termination function and leads to an increase of nonsense codon readthrough. The PrD of Sup35p has itself a non-prion function, facilitating mRNA turnover. Ure2p has an architecture similar to that of Sup35p, including an N-terminal PrD (residues 1-65) responsible for [URE3] formation[135] and a C-terminal domain (Ure2C) structurally similar to glutathione transferase but involved in repression of nitrogen catabolism genes by binding to the transcription factor Gln3p, an interaction that is significantly impeded in the amyloid state.[136]

The modular structure of yeast prions allows to dissect them in order to assess the prionogenic potential of other proteins and, in this way, identify new prions. This is performed by replacing the PrD or globular domain, via genetic engineering, with any desired polypeptide or reporter protein, such as the green fluorescent protein (GFP), to study the ability of the fusion protein to induce [PSI+]-like appearance or to visualize protein inclusions formation in cells.[137] These studies demonstrated that yeast prions could be easily engineered, immediately suggesting that they can be exploited in the design of self-assembling functionalized materials.

5.2. Prion-like nanomaterials obtained via genetic protein fusion

It has been reported that the yeast prion Ure2p exhibits glutathione peroxidase activity in both the native and fibrillar forms.[138] In addition, exogenous globular domains fused to the Ure2p PrD retain a native-like fold and show catalytic activity within *in vitro* formed amyloid filaments.[124] This ability to combine amyloid and globular folds in the aggregated state seems to be generic of yeast prions since the analysis of Sup35p amyloid filaments by cryo-EM, STEM[139] and solid-state NMR,[140] revealed that

filaments of full-length Sup35p indicated a thin fibril backbone surrounded by globular C-domains. The formation of a spatially-defined fibrillar backbone and the segregation of folded globular domains are central for the design of prion-like self-assembling functionalized nanomaterials. In contrast to the fast and challenging to control aggregation reactions of pathogenic amyloids, prion and prion-like proteins generally display slow and tunable aggregation kinetics, which is considered optimal for the design of self-assembling materials.[70] Actually, different enzymes such as alkaline phosphatase (AP),[141] carbonic anhydrase (CA),[124] methyl-parathion hydrolase (MPH)[142] or other globular proteins like protein G and Z domain[143] have been genetically fused to the PrDs of yeast prion Sup35 or Ure2p to generate functionalized nanomaterials in which the PrDs constitute the spine of the fibrils, whereas the globular domains hang from them in a folded and functional conformation. These materials are biocompatible, degradable, and environmentally friendly and can be employed in applications encompassing from catalytic nanowires[141] to nanosensors.[142] PrDs and PrLDs in functional prions in yeast with versatile features are being exploited to generate genetically encoded functionalized nanomaterials displaying their potential applications. The general strategy to obtain the building blocks for such materials consists in the genetic fusion of a yeast prion PrD to a protein of interest and its recombinant production in microbial cell factories (**Figure 7**). The purified soluble fusion protein can self-assemble spontaneously into nanofibrils, which exhibit the function of the appended globular domain, under appropriate conditions of temperature, pH and agitation, in certain cases, requiring the presence of detergents.

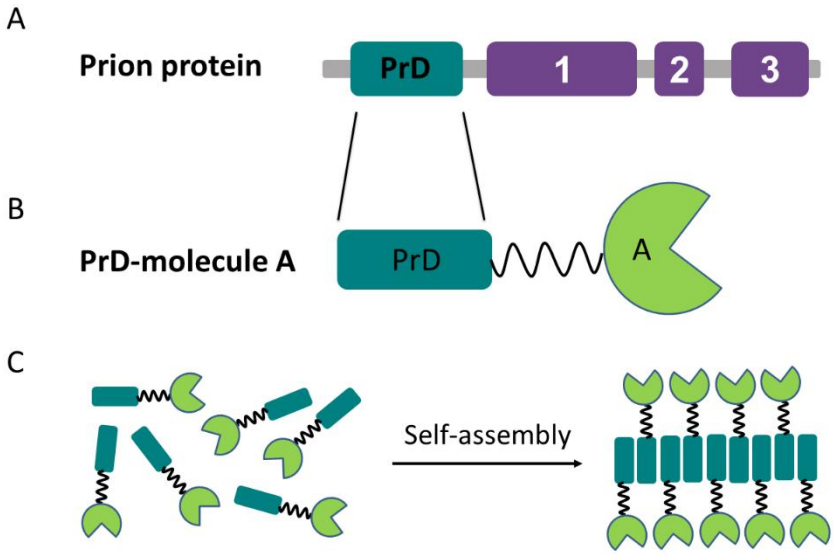

Figure 7. The general genetic fusion approach for functionalized nanomaterials. (A) Domain organization of the prion protein: the prion domain (PrD) (green) and the respective functional domains (purple) in a prion protein are shown. **(**B**)** Cartoon of the prion domain fused to molecule A, which represents a globular functional protein. **(**C) Soluble fusion protein self-assembly into nanofibrils, which preserve the function of molecule A.

Biosensors. The first prion-like nanomaterial obtained via genetic protein fusion was intended to work as a biosensor. Dong Men et al[142] generated bifunctional protein nanowires (bFPNw) that displayed two different globular proteins: protein G and methyl-parathion hydrolase (MPH), which provide antibody binding and catalytic activities, respectively. They aimed to obtain fibrils with a high enzyme to protein G ratio. Towards this objective, both protein G and MPH were fused individually to the N-terminus of the Sup35p PrD (residues 1-61) and expressed as soluble monomers in *E.coli*. The fibrils formed by the self-assembly of the $Sup35^{1-61}$-protein G fusion were broken into small fragments to be used as seeds. The bifunctional nanowires exhibiting two different functions with different ratios were obtained by mixing these seeds and soluble $Sup35^{1-61}$-MPH in different proportions; this also allowed to tune the length of the fibrils. The biological activity of these bifunctional nanowires was tested by a typical ELISA assay, implementing them instead of the classical enzyme-conjugated secondary antibody. As a general trend, the longer the nanowires are, the higher the

signal amplification they provide. In this work, optimized nanowires of 500 nm in length (ratio of protein G to MPH was 1:8) showed a sensitivity for the detection of the F1 protein from *Yersinia pestis* 1000-fold higher than that of a protein G-MPH fusion and 100-fold higher than that of the traditional HRP-conjugated antibody-based ELISA.

In subsequent work, the same group designed auto-biotinylated bFPNw by introducing a biotin-avidin binding system, which can efficiently attach any avidin-labeled commercial enzyme, such as HRP or AP, to the nanowires backbone, surpassing the need to produce and purify a different fusion for each intended application.[144] In particular, the biotin acceptor peptide tag (BAP) and the IgG-binding domain, C1, of protein G from *Streptococcus* (SPG) were individually fused to the Sup35p PrD described above. The construct $Sup35^{1-61}$-BAP was co-expressed in *E. coli* with another construct encoding for the biotin holoenzyme synthetase (BirA). In this way, the BAP tag was biotinylated by BirA during the expression of the $Sup35^{1-61}$-BAP fusion. Instead of inducing the formation of bifunctional nanowires by seeding, which they found reduced the IgG-binding activity of protein G because of the sonication processes, on this occasion, they just mixed the pre-formed biotinylated $Sup35^{1-61}$-BAP fibrils with the $Sup35^{1-61}$-SPG monomer, which was templated and incorporated at the end of the fibrils. The function of the bFPNw was evaluated by performing a typical ELISA in the presence of Streptavidin-HRP instead of the typical secondary IgG-HRP conjugate. For the detection of *Y. pestis* F1 antigen, the auto-biotinylated bFPNw showed stronger signal intensity, lower background noise, and better signal stability than that of previous seeding-promoted bFPNw. More importantly, they showed 2000- to 4000-fold higher sensitivity than that of a conventional ELISA.

The lessons learned in these two previous works were exploited to construct a highly sensitive fluorescent molecular biosensor. First, an F64L/S65T/T203Y/L231H green fluorescent protein mutant (E^2GFP) was fused to MPH.[145] E^2GFP shows a distinct excitation and emission spectra depending on the pH, which makes it an effective ratiometric pH indicator. The chimeric E^2GFP-MPH protein allowed the detection of the pesticide methyl parathion (MP) by sensing the H^+ ions released during enzymatic

reaction. Interestingly enough, when this chimeric protein was further fused to the Sup35p PrD and incorporated into fibrils through its self-assembly, the obtained nanowires showed dramatic enhancement of sensitivity and allowed detection of MP at a concentration as low as 1pmol mL^{-1}, which was about 10000-fold that of the soluble fusion protein biosensor E^2GFP-MPH and 10^7-fold higher than the one of an equimolar mixture of free E^2GFP and MPH moieties. These nanostructures might become important tools for the detection of pesticides in the food industry.

Enzyme immobilization. Immobilization can enhance the value and usage of commercial enzymes, providing advantages such as an easy separation of the product and the reuse of the enzyme. The lab of Sarah Perrett has recently described PrD-based enzyme immobilization systems.[141] On this occasion, it was the PrD of Ure2p that was exploited to drive the self-assembly, whereas two different enzymes, AP and HRP, were individually fused to it. Both fusion proteins formed active nanofibrils, where the enzymes become immobilized. Although the fibrils exhibited slightly lower activity than that of the soluble wild type proteins, they could be recycled both in a conventional enzymatic setup and in the context of a continuous microreactor. In a follow-up study, the authors encapsulated the soluble PrD-AP protein fusion into a uniform droplet using microfluidic techniques.[146] The encapsulated soluble protein spontaneously self-assembled into a three-dimensional fibrillar network within the droplet, immobilizing AP molecules in their active form. As a result, a catalytically active microgel with a porous architecture was obtained. These hydrated functional nanoparticles, when combined with microfluidics, are ideal entities to implement microscale analytical assays.

Bioelectrodes. Vincent Forge and colleagues designed a protein-only redox biofilm inspired by the architecture of bacterial electroactive biofilms.[147] Briefly, the formation of the nanowires was induced by the self-assembly of the PrD of HET-s *P. anserine* prion, which was genetically fused to rubredoxin (Rd), a redox protein that acts as an electron carrier. The formation of the fibrillar structures allowed the alignment of Rd along the fibril axis in an array, which facilitates the transportation of electrons through

the network resulting in an exceptional conductivity. As a proof of concept, such redox-active fibrils were successfully applied to mediate the electron transfer towards the multi-cooper enzyme laccase for catalytic oxygen reduction. Therefore, these protein nanowires are expected to find applications as effective bioelectrodes.

Antibody purification. The increasing demand of antibodies for diagnostic and therapeutic purposes requests cost-efficient means for their purification. In recent work, Torleif Härd and co-workers exploited the small size and high affinity to IgG of the Z-domain from staphylococcal protein A to build up antibody-capturing fibrils. A Z-domain dimer (ZZ) was fused to the PrD of either Sup35p (Sup35-ZZ) or Ure2p (ZZ-Ure2). Both of them promoted the fibrillation of the chimeric protein.[143] These fibrils bind to IgG, but their loading capacity was limited by steric hindrance. The PrDs of Sup35p and Ure2p were then co-fibrillated with chimeric Sup35-ZZ and ZZ-Ure2 proteins, respectively, to maximize the accessibility of antibodies to ZZ. This fibril doping strategy resulted in co-fibrils in which ZZ globular domains are more spaced in the fibrillar structure, thus maximizing the fibrils binding capacity (**Figure 8**). Different doping frequencies were assayed, and in the optimal PrD:protein fusion ratio, the resulting fibrils yielded a capture capacity of 1.8 mg of antibody per mg fibrils, which is 20-fold higher than the one of commercial protein A agarose.

Despite the exceptional binding capacity of ZZ functionalized fibrils, this material should be produced and engineered at large-scale before it can commercially compete with the present well-positioned methods for purification of antibodies. Towards this objective, the Härd lab integrated both the Sup35-PrD and the chimeric Sup35-ZZ genes in the genome of *Komagataella pastoris*.[148] This yeast continuously expressed and secreted both proteins into the extracellular medium, where they co-fibrillated into ready-to-use functionalized nanofibrils, with an impressive production yield of 35 mg/L culture. Importantly, the separation of the fibrillar material from the cell culture required only centrifugation and their resuspension in a saline buffer. The high yield, high homogeneity, and high stability of the purified material, together with minimal equipment requirement and the low hands-on time of this strategy, allows foreseeing,

for the first time, the possibility of producing functional prion-like fibrils at an industrial scale.

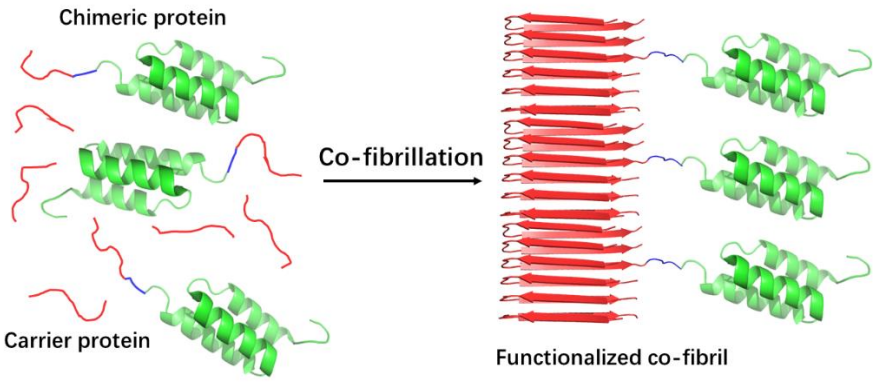

Figure 8. Schematic illustration of co-fibrillation to immobilize globular domains into amyloid fibrils with defined doping frequencies. The chimeric protein consists of a prion-like domain (red) and an Z domain (green, antibody binding domain from Protein A, PDB 1Q2N), linked via a flexible linker (blue). The chimeric protein is fibrillated with the carrier protein (the prion-like domain alone), in order to allow enough space between globular domains in the final fibrils. This reduces steric constraints and maximizes the functionality.

Biocatalytic cascades. As discussed above, the chemistry-free covalent enzyme immobilization and the high surface to volume ratio of prion-based fibrils make of these nanostructures optimal recyclable catalyzers. However, many industrially relevant processes require two or more coupled reactions, and, in most cases, only single-enzyme functionalized fibrils have been characterized. Mats Sandgren and co-workers addressed this issue by building up catalytic cascade by genetically fusing xylanase A, β-xylosidase, and aldose sugar dehydrogenase to the Sup35 PrD to create three different Sup35–enzyme chimeras.[149] As in the case of the Z-domain, the fibrils were doped with the PrD alone to introduce enough space between the functional enzyme molecules to eliminate steric restrictions. The objective was to convert beechwood xylan to xylonolactone, a valuable chemical for versatile applications, in three discrete enzymatic steps. Each type of functional fibril was formed individually, and then they were mixed in different proportions to maximize the catalytic efficiency. These

investigations revealed that a sequential cascade in which fibrils of xylanase A and β-xylosidase were first mixed, the resulting product of the coupled reaction removed and then incubated with fibrils of aldose sugar dehydrogenase was more productive than a mixture containing all the three kinds of fibrils together since the immobilized enzymes displayed hardly compatible stabilities. Despite its limitations, this work constitutes of nice proof-of-concept for prion-inspired biocatalytic cascades.

6 Soft amyloid cores in PrDs and their use in nanomaterials

The absence of highly amyloidogenic sequences allows PrD to transit between soluble and aggregated states.[150] The aggregation of prions seems to depend on the establishment of a large number of weak interactions distributed along the complete PrD.[151] Accordingly, natural PrDs often consist of > 100 residues and 60 residues was traditionally considered to be the minimum length required to attain an efficient PrD self-assembly at moderate protein concentrations.[115] With this size, the PrD might constitute a significant fraction of the fusion protein, especially when it is fused to small globular domains, like the Z-domain. This results in steric restrictions in the fibrillar state since the amyloid spine may have a larger transversal dimension that the adjacent protein, a property that has been reported to reduce the diffusion of substrates and limit the functionality of the fibril.[139] Besides, large PrDs often compromise the solubility and stability of the appended globular domains in the monomeric fusions state, impacting the recombinant production of the chimeric proteins and the storage after their purification.

We have recently proposed that inner weak amyloidogenic sequence stretches within PrD might contribute to nucleate the conformational conversion of prionic proteins into a cross-β structure.[152] These soft amyloid cores (SACs) (**Figure 9**), were first identified in yeast prions,[153] but they are also present in a significant number of human prion-like proteins.[154] SACs differ from the classical short amyloid cores of pathogenic proteins, which hold a high aggregation propensity and are typically enriched in hydrophobic

residues. SACs are slightly longer and more polar, resulting in a less concentrated aggregation potential. This allows the PrD to remain soluble under most physiological conditions, but also to hold a cryptic aggregation propensity that might facilitate its efficient self-assembly in response to cellular changes.[150] The SAC from Sup35p PrD consists of 21-residues and self-assembles spontaneously into highly ordered amyloid fibrils. These fibrils can seed the amyloid formation of the complete PrD *in vitro*. Also, when the *in vitro* formed SAC fibrils are added to yeast cells, they promote the aggregation of the endogenous full-length Sup35p and the subsequent emergence of a prionic phenotype.[155] Moreover, this region is necessary for the induction, propagation, and inheritance of the prion state of Sup35p in mammalian cytosol.[156] All these data converge to suggest that this short sequence stretch might have prion-like properties, which lead us to hypothesize that SACs can substitute the role of complete PrDs and recapitulate the self-assembly properties of the larger domain in the context of modular fusion proteins.

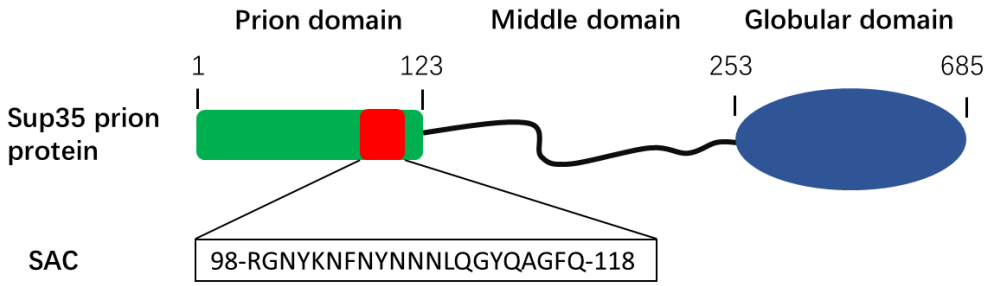

Figure 9. Soft amyloid core within prion domain of Sup35p. Sup35p consists of a prion domain (green box), charged middle domain (black line) and C-terminal globular domain (blue oval). The identified soft amyloid core (SAC, red box) corresponds to 21-residues long sequences with Q/N-rich composition.

Importantly, not only a vast diversity of natural SACs can be identified computationally,[157] but these stretches can also be artificially engineered and minimized, provided that we keep their main physicochemical traits. For instance, we have designed a family of minimalist SAC-inspired polar heptapeptides that self-assemble

into amyloid fibrils. The nice thing of these nanowires is that, if they are adequately designed, they can be endorsed with intrinsic catalytic activity, without the need for an adjacent globular domain.[158]

The genetic protein fusion strategy has resulted in the development of PrD-based functionalized nanomaterials with exceptional properties as enzyme immobilizers, biosensors, bioelectrodes or for antibody purification and antigens detection. We are convinced that the application of these unique materials in fields like nanomedicine or nanotechnology will expand significantly in the forthcoming years, mainly because the use of natural and artificial SACs might allow us to potentially endorse nanofibrils with a new set of previously inaccessible functionalities.

Research objectives

Chapter I

Conventional PrDs are challenging to use in order to obtain functional amyloids, because their redundant sequences and large sizes compromise the solubility of the adjacent globular proteins and result in steric restrictions in the fibrillar state, reducing substrates diffusion and significantly limiting the activity of the target proteins. In our previous studies, we showed that PrDs could be substituted by their shorter soft amyloid cores (SACs), in such a way that the SACs will recapitulate the self-assembly properties of the much longer sequence. To develop an application for such property, we will genetically fuse the SAC within Sup35p yeast prion to three globular proteins displaying different structures, namely the all-α FF domain, the all-β green fluorescent protein, and the α/β carbonic anhydrase enzyme.

1. Design of Sup35-FF fusion protein and computational prediction of its solubility.
2. Expression and conformation characterization of Sup35-FF fusion protein.
3. Stability and folding kinetics characterization of Sup35-FF fusion protein.
4. Characterization of the amyloid properties of incubated Sup35-FF fusion protein.
5. Use of a Sup35-GFP fusion protein to obtain fluorescent amyloid fibrils.
6. Use of Sup35-CA fusion protein to generate biocatalytic amyloid fibrils.
7. Building up a molecular model for Sup35-GFP fusion as a proof-of-concept.

Chapter II

A relationship between the size of globular domains and the length of the minimalist linker that allows the formation of amyloid fibrils when fused to an amyloidogenic region (AR) has been previously established by using molecular and mesoscopic modeling. In this study, we wanted to provide further insight into this relationship. The objective of the present work was to computationally study the self-assembly of ARs when the linker between this segment and the globular domain is shorter than the one allowing the formation of the infinite fibrils.

1. Modeling the self-assembly of a hybrid protein ($A\beta_{17-42}$-GFP) containing an AR and typical globular domain by using full-atom and coarse-grained targeted molecular dynamics.
2. Using rigid body simulation, as a simplified approach, to develop a more general case and determine the relationship between the number of subunits in the assembly and the number of residues in the linker.

Chapter III

As we have proposed in Chapter II, a hybrid protein consisting of an amyloidogenic region (AR) and a globular domain would likely self-assemble into oligomers instead of amyloid fibrils when the linker between the AR and the globular domain is shorter than the one allowing the formation of the infinite fibrils. This study's objective was to obtain amyloid assemblies with different mesoscopic structures by playing with the length of the linker between the soft amyloid core (SAC) and the globular domain. To this aim, we will build up two different constructs consisting of a Sup35 SAC and dihydrofolate reductase (DHFR) separated by either a 5-residue or 8-residue long linker. Then, we will exploit the potential of these amyloid assemblies for building up multifunctional materials for targeted drug delivery by using a tandem fusion protein also containing the antibody binding Z-domain.

1. Expression and conformation characterization of the fusion proteins.
2. Characterization of the amyloid properties and morphology of the incubated fusions.
3. Characterization of the catalytic activity of DHFR when embedded in the amyloid assemblies.
4. Design of a bifunctional fusion by incorporation of the Z domain of protein A: Sup35-5aa-DHFR-Z
5. Amyloid properties and morphology characterization of incubated Sup35-5aa-DHFR-Z fusion.
6. Cargo and antibody binding abilities of Sup35-5aa-DHFR-Z amyloid nanoparticles.
7. Targeting, release, and toxicity for cancer cells of DHFR inhibitors transported by the Sup35-5aa-DHFR-Z nanoparticles.

Chapter IV

Bispecific antibodies (BsAbs) have been widely developed as therapeutic drugs for the treatment of different diseases. In this study, we will build up specific antibody conjugated amyloid fibrils, aiming to act as mimetics of BsAbs, using an antibody binding hybrid protein consisting of Sup35 soft amyloid core (SAC) and the antibody binding Z domain (Sup35-Z). As in our previous studies, we expected Sup35-Z to self-assemble into amyloid fibrils in which the Z domain would remain in its native structure and preserve its antibody binding activity. Ideally, two monospecific antibodies (anti-EGFR and anti-CD3) conjugated amyloid fibrils would direct T cells to HeLa cells and prove the usefulness of these proteins for immunotherapeutic approaches.

1. Expression, conformation, and stability of Sup35-Z.
2. Aggregation properties and morphology of incubated Sup35-Z fusion.
3. Antibody binding affinity and capacity of Sup35-Z fibrils.
4. Functionality of conjugated antibody on Sup35-Z fibrils.
5. Building up double conjugated nanofibrils.
6. Engineering of the size of the fibrils to obtain rod-like nanofibrils and assessing the cytotoxicity of the nanorods.
7. Assessing if single antibody-conjugated Sup35-Z nanorods target the desired antigens in cells.
8. Testing if double antibody-conjugated Sup35-Z nanorods direct T cells to Hela cells.

Chapter I

Prion soft amyloid core driven self-assembly of globular proteins into bioactive nanofibrils

*Weiqiang Wang[1], Susanna Navarro[1], Rafayel A. Azizyan[2], Manuel Baño-Polo[1], Sebastian A. Esperante[1], Andrey V. Kajava[2] and Salvador Ventura[1]**

[1]Institut de Biotecnologia i de Biomedicina and Departament de Bioquímica i Biologia Molecular; Universitat Autònoma de Barcelona; 08193 Bellaterra (Barcelona), Spain.

[2]Centre de Recherche en Biologie cellulaire de Montpellier, UMR 5237 CNRS, Université Montpellier, 1919 Route de Mende, 34293 Montpellier, Cedex 5, France

E-mail: salvador.ventura@uab.es

Abstract

Amyloids have been exploited to build up amazing bioactive materials. In most cases, short synthetic peptides constitute the functional components of such materials. The controlled assembly of globular proteins into active amyloid nanofibrils is still challenging, because the formation of amyloids implies a conformational conversion towards a β-sheet-rich structure, with a concomitant loss of the native fold and the inactivation of the protein. There is, however, a remarkable exception to this rule: the yeast prions. They are singular proteins able to switch between a soluble and an amyloid state. In both states, the structure of their globular domains remains essentially intact. The transit between these two conformations is encoded in prion domains (PrDs): long and disordered sequences to which the active globular domains are appended. PrDs are much larger than typical self-assembling peptides. This seriously limits their use for nanotechnological applications. We have recently shown that these domains contain soft amyloid cores (SACs) that suffice to nucleate their self-assembly reaction. Here we genetically fused a model SAC with different globular proteins. We demonstrate that this very short sequence act as minimalist PrDs, driving the selective and slow assembly of the initially soluble fusions into amyloid fibrils in which the globular proteins keep their native structure and display high activity. Overall, we provide here a novel, modular and straightforward strategy to build up active protein-based nanomaterials at a preparative scale.

Keywords: protein self-assembly, amyloid fibrils, prion domain, soft amyloid core, functional amyloids, nanomaterials.

Introduction

The formation of amyloid fibrils is associated with the onset of a range of protein misfolding diseases.[1] However, amyloid structures are also exploited for functional purposes by different organisms.[2,3] The inner intermolecular β-sheet structure of the fibrils makes these protein assemblies very stable,[4] even in harsh environments. This, together with their tuneable assembly under physiological conditions, make them attractive modules to build up nanomaterials for biomedical and biotechnological applications, including catalysis,[5] biosensors,[6] electronics,[7] tissue engineering[8] or drug delivery.[9] Most efforts so far have been focused on the use of short synthetic peptides as the bioactive components of such materials,[10, 11] while an analogous approach for inducing globular proteins to assemble into functional nanofibres has been much less explored.

The main limitation to create amyloids that display functional proteins comes from the connection between protein function and the attainment and maintenance of a defined folded state. Protein folding and aggregation are two competing reactions which depend on overlapping physicochemical properties.[12] In a large majority of cases, the aggregation of the protein into amyloid fibrils implies a conformational conversion in which it losses the native fold and gains β-sheet structure,[1] becoming thus inactive in the aggregated state.

The design of fibrillar structures containing properly folded domains appears as a challenging task, since the protein of interest should fulfil contradictory properties: it should remain soluble and folded during recombinant expression and subsequent purification and storage, but, at the same time, it should be able to self-assemble into β-sheet rich amyloid-like structures and, moreover, this must occur without a structural conversion of the globular domains. A way to bypass these limitations consists in the design of hybrid structures in which purified folded domains are chemically linked to preformed fibrillar structures.[13] However, this strategy is deterred by the limited

available polypeptide conjugation chemistries, their cross-reactivity and the unavoidable reduction in the proportion of conformationally active molecules in the assembly.

Nevertheless, functional proteins with intrinsic self-assembling properties already exist in nature, among them, the yeast prions. Yeast prions are proteins expressed and stored in the cell in a soluble state, but they are able to self-associate into amyloid structures under certain conditions.[14] A common feature of these proteins is the presence of a prion domain (PrD). PrDs correspond to intrinsically unstructured sequences of low complexity highly enriched in asparagine (N) and/or glutamine (Q) residues[15] and are accompanied by one or more globular domains at their C- or N-terminus.[16] The PrD is both necessary and sufficient for self-assembly, whereas the globular domains account for the protein activity.[17] The nice thing here is that there are evidences that only the PrD is integrated in the core of the fibril, whereas the globular domains hang from the fibril in a folded conformation.[18]

Because yeast prions are modular, one can, in principle, fuse any globular domain to a given PrD and potentially obtain a functionalized fibrillar nanomaterial. In this way, enzymes like alkaline phosphatase (AP)[19] or carbonic anhydrase (CA)[20] have been genetically fused to the PrD of the yeast prion Ure2p and methyl-parathion hydrolase (MPH) to the PrD of Sup35 prion.[21] The PrDs were shown to drive the association of the correspondent fusion proteins into amyloid fibrils, generating catalytic nanomaterials that displayed folded enzyme moieties.

The required equilibrium between solubility and aggregation propensities explains the absence of highly amyloidogenic sequences in PrDs,[22] fibrillation being thought to rely on the establishment of a large number of weak interactions distributed along the complete low complexity sequence.[23] As a consequence, PrDs are much larger than the majority of self-assembling peptides used for nanotechnological applications.[24] Traditionally, a length of at least 80 residues has been considered necessary for the

conversion of PrDs into amyloids.[25] Thus, the PrD would constitute a significant fraction of any protein fusion, especially when fused with small globular domains. In addition, the large size of PrDs results in steric hindrance in the fibrillar state, reducing substrates diffusion and significantly limiting the activity of the adjacent domains, relative to their soluble counterparts.[19] Indeed, in the natural Ure2p yeast prion, the globular GST domain becomes inactive in the fibrils, despite its native-like structure.[20]

We have recently shown that, in addition to a distinctive amino acidic composition, PrDs contain inner weak amyloidogenic sequence stretches that contribute to trigger the initial protein self-assembly reaction.[26] These cryptic amyloids promote conformational conversion in bona fide yeast prions,[27] but they also exist in human prion-like proteins.[28] The soft amyloid cores (SAC) embedded within PrDs can be identified computationally. [29] They differ from the classical amyloid cores of pathogenic proteins in that they are slightly longer and more polar, in such a way that the amyloid potential is less concentrated, allowing the containing PrD to remain soluble under most physiological conditions, while still keeping a certain amyloid propensity that might facilitate its assembly in certain circumstances.[30]

We hypothesized here, that complete PrDs can be substituted by their SACs, in such a way that these shorter sequences would recapitulate the larger domain self-association properties in the context of modular fusion proteins. It is expected that the polar nature of SAC would prevent aggregation and misfolding of the adjacent globular domain during recombinant expression, purification and storage, while still being able to induce its self-assembly under controlled conditions. Ideally, in the resulting amyloids fibrils, the SAC would form the core of the fibril, whereas the attached globular domain would remain folded and active. As a proof of principle, we selected the canonical yeast prion Sup35, an eukaryotic translation release factor.[31] Its SAC corresponds to a 21-residues long sequence stretch that autonomously self-assembles into highly ordered amyloid fibrils, displaying a typical cross-β diffraction pattern. This

sequence is able to seed amyloid formation by the entire PrD, *in vitro*, and of the complete protein, *in vivo*,[32] and is indispensable for the induction, propagation and inheritance of the prion state in the mammalian cytosol.[33] We show here, how the properties of this short sequence stretch can be exploited to obtain highly functional amyloid-like nanofibers. We illustrate the wide applicability of the approach by fusing this SAC to three globular proteins displaying different structure; namely the all-α FF domain, the all-β green fluorescent protein and the α/β carbonic anhydrase enzyme.

Results and discussion

Design of a self-assembling protein fusion consisting of Sup35-SAC and a small all α-helical domain

The 21 residues-long SAC of Sup35 PrD (Sup35-SAC) corresponds to residues 98-118 of the yeast protein (**Figure 1A**). We decided to exploit the prion-like characteristics of this small segment to design prion-inspired self-assembling fusion proteins, consisting of a chimera of Sup35-SAC and different globular proteins. As a proof of principle, we used the FF domain of the URN1 splicing factor.[34]

FF domains are small protein-protein interaction modules of 50-70 residues, characterized by a fold that consists of three α-helices arranged as an orthogonal bundle with a 3_{10} helix in the loop connecting the second and the third helices (**Figure 1B**).[35] FF domains are highly soluble and the folding landscape of several of these proteins, including the URN1-FF domain (57 residues), have been characterized in detail.[36,37] Thus, it constitutes a perfect model protein to assess: *i*) whether the fusion of Sup35-SAC to a globular domain impacts its solubility, conformation, thermodynamic stability and/or folding kinetics and, *ii*) if this short segment suffices to drive the self-assembly of a soluble globular protein into amyloid-like structures, allowing to maintain its native structure in the aggregated state.

We fused Sup35-SAC to the N-terminus of the FF domain. A flexible linker consisting of SG_3SG_2S was incorporated between both protein moieties (**Figure 1C**) and a His6 tag at the FF domain C-terminus. In this fusion, the ratio between the size of the self-assembling sequence and the globular domain is 0.4, whereas in the case of the full-length Sup35 PrD (114 residues) instead, this ratio would have increased to 2.0.

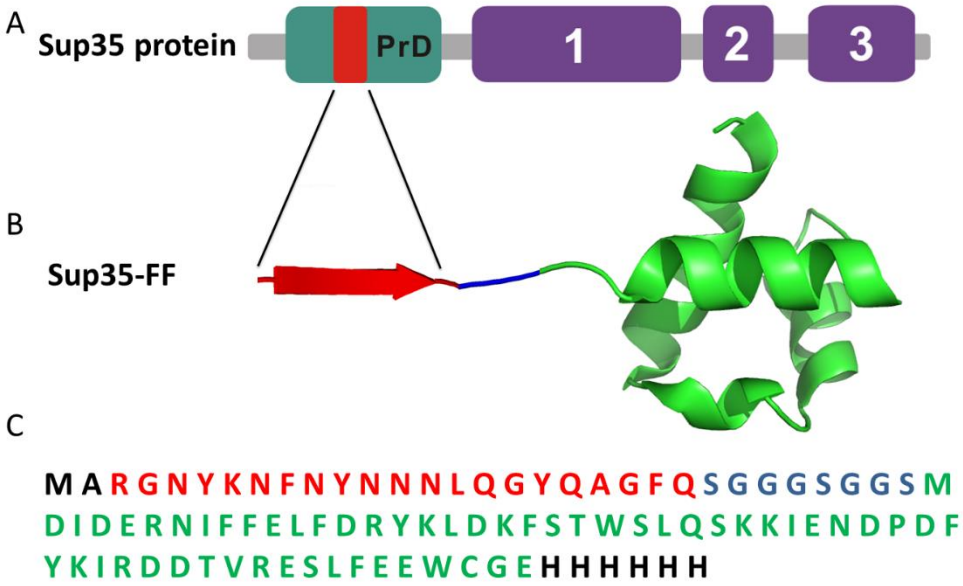

Figure 1. Schematic representation of Sup35-FF fusion protein. (A) Domain organization of Sup35 protein: soft amyloid core (SAC) (red) within the PrD (green) and the respective elongation factor functional domains (purple) in Sup35 are shown. (B) Cartoon of Sup35 soft amyloid core fused to FF domain shown as ribbon representation (PDB: 2JUC). (C) Sequence of the Sup35-FF fusion protein. Sup35 soft amyloid core, spacer linker, FF domain and His6 tag are shown in red, blue, green and black, respectively.

Previous studies showed that proteins appended with β-sheet forming peptides aggregated into insoluble inclusion bodies during expression.[38,39,40] Thus, we first assessed computationally if the addition of Sup35-SAC to the FF domain would impact its inherent solubility, compromising thus its expression, using three different aggregation propensity algorithms: AGGRESCAN[41], FoldAmyloid[42] and PASTA[43], all them predicting Sup35-SAC as having low aggregation propensity, compared with classical amyloid cores (**Table S1A**). Indeed, 57 % of the residues in Sup35-SAC are polar with only 19 % of them being strictly hydrophobic. The rest correspond to a 14 % of Tyr, sharing both characters, and another 14 % of Gly, a residue with very low aggregation propensity. Accordingly, Sup35-SAC is not expected to impact

significantly the solubility of the globular protein in the fusion (**Table S1A**). Therefore, we proceeded to construct the fusion protein (Sup35-FF) and expressed it in *E. coli*. As predicted, the protein was almost completely located in the soluble cell fraction (**Figure S1**), from which it was purified by IMAC and gel-filtration chromatography (**Figure S2A**). The yield was 68 mg of purified fusion per L of culture, much higher than that reported for other fusions between amyloidogenic peptides of similar size and globular proteins, where a significant fraction of the protein remained insoluble.[44]

Sup35-SAC does not affect the conformation of the FF domain in the protein fusion

A first requirement to use Sup35-FF for building up functional nanofibers is that the N-terminal Sup35-SAC does not alter the conformation, stability or folding properties of the globular domain.

We compared the conformational properties of Sup35-FF and the FF domain alone (FF-wt) at pH 7.4 and 25 ºC by monitoring the far-UV CD spectra and Trp intrinsic fluorescence. The far-UV CD spectra of the two proteins at different concentrations closely resemble and are dominated by α-helical signals (**Figure S2B**). The intrinsic fluorescence spectrum of FF-wt exhibits an emission maximum at 337 nm. In Sup35-FF, it is shifted to 334.5 nm (**Figure S2C**), suggesting the Trp being in a slightly more protected environment. The FF domain possesses two Trp residues at positions 27 and 56. Trp27 is buried, while Trp56 is structurally adjacent to the N-terminal α-helix and partially exposed to solvent. Four different disorder prediction algorithms: PONDR (VSL2)[45], GlobPlot[46], PASTA[43], and IUPred[47] suggest that the Sup35-SAC and the linker region will remain disordered in the context of the fusion protein (**Table S1B**). Therefore, it is likely that the structural fluctuations of the disordered N-terminal tail might shield, at least partially, the exposed Trp56 side chain in the FF domain, as we already observed for human SUMO domains.[48] In any case, the shift towards the blue

region of the spectrum is indicative of a properly folded domain, since upon destabilization and partial unfolding, the Trp spectrum of the FF domain shifts to the red, displaying a large increase in its fluorescence emission.[36]

Sup35-SAC does not affect the thermodynamic stability of the FF domain in the protein fusion

Despite the native-like conformation of the FF domain in the protein fusion, the adjacent exogenous sequence might compromise the protein stability, promoting partial unfolding. To the best of our knowledge, the stability of globular domains when attached to exogenous PrDs or amyloidogenic sequences has not been yet addressed, likely because these fusions begin to aggregate soon after purification.

The equilibrium unfolding of Sup35-FF and FF-wt proteins was analysed at pH 7.4 and 25 °C. The urea denaturation curves at equilibrium were obtained recording the changes in Trp intrinsic fluorescence at 350 nm (**Figure 2C**) and in molar ellipticity at 222 nm (**Figure 2D**) at increasing denaturant concentrations. Both proteins displayed a single visible transition indicative of a cooperative unfolding process. The main thermodynamic parameters of the unfolding reaction were calculated from the equilibrium curves assuming a two-state model (R > 0.99) (**Table S2**). The stabilities of Sup35-FF calculated from fluorescence and CD measurements were similar with $\Delta G_{U-F}^{H_2O}$≈ 4.03 kcal/mol and a $[\text{Urea}]_{50\%}$ of ≈ 5.47 M. FF-wt shows similar values, with a $\Delta G_{U-F}^{H_2O}$ ≈ 4.15 kcal/mol and $[\text{Urea}]_{50\%}$ of ≈ 5.57 M, indicating that the N-terminal extension has a minor impact in the domain's chemical stability (**Table S2**).

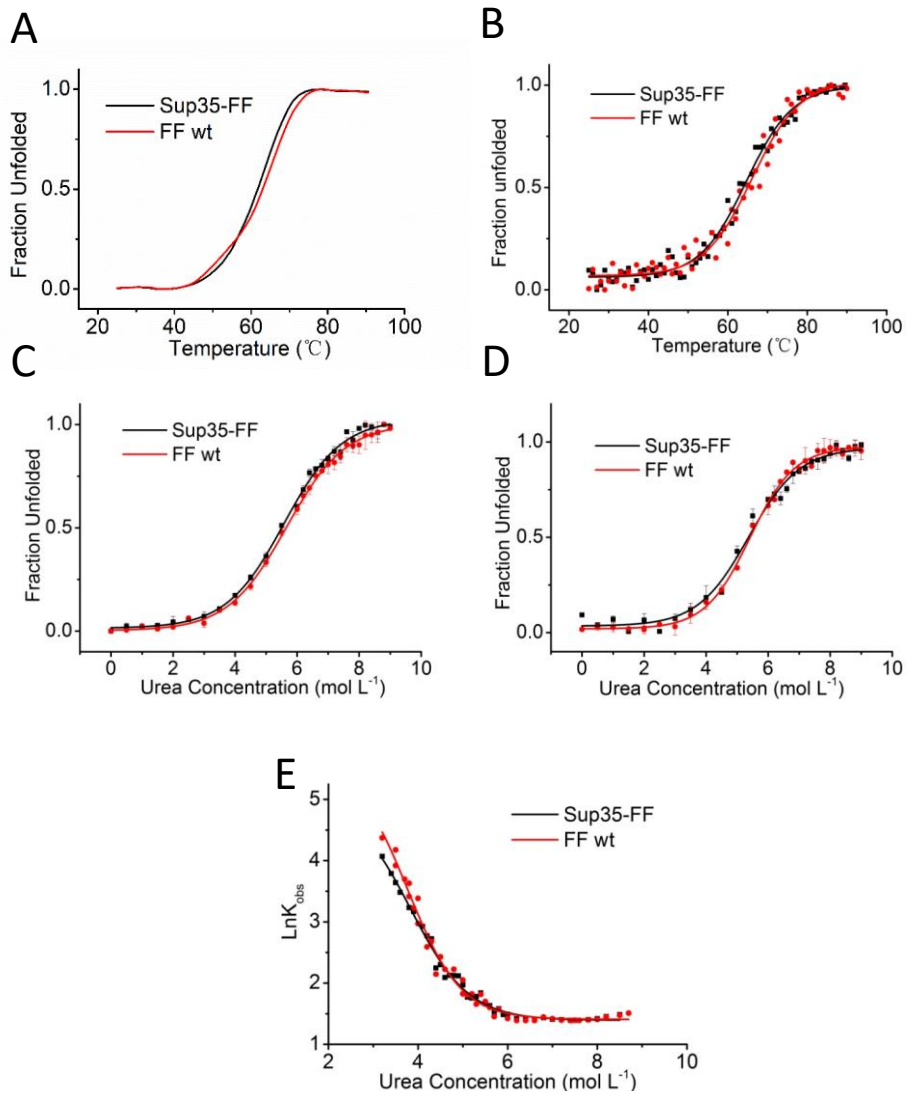

Figure 2. Stability and folding properties of Sup35-FF and FF-wt proteins. Thermal stabilities were analysed by (A) Trp intrinsic fluorescence emission at 350 nm and, (B) far-UV CD signal at 222 nm. Chemical equilibrium curves with urea were followed at 25 ºC by (C) Trp intrinsic fluorescence at 350 nm and, (D) far-UV CD at 222 nm. (E) The kinetics of folding and unfolding were measured by Trp intrinsic fluorescence at 25 ºC performing stopped-flow experiments. The rate constants were calculated under conditions of apparent two-state folding.

Thermal unfolding of Sup35-FF and FF-wt at pH 7.4 was followed by Trp intrinsic fluorescence (**Figure 2A**) and far-UV CD (**Figure 2B**), monitoring the changes in molar ellipticity and fluorescence emission at 222 nm and 350 nm, respectively.

Thermal denaturation curves show a single cooperative transition. The melting temperatures obtained for Sup35-FF were T_m of 64.7 ºC and 62.7 ºC by far-UV CD and intrinsic fluorescence, respectively. For FF-wt we recorded T_m of 66.1 ºC and 64.6 ºC by far-UV CD and intrinsic fluorescence, respectively. Therefore, Sup35-SAC has only a minor effect on the thermal stability of the FF domain (**Table S3**).

Sup35-SAC does not affect the folding and unfolding kinetics of the FF domain in the protein fusion

As far as we know, the folding properties of a globular domain when embedded in a fusion with an aggregation-prone segment have never been assessed. Here, we determined the kinetics of folding and unfolding of FF domains by stopped-flow at pH 7.4 and 25 ºC by monitoring the changes in Trp intrinsic fluorescence under a wide range of urea concentrations. For both Sup35-FF and FF-wt, the folding and unfolding fluorescence traces fit well to a single exponential function. Chevron plots appear to be linear in the entire range of denaturant concentrations studied and fit well to a two-state model, indicating the absence of detectable intermediates (**Figure 2E**). 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 1**. The folding and unfolding rates of Sup35-FF are only a 24% slower and an 8% faster than that of FF-wt, respectively. According to this kinetic rates, Sup35-FF is only about 0.2 kcal/mol more stable alone than in the fusion, in good agreement with the equilibrium data.

Table 1. Folding kinetics parameters of Sup35-FF and FF-wt proteins.

	$\Delta G_{U-F}^{H_2O}$ [a] (kcal mol^{-1})	m_{U-F} [b] (kcal mol^{-1} M^{-1})	C_m [c] (M)	k_F (s^{-1})	k_U (s^{-1})	RTm_U (kcal mol^{-1} M^{-1})	RTm_F (kcal mol^{-1} M^{-1})
Sup35-FF	4.32±0.13	0.79±0.08	5.50±0.56	2835±260	1.93±0.25	0.057±0.009	0.728±0.083
FF-wt	4.53±0.16	0.78±0.04	5.74±0.40	3714±733	1.78±0.12	0.063±0.004	0.719±0.043

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

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

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

Overall, the data presented in this and the previous sections clearly indicate that, in the soluble state, the FF domain maintains its fold, stability and folding properties almost intact when fused to Sup35-SAC.

Sup35-SAC promotes the assembly of the Sup35-FF fusion into amyloid fibrils under mild conditions

We used the amyloid-specific dyes Thioflavin-T (Th-T) and Congo Red (CR) to asses if the Sup35-FF protein fusion self-assembles with time as amyloid-like structures under mild conditions. To this aim, Sup35-FF and FF-wt were incubated at pH 7.4 and 37 ºC for one week. Th-T is a dye which fluorescence emission maximum at 488 nm increases in the presence of amyloid-like structures.[49] The presence of incubated Sup35-FF promoted a large increase of Th-T fluorescence emission signal, whereas FF-wt had a negligible effect (**Figure 3A**). In agreement with these results, CR binding was observed for Sup35-FF, resulting in a clear red-shift of CR absorption spectrum, indicative of the dye binding to an amyloid structure[50], whereas FF-wt did not promote any spectral shift (**Figure 3B**). The morphological analysis of the two protein solutions by negative-staining and transmission electron microscopy (TEM) confirmed the

presence of characteristic long and unbranched amyloid fibrils of 16.0±1.7 nm in width for Sup35-FF (**Figure 3D**), whereas the FF-wt solution did not show any detectable ordered aggregate (**Figure 3C**).

We further analysed the assembly of Sup35-FF into amyloid fibrils by monitoring the changes in Th-T signal with time. To this aim, the reaction was set up in 96 well plates with constant agitation in the presence of Teflon beads, a condition that accelerates aggregation and allows a continuous reading of the dye signal.[51] The kinetics of amyloid fibril formation can be usually adjusted to a sigmoidal curve, reflecting the existence of a nucleation-dependent growth reaction.[52] The aggregation of Sup35-FF under native conditions follows this kinetic scheme, exhibiting a lag phase of ≈ 6 h and being completed after ≈ 17 h (**Figure S3**). A characteristic of most amyloid assemblies is that they are able to seed and accelerate the reaction of their soluble counterparts. Importantly, the presence of 2% Sup35-FF preformed fibrils was enough to dramatically accelerate the aggregation reaction, reducing the lag phase to ≈ 1.5 h and allowing it to complete in ≈ 7 h (**Figure S3**).

Overall, the data converge to indicate that Sup35-SAC is able to specifically promote the self-assembly of the Sup35-FF fusion, containing the otherwise soluble FF domain, into ordered nanofibrils.

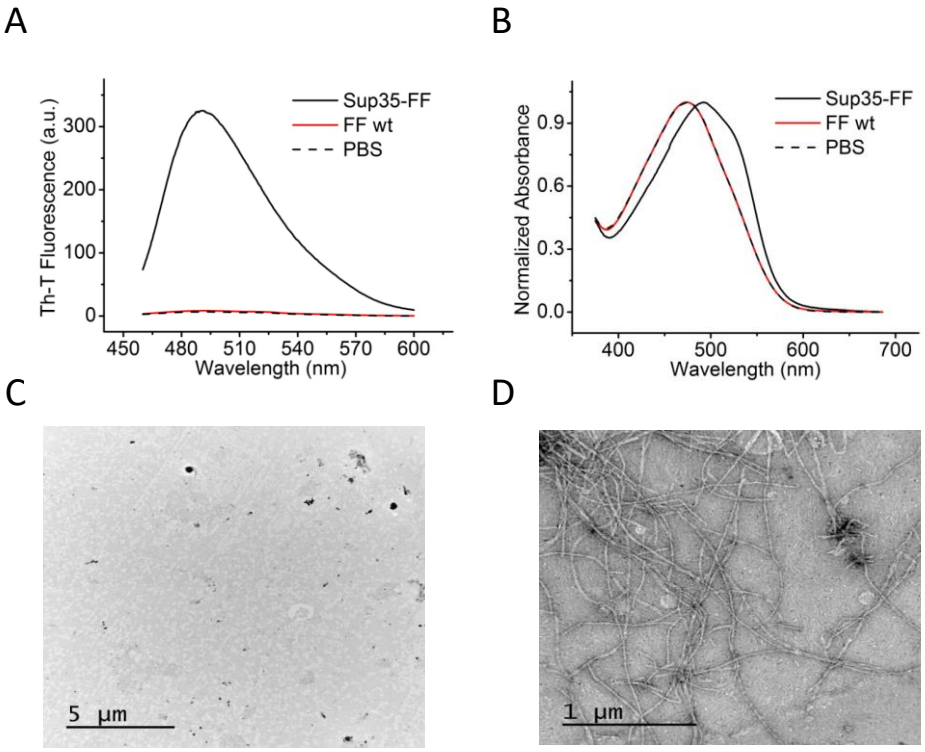

Figure 3. Characterization of Sup35-FF fibrils. Sup35-FF and FF-wt solutions were incubated for one week and analysed by measuring (A) Th-T fluorescence emission and (B) Congo red absorbance. FF-wt and Sup35-FF are shown in red and black, respectively. PBS without protein was included as a control (dashed line). Representative TEM micrographs of incubated proteins upon negative staining: (C) FF-wt and, (D) Sup35-FF. The scale bar represents 5 µm and 1 µm, respectively

Secondary structure of Sup35-FF amyloid fibrils

The modular nature of yeast prions accounts for the fact that globular domains retain its original fold in the amyloid state.[53] We wondered whether Sup35-SAC might act as a mimic of the 5-fold longer Sup35 PrD, driving the self-assembly of protein fusions into amyloid structures, but allowing to maintain the native structure of the adjacent globular domain. Because the conversion of a globular protein into an amyloid implies a conformational conversion into a β-sheet rich structure, independently of the initial 3D-conformation, we took profit of the all α-helical nature of the FF domain to

monitor if it remains folded in the amyloid assembly, using Attenuated Total Reflectance Fourier Transform Infrared spectroscopy (ATR-FTIR).

We recorded the infrared spectra of Sup35-FF amyloid fibrils in the amide I region of the spectrum (1700-1600 cm^{-1}) (**Figure 4A**). This region corresponds to the absorption of the carbonyl peptide bond group of the protein main chain and is sensitive to the peptide conformation. Deconvolution of the spectra allows to assign the secondary structure elements and their relative contribution to the main signal (**Table S4**). The spectra of Sup35-FF displayed signals indicative of the formation of β-sheet structure, coming from the intermolecular β-sheet region at 1624 cm^{-1} contributing to 27.25 % of the area. However, the largest contribution to the spectra, accounting a 59.5 % of the area, comes the band at 1649 cm^{-1}, compatible with the presence of native-like α-helices,[54] in the Sup35-FF amyloid fibrils. This putative helical band fits well with the major band at 1650 cm^{-1} observed in the ATR-FTIR spectra of the soluble FF-wt domain (**Figure 4A and 4B** and **Table S4**). No band indicative of antiparallel β-sheet (1685-1690 cm^{-1}) was observed.

In the Sup35-FF fusion the two Trp residues are located in the FF domain α-helices; because the core of the amyloid fibrils is expected to be resistant to urea denaturation, or at least more resistant than the helices, following urea denaturation by intrinsic fluorescence should allow us to assess if the Trp residues remain in a native-like context in the fibrillar state. As it can be seen in **Figure 4C** the denaturation curve of aggregated Sup35-FF is cooperative and resembles that of soluble FF-wt, despite the calculated $[Urea]_{50\%}$ is slightly higher for the fibrils (5.86 ± 0.11 M) than for the soluble wild type protein (5.68± 0.04 M), both calculated from the changes in intrinsic fluorescence. These data support the globular domain remaining folded in the amyloid assembly.

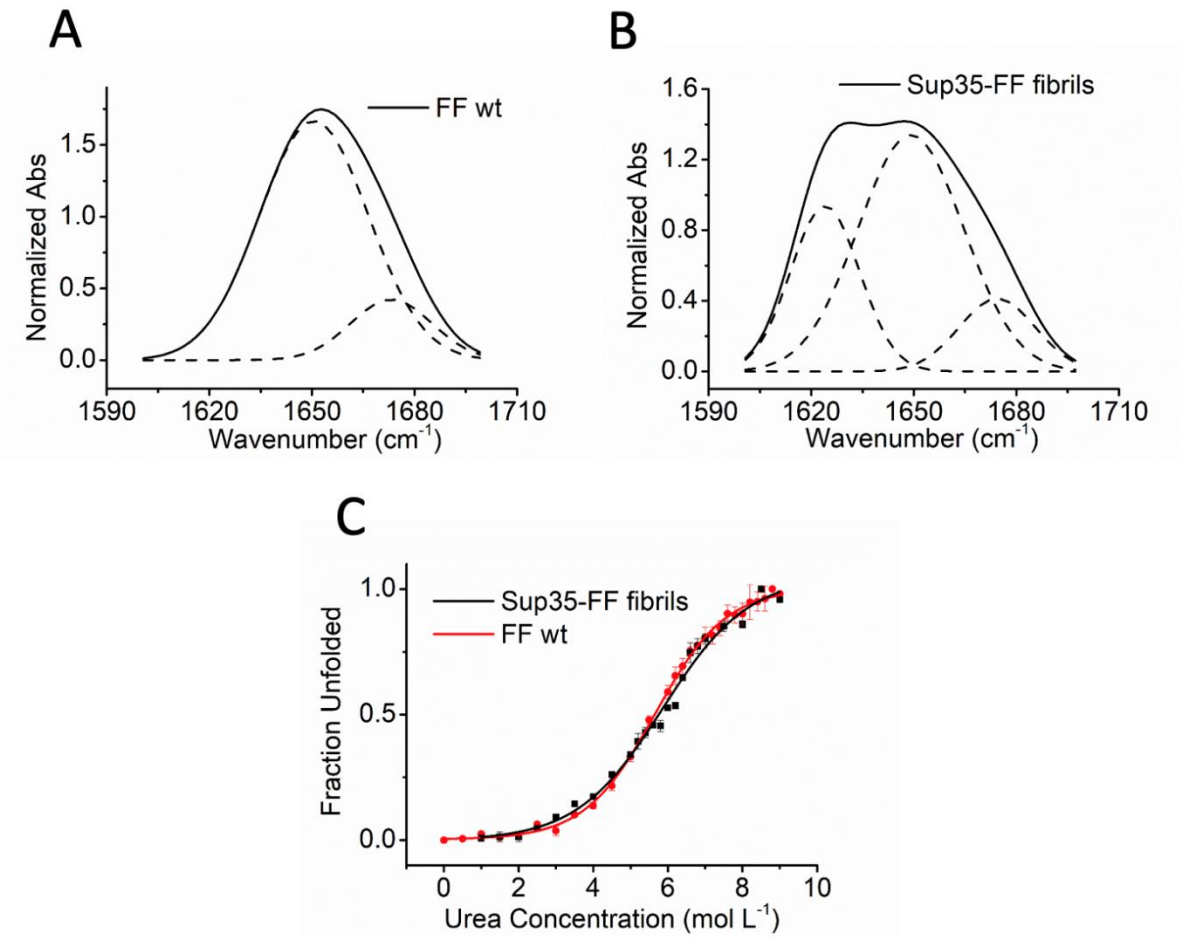

Figure 4. Conformational properties of Sup35-FF fusion protein fibrils. Sup35-FF proteins solutions were incubated for one week. The absorbance spectra of (A) Soluble FF-wt and (B) Sup35-FF fibrils in the amide I region (solid line) and the components bands (dashed lines) are shown. (C) The chemical equilibrium curve with urea for aggregated Sup35-FF was followed at 25 ºC by Trp intrinsic fluorescence at 350 nm. Soluble FF-wt was measured in the same conditions as a control.

Sup35-SAC drives the formation of GFP fluorescent amyloid fibrils

The above described data indicates that the short Sup35-SAC allows the expression and purification of soluble fusions in which the adjacent C-terminal globular domain remains properly folded as well as the self-assembly of the protein into nanofibers that retain detectable native structure. On this basis, we explored whether

we can obtain fluorescent amyloid fibrils by fusing Sup35-SAC to GFP (Sup35-GFP) (**Figure S4A).**

Sup35-GFP was purified from the soluble cellular fraction (**Figure S5**) with a yield of 85 mg/L culture. We characterized the conformational features of the GFP moiety in the fusion by monitoring GFP absorption and fluorescence emission spectra (**Figure 5A and 5B**) as well as by far-UV CD (**Figure 5C**) at pH 7.4 and 25 ºC. Sup35-GFP showed spectral properties indistinguishable from that of GFP alone, with an absorption maximum at ~ 490 nm, maximum fluorescence emission at ~511 nm, and a far UV-CD spectrum exhibiting a single minimum at ~218 nm characteristic of β-sheet proteins.

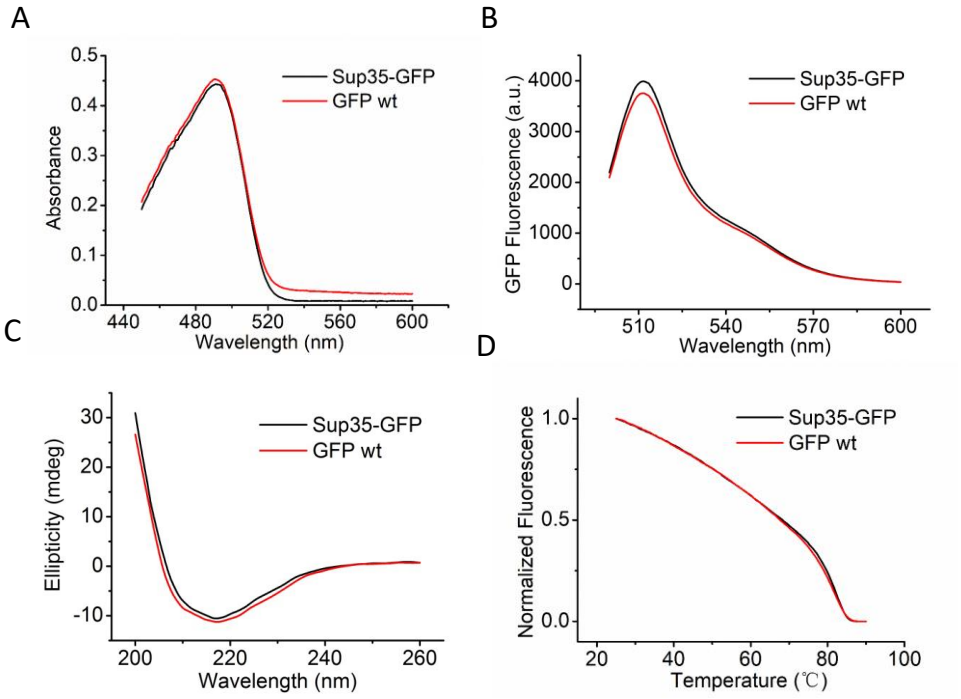

Figure 5. Conformational properties and thermal stabilities of Sup35-GFP. Emission (A) excitation (B) and far-UV CD (C) spectra. (D) Thermal unfolding curves followed by monitoring GFP fluorescence at 515 nm from 25 ºC to 90 ºC with a heating rate of 1 ºC/min. Sup35-GFP and GFP are shown in black and red, respectively.

Next, we monitored the thermal stability of Sup35-GFP and GFP by following the changes in GFP fluorescence emission at 515 nm with the temperature (**Figure 5D**). We obtained cooperative, superimposable, denaturation curves with $T_m = 81.0 \pm 0.2$ ºC and 81.9 ± 0.2 ºC for Sup35-GFP and GFP, respectively. Altogether the data indicate that, as in the case of the FF domain, Sup35-SAC does not alter the fold or the stability of GFP.

Finally, we incubated Sup35-GFP and GFP at pH 7.4 and 37 °C for one week. Because Th-T and CR cannot be used to monitor amyloid structure when GFP is present, we directly characterized the morphology of the potential assemblies using TEM, confirming the presence of long fibrillar assemblies with a diameter of 20.1±1.1 nm for Sup35-GFP (**Figure 6B**) and the absence of any ordered aggregate for GFP (**Figure 6A**), indicating that Sup35-SAC drives specifically the amyloid assembly in the context of the Sup35-GFP fusion.

To assess if GFP maintains a functional conformation when embedded in the fibrils, we recorded the GFP fluorescence emission spectra of the aggregated material. It exhibited, the characteristic maximum at ~511 nm (**Figure 6C**) and, indeed, when this material was imaged in a fluorescence microscope using an FITC filter (excitation at 465-495 nm) the presence of large green fluorescent fibrillar structures was directly observed (**Figure 6D**). A quantitative analysis of the fluorescence emission of Sup35-GFP in the soluble and fibrillar states indicated that the specific fluorescence emission of the fibrils was 2.6 times lower (**Figure 6C**). Whether this decrease in activity responds to a fraction of the domains being inactive in the assembly or it results instead from self-quenching, owing to the expected dense packing of the domains and close proximity of the fluorophores in the fibrils, should be further examined.

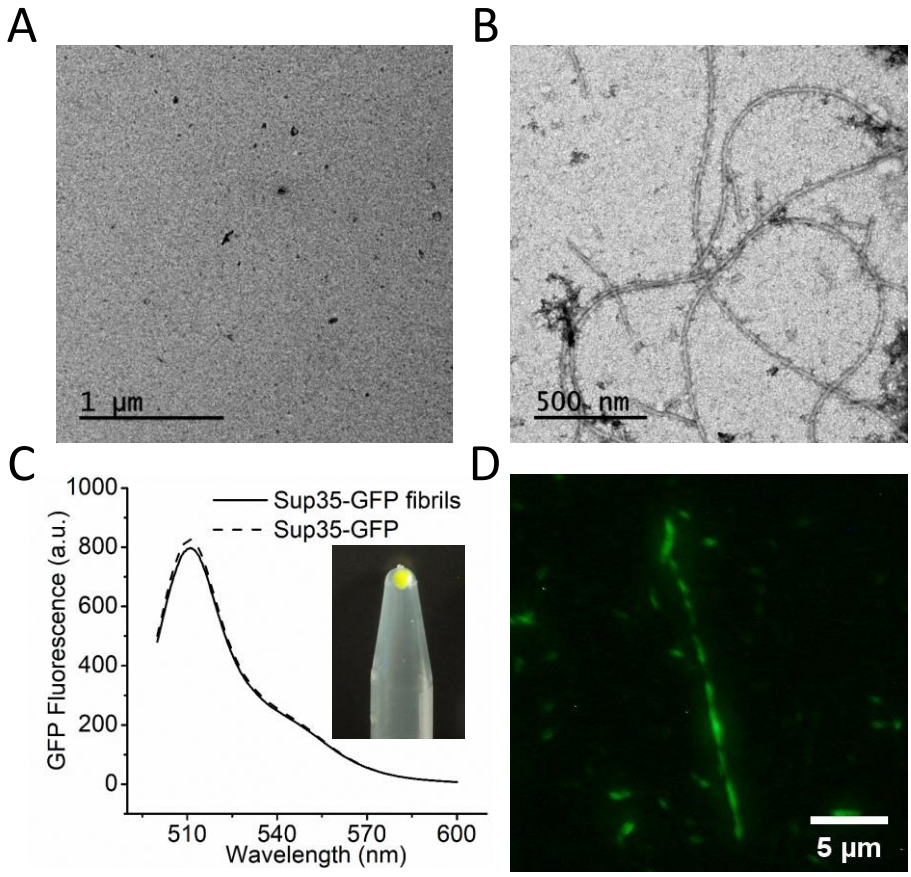

Figure 6. Morphology and functionality of Sup35-GFP fibrils. Sup35-GFP and GFP solutions were incubated for one week. (A) and (B) show representative TEM micrographs of GFP and Sup35-GFP, respectively. Scale bar represents 1 µm and 500 nm. (C) Emission fluorescence spectra of Sup35-GFP fibrils at 5 µM and photograph of precipitated fibrils. The fluorescence emission of soluble Sup35-GFP at 2 µM is shown for comparison. (D) Fluorescence microscopy image of Sup35-GFP fibrils. Scale bar represents 5 µm.

Sup35-SAC drives the formation of catalytically active amyloid fibrils

To further confirm the idea that Sup35-SAC can be exploited to obtain modular and functional nanofibrillar assemblies, this short sequence stretch was fused to the carbonic anhydrase enzyme (Sup35-CA) (**Figure S4B**).

Again, Sup35-CA was purified from the soluble cell fraction (**Figure S6**), with a yield of 98 mg/L culture. As in the case of the FF and GFP proteins, the comparison of the conformational properties of Sup35-CA with those of wild type CA (CA-wt) indicated that the fusion of the 21-residues peptide does not alter significantly the adjacent globular structure (**Figure S7A and S7B**), neither its stability, since CA-wt and Sup35-CA exhibit similar cooperative denaturation curves with $[Urea]_{50\%}$ of 4.24 ± 0.03 M and 4.40 ± 0.04 M, respectively, as monitored by intrinsic fluorescence (**Figure S7C**).

To analyse the self-assembling capacities of Sup35-CA, we incubated the protein at pH 7.4 and 25 °C for one week and analysed the presence of amyloid-like material using Th-T (**Figure 7A**) and CR assays (**Figure 7B**). In the presence of incubated Sup35-CA, Th-T exhibits a high increase in fluorescence emission; likewise, the protein promotes a red shift of the CR spectra, both proves indicating an amyloid-like nature. The morphology of the Sup35-CA assemblies was analysed by TEM (**Figure 7D**), which allowed to observe the presence of long amyloid fibrillar structures of 23.0±1.5 nm in diameter. No aggregate was detected in a fresh Sup35-CA protein solution (**Figure 7C**), consistent with its inability to bind Th-T and CR (**Figure 7A and 7B**).

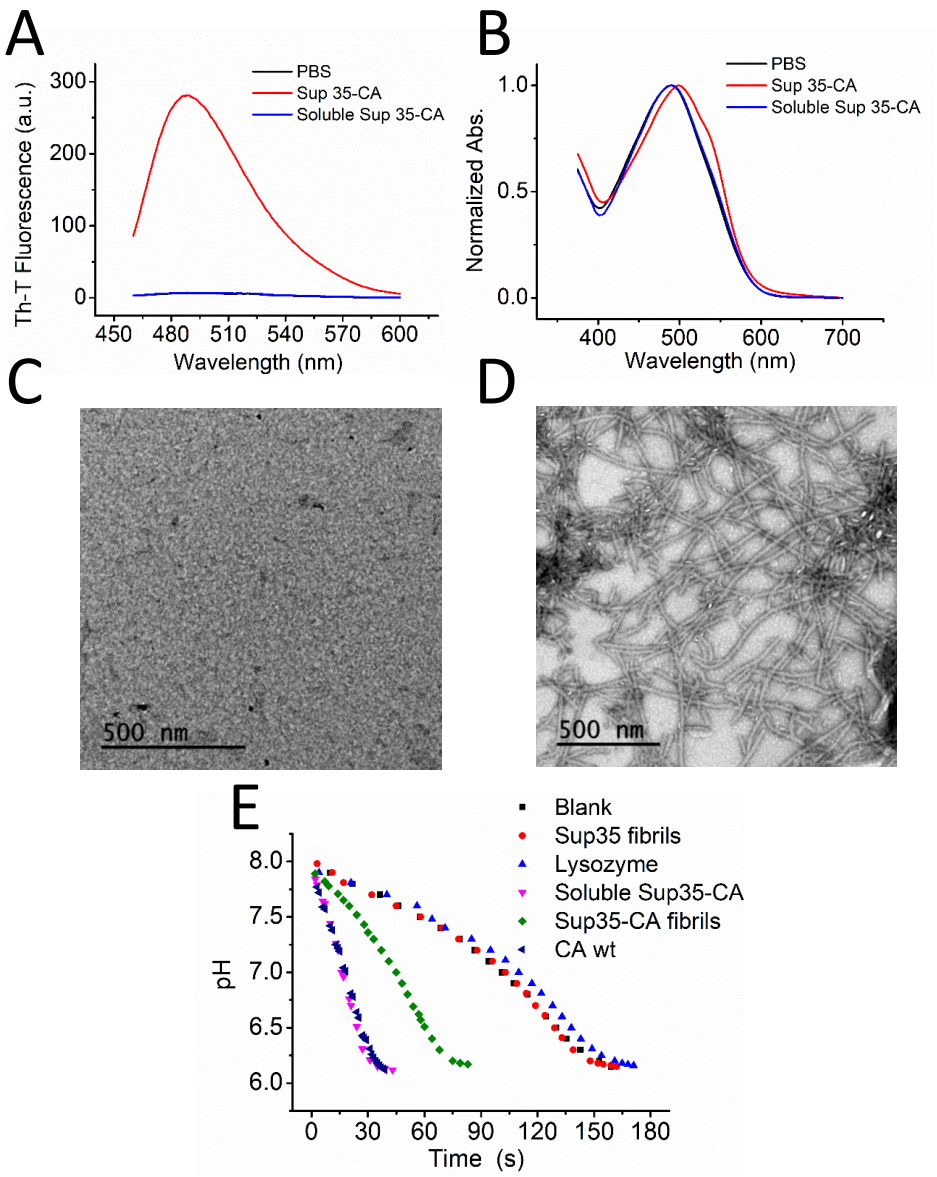

Figure 7. Tinctorial properties, morphology and activity of Sup35-CA fibrils. Sup35-CA was incubated for one week. (A) Th-T and (B) CR binding assays in the absence (black) or presence (red) of incubated Sup35-CA protein. Soluble Sup35-CA was used as a control (blue). Panels (C) and (D) show representative TEM micrographs of soluble and incubated Sup35-CA protein, respectively. Scale bar represents 500 nm. (E) CO_2 hydration activity of Sup35-CA fibrils. Buffer alone (black), Sup35-SAC fibrils (red) and lysozyme (blue) were used as negative controls while CA-wt (dark blue) was used as positive control. pH measurements showed that carbonic anhydrase moieties embedded in the Sup35-CA fibrils (green) or in the soluble fusion (pink) were active.

Finally, we characterized the catalytic activity of Sup35-CA fibrils by modified Wilbur-Anderson method[55] (**Figure 7E**). In this method, the activity of CA is measured by monitoring the solution pH acidification caused by the CA catalysed conversion of carbon dioxide and water into bicarbonate and hydrogen ions, resulting in a decrease of pH. The kinetics of the reaction in the pH range 8.1- 6.3 were recorded. Before measuring the activity of fibrils, soluble protein was removed by repeated centrifugation and washing steps. The activity of the protein solutions was calculated using the equation: WAU= $(t_0-t)/t$ where t is the time required for the pH change when the protein is present and t_0 is the time required in buffer alone. Bromothymol blue (BTB) was used as a pH indicator, to confirm the solution acidification. When the reaction was performed in the presence of Sup35-CA fibres, an activity of 1.2 WAU was recorded, and the solution colour in the presence of BTB changed from blue to yellow, confirming that at least a fraction of the CA domains in the fibrils was functional. No activity was detected when lysozyme or fibrils formed by the Sup35-SAC peptide alone, therefore devoid of CA, where used; indicative of the specificity of the reaction. The activity of CA in the fibrils was 3.4 times lower that the one exhibited by an equivalent amount of either soluble Sup35-CA (4.1 WAU) or CA-wt (4.0 WAU). This reduction of CA activity in the fibrillar state has been already observed for a fusion between the complete PrD of the yeast Ure2p and CA,[20] despite in that case the reduction of activity in the filaments relative to the soluble form was higher, about 10-fold. Two explanations may account for these differences in activity: *i*) when folded, CA displays the same activity in both fibrils, but the amount of properly folded domains is higher in Sup35-CA fibrils. *ii*) CA is more active in Sup35-CA fibrils.

CA is a near-diffusion-limited enzyme.[56] This implies that the reaction rate depends on the speed at which substrate diffuses to the active size. Aggregation might decrease the apparent activity of CA because the fibrillar structure reduces the diffusion rate of the substrate. In this case, it is expected that the fibrils formed by smaller self-assembling domains, such as the Sup35-SAC, would impose lower restrictions to

diffusion than those of the larger PrDs. The same rational would apply in case the reduction in activity is caused by steric hindrance. Indeed, an inverse relationship between the size of the fibril-forming component and the activity of the adjacent enzyme in the aggregated state was reported for a fusion of the Ure2p PrD with horseradish peroxidase (HRP).[19] The use of the 93-residues full-length PrD rendered very low HRP activity, whereas shortening it to 80 residues permitted to double the catalytic activity of the enzyme in the fibrillar form.

In any case, it seems clear from our experiments that the short soft amyloid core of Sup35 PrD can be employed to generate catalytically active fibrils.

Modelling of amyloid fibrils containing Sup35-SAC linked to a globular domain.

The distance between β-strands in a typical amyloid β-sheet is ≈ 4.8 Å; whereas the size of the globular domains is typically ranged between 30-40 Å. Therefore, when located in close vicinity to amyloid regions along the chain, globular domains might prevent the formation of ordered amyloids because of the steric repulsion. This is especially true for the parallel arrangement because the equivalent c-terminal β-sheet positions to which globular domains are appended are separated only by 4.8 Å, while in the antiparallel arrangement this distance is twice longer (9.6 Å). To ensure our conclusion that a Sup35-SAC amyloid core surrounded by globular domains can be formed without significant steric tension, we applied a molecular modelling approach as previously described.[57] In this work, as a model, we used a hybrid protein containing the amyloid-forming Sup35-SAC peptide and GFP linked by 8-residues linker, as in our Sup35-GFP construct. A number of experimental data suggest that the prion domain of Sup35 and its fragments form amyloid fibrils with the parallel and in-register cross-β structure[58], which is in agreement with the ATR-FTIR spectra of Sup35-FF fibrils. Furthermore, it is known that the amyloid protofibrils usually have a slight left-handed twist. Therefore, we constructed these amyloid fibrils with an axial twist angle of 2°

per 4.8-Å step (**Figure 8**). For the parallel and in-register cross-β amyloid core we used two structural arrangements: stacks of linear β-strands and of β-arches. The 21-residues Sup35-SAC is long enough to form β–strand-β–arc-β-strand elements called 'β-arches' that have been found in a significant number of disease-related and functional amyloid fibrils. The possibility of a β-arch arrangement for Sup35-SAC was correctly predicted by the ArchCandy algorithm.[59] Our molecular modelling confirmed that the hybrid molecules experimentally tested in this work can form parallel and in-register cross-β amyloids (with Sup35-SAC either in the linear or β-arch conformations) decorated by globular domains (**Figure 8**). The structural models were energy-minimized using steepest descent algorithm of GROMACS package version 4.6.7 [60] and their steric tension and overall stereochemistry were evaluated by using PROCHECK package.[61] In the final structures (**Figure 8**), the GFP domains do not have steric clashes. The conclusion of the molecular modelling can be extended to the antiparallel cross-β amyloids of the hybrid molecules because, as detailed above) these fibrils would have even more space to accommodate the globular domains. The diameter of the modelled fibrils are 17.7 nm and 17 nm for the linear and β-arch conformations, respectively; in very good agreement with the 20 nm we measured for Sup35-GFP amyloid fibrils.

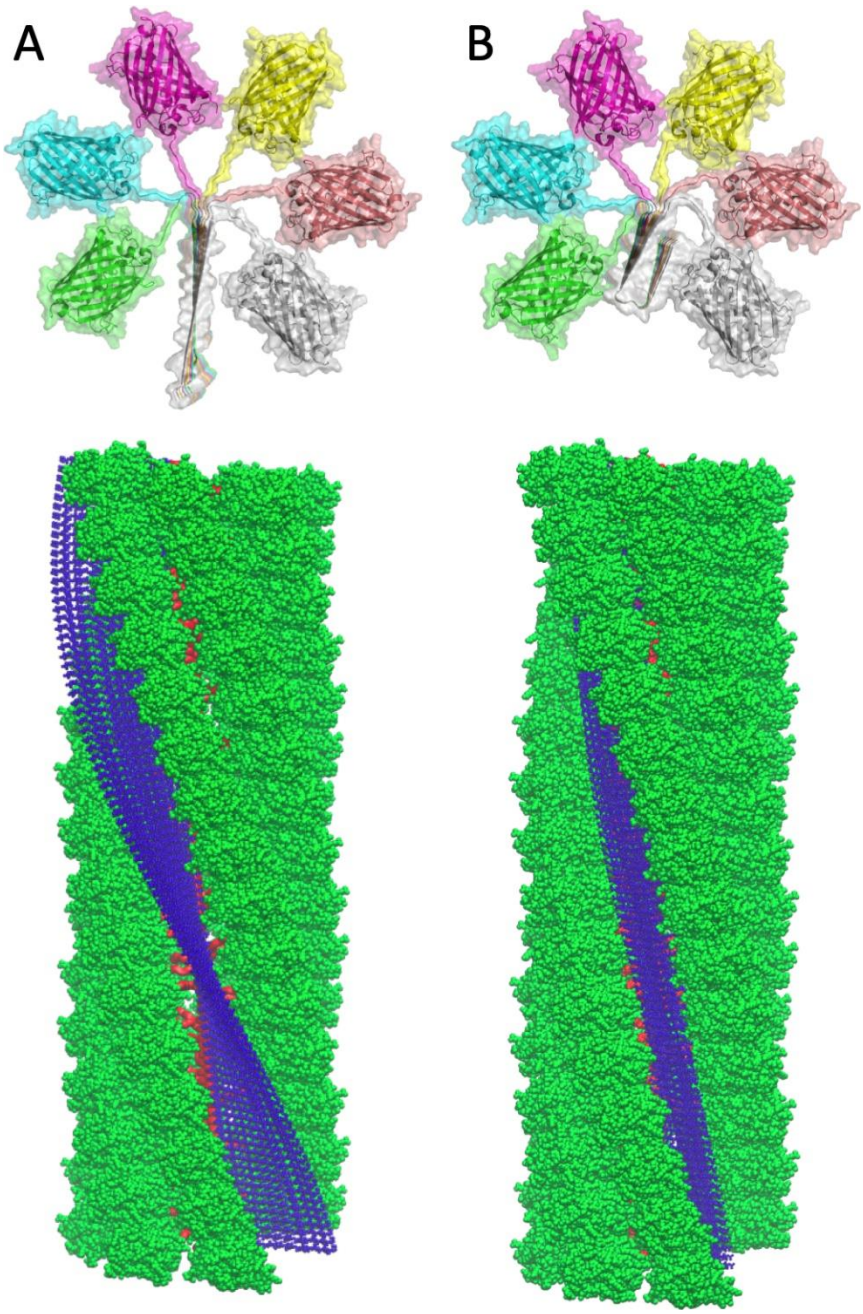

Figure 8. Structural models of amyloid fibrils formed by Sup35-GFP (A) Fibrils formed by stacking of linear β-strands of Sup35-SAC. Axial projection of the repetitive element (top). Lateral projection of the 90-mer fibril that corresponds to a half helix turn (180°). The fibril was generated by the 28.8 Å translation of the hexameric element along the fibril axis (bottom). B. Fibrils formed by stacking of β-arches of Sup35-SAC. Axial projection (top) and lateral projections (bottom). Images were generated by using PyMol software.[62]

Conclusions

We have developed different functional nanofibrils exploiting our previous discovery that the self-assembling properties of PrDs can be mimicked by their much shorter SACs. In particular, we have demonstrated that the conjugation of globular proteins to Sup35-SAC induces a controlled self-assembly into amyloid-like supramolecular structures. In contrast to other fusions between β-sheet forming peptides and globular domains, our proteins could be obtained at high yield in a soluble state, without any need for refolding procedures. This is likely because Sup35-SAC remains largely disordered when soluble, despite its ability to adopt a β-sheet conformation in the fibrillar state. Moreover, in Sup35-SAC fusions the adjacent globular proteins display conformational, stability and folding properties almost indistinguishable of those of the original proteins. Importantly, the globular domains within our fusion proteins seem to maintain their structure preserved in the fibrillar state. Indeed, the diameter of the individual fibrils correlates well with that of the conjugated globular domain: 16, 20 and 23 nm for the FF domain (8 kDa), GFP (27 kDa) and CA (30 kDa) fibrils, respectively. This trend suggests that the packing of FF and CA domains in the fibrils might be similar to the one observed in the Sup35-GFP fibrils derived models. Therefore, Sup35-SAC appears as a module that can be readily used to immobilize bioactive proteins of different sizes and structures. The results agree well with a previous molecular modelling study that allowed us to establish a more general relationship between the size of the globular domains and the length of the linkers in the parallel and in-register cross-β fibrils.[57]

The modular genetic fusion approach described here can be applied to decorate fibrils with different functionalities, including active enzymes. Altogether, in addition to validate the prion-like proteins of short SACs, the present work illustrates a straightforward strategy to obtain novel bionanomaterials displaying immobilized functional proteins of biological or chemical interest.

Materials and methods

Reagents and enzymes. Reagents and enzymes were purchased from Sigma-Aldrich (UK), unless otherwise stated. Carbon grid (400 square mesh copper) were purchased from Micro to Nano (Netherlands) and the uranyl acetate solution were provided by the microscopy service (Universitat Autònoma de Barcelona). Sup35-SAC 21-residues peptides were purchased from CASLO ApS (Scion Denmark Technical University).

Expression and Purification of Proteins. The FF domain, corresponding to residues 212-266 of yeast URN1, was cloned into a pET-28(a) plasmid (Addgene, USA), that previously contained the soft amyloid core Sup35 sequence, resulting in a plasmid encoding for a chimeric protein (Sup35-linker-FF domain) with an His6 tag. The cDNAs of Sup35-GFP (folding reporter green fluorescent protein[63]) and Sup35-CA (Carbonic anhydrase) cloned in the plasmid pET28(a) with a His6 tag were acquired from GenScript (USA). *E.coli* BL21 (DE3) competent cells were transformed with the correspondent plasmids. Then, transformed cells were grown in 10 mL LB medium containing 50 μg/mL kanamycin, overnight at 37 °C, and transferred into 1 L fresh LB media containing 50 μg/mL kanamycin. After reaching an OD_{600} of 0.6, the culture was induced with 0.4 mM IPTG and grown either at 20 °C for 16 h or at 37 °C for 4h. At both temperatures, the proteins were found mainly in the soluble cell fraction; however, 20 °C was selected for all preparative productions, since the soluble/insoluble ratio was slightly higher. Cells were collected by centrifugation at 5000 rpm for 15 min at 4 °C. The collected pellet was resuspended into 20 mL PBS pH 7.4 containing 20 mM imidazole, 1 mg/mL lysozyme and 1 mM PMSF. The solution was incubated on ice, followed by sonication for 20 min. The supernatant was collected by centrifugation at 15000 rpm for 30 min at 4 °C and, purified in an His-tag column, according to the manufacturer's protocol, followed by a gel filtration onto a HiLoad™ Superdex™ 75 prepgrade column (GE Healthcare,USA). The purified proteins were frozen with liquid nitrogen and stored at -80 °C. The purity of the sample was confirmed by SDS-PAGE.

The concentration of the proteins: Sup35-FF, Sup35-GFP and Sup35-CA was determined by UV absorption using a ε value of 18450 $L \cdot mol^{-1} \cdot cm^{-1}$, 24995 $L \cdot mol^{-1} \cdot cm^{-1}$ and 54890 $L \cdot mol^{-1} \cdot cm^{-1}$, respectively. The protein was found mainly in the soluble cell fraction, independently if the expression was done at 37 °C or at 20 °C; however, this last temperature was used for all preparative protein productions since the soluble/insoluble ratio was slightly higher.

Prediction of Aggregation and Disorder. The protein sequence of FF domain, Sup35-FF, Sup35-SAC, and Sup35-linker were used to predict aggregation and disorder. The aggregation propensity was predicted by three different methods: Aggrescan[41], FoldAmyloid[42] and PASTA 2.0 [43]. The disordered regions were predicted using four different algorithms: PONDR (VSL2)[45], GlobPlot[46], PASTA 2.0[43] and IUPred[47]. The parameters used for the predictions were the following: Aggrescan and PONDR (VSL2) were run with default parameters. For FoldAmyloid, the scale option was “expected number of contacts 8Å” the "averaging frame" was 5 and threshold was 21.4. For PASTA 2.0, the “custom” mode was used and the “top pairing energies” was 20, the “energy threshold” was -5 (1.0 Pasta Energy Unit corresponding to 1.192 Kcal/mol), and the TPR (sensitivity) and FPR (1- specificity) were 40.5% and 4.7%, respectively. For GlobPlot, the mode of propensities was “Russell/Linding (recommended P=RC-SS)”, for disorder prediction, minimum peak width was 5 and maximum join distance was 4; the globular domain hunting was performed using domain prediction “SMART/Pfam” parameter with minimum peak width 74 and maximum join distance 15, and the plot was smoothed using 1st and 2nd derivative Savitzky-Golay with frame 10. For IUPred, the “short disorder of prediction type” was used and the output type was “generate plot” with window size 500.

Conformational Characterization. Proteins were dissolved at a final concentration of 20 μM for Sup35-FF, Sup35-GFP and Sup35-CA in PBS pH 7.4 buffer, then samples were filtered through a 0.22 μm Millipore filter and immediately analysed. Far-UV CD

spectra were recorded from 260 nm to 200 nm at 1 nm bandwidth, response time of 1 second, and a scan speed of 100 nm/min in a Jasco-710 spectropolarimeter (Jasco Corporation, Japan), thermostated at 25 °C. Ten accumulations were averaged for each spectrum. Trp intrinsic fluorescence and GFP intrinsic fluorescence spectra were measured at 25 °C on a Jasco FP-8200 Spectrofluorometer (Jasco Corporation, Japan). Three averaged spectra were accumulated using an excitation wavelength of 280 nm and 485 nm and recording emission from 300 to 400 nm and 500 to 600 nm for Trp and GFP emission fluorescence, respectively, with slit widths of 5 nm. Three averaged absorbance spectra of Sup35-GFP were recorded from 450 to 600 nm on a SPECORD 200 plus spectrophotometer (Analytik Jena, Germany), with scan speed of 20 nm/s. As controls, 20 μM FF-wt, GFP and CA-wt in PBS buffer, pH 7.4, were measured in all the assays under the same conditions. The fluorescence emission spectra of Sup35-GFP fibrils was recorded as described for the soluble fusion at 5 μM final concentration. The amount of protein in the fibrillar solution was titrated after denaturation by SDS-PAGE against a concentration standard of soluble Sup35-GFP.

Thermal and chemical denaturation. FF-wt and Sup35-FF were dissolved at 20 μM in PBS, pH 7.4. Trp intrinsic fluorescence was monitored in a Jasco FP-8200 Spectrofluorometer from 25 °C to 90 °C, with an excitation wavelength of 280 nm and recording emission at 340 nm, with an interval of 0.5 °C and a heating rate of 0.5 °C/min. Ellipticity was recorded at 222 nm each 1 °C with a heating rate 0.5 °C/min from 25 °C to 90 °C, using a Jasco-710 spectropolarimeter. In the case of Sup35-GFP and GFP, soluble proteins were prepared at 10 μM in PBS, pH 7.4. GFP intrinsic fluorescence was recorded from 25 °C to 90 °C, with excitation wavelength of 485 nm and emission recorded at 515 nm each 0.5 °C with a heating rate of 0.5 °C/min.

For chemical denaturation, samples were prepared at 10 μM in PBS, pH 7.4 in the presence of different concentrations of denaturing agent (0-9 M of urea), and the reaction was left to equilibrate for 20 h at room temperature. Each sample was prepared

in duplicate. Trp intrinsic fluorescence spectra and far-UV CD spectra were collected at 280/360 nm (excitation/emission) and 222 nm, respectively. One-week aggregated Sup35-FF was analysed in the same manner at 5 μM final concentration. For Sup35-CA and CA-wt the shift in the center of mass (CM) of the Trp emission spectrum was calculated and plotted against the urea concentration.

The results 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 by Origin 8.5 (OriginLab Corporation).

Folding Kinetics. Kinetics of unfolding and refolding were analysed by recording intrinsic Trp fluorescence changes upon transition at 25 °C in a Bio-Logic SFM-3 stopped-flow instrument (Bio-Logic Science Instruments), employing an excitation wavelength at 280 nm and a 320 nm cut-off filter. For unfolding reactions, the protein solution was prepared in PBS pH 7.4, at 10 μM and appropriate volumes of the same buffer containing concentrated urea were added to initiate the reaction. For the refolding reaction, selected volumes of denaturant free buffer were added to the initial protein solution at 10 μM in 9M urea. All fluorescence traces were fitted to an exponential function, and kinetic parameters were calculated considering a two-state folding model, using Origin 8.5 program (OriginLab Corporation). Kinetic and free energy values were determined using the following equations:

$$\Delta G_{U-F}^{H_2O} = -RT \cdot \ln\left(k_U^{H_2O} / k_F^{H_2O}\right)$$

$$m_{U-F} = RT\left(m_F + m_U\right)$$

where k_F and k_U are the rate constants for folding and unfolding, respectively, and the m_F and m_U values correspond to the slopes of the respective folding and unfolding regions.

Fibril Formation. Sup35-FF, Sup35-GFP and Sup35-CA proteins were prepared at 50 μM, 1.5 mM and 1.0 mM in PBS pH 7.4, and filtered through a 0.22 μm filter. The Sup35-FF and Sup35-GFP samples were incubated at 37 °C, and Sup35-CA at 25 °C, with agitation at 300 rpm for one week. FF-wt and GFP were incubated at the same concentrations and conditions as controls.

Amyloid dyes binding. Thioflavin T (Th-T) and Congo red (CR) was used to determine the formation of amyloid fibrils. For the binding assay, incubated proteins were diluted to a final concentration of 20 μM in PBS pH7.4, in the presence of 25 μM Th-T. Emission fluorescence was recorded in the 460-600 nm range, using an excitation wavelength of 445 nm and emission bandwidth of 5 nm on a Jasco FP-8200 Spectrofluorometer (Jasco Corporation, Japan). For the CR binding assay, incubated proteins were prepared at final concentration of 20 μM and, CR was mixed to a final concentration of 20 μM. Optical absorption spectra were recorded in the range from 375 to 700 nm in a Specord 200 Plus spectrophotometer (Analytik Jena, Germany). Spectra of protein alone and buffer were acquired to subtract protein scattering.

Transmission Electron Microscopy (TEM). For TEM samples preparation, 10μL of the incubated proteins were deposited on a carbon-coated copper grid for 10 min and the excess liquid was removed with filter paper, followed by a negative stain with 10 μL of 2 %(w/v) uranyl acetate for 1min. Grids were exhaustively scanned using a JEM 1400 transmission electron microscope (JEOL ltd, Japan) operating at 80 kV, and images were acquired with a CCD GATAN ES1000W Erlangshen camera (Gatan Inc., USA). The width of fibrils was analysed by Image J (National Health Institute), averaging the measures of 10 individual fibrils for each fusion protein.

Fourier Transform Infrared Spectroscopy (FTIR). FF-wt was dissolved at 180 μM in PBS, pH 7.4, and filtered through a 0.22 μm filter. 30 μL of the prepared Sup35-FF fibrils at 50 μM were centrifuged at 12000g for 30 min and resuspended in 10 μL of

water. Samples were placed on the ATR crystal and dried out under N_2 flow. The experiments were carried out in a Bruker Tensor 27 FTIR (Bruker Optics, USA) supplied with a Specac Golden Gate MKII ATR accessory. Each spectrum consists of 32 acquisitions measured at a resolution of 1 cm^{-1} using the three-term Blackman-Harris Window apodization function. Data were acquired and normalized, using the OPUS MIR Tensor 27 software (Bruker Optics, USA), with the Min/Max normalization method, which scales spectrum intensities to the effect that the minimum absorbance unit will be 0 and the maximum 2. The analysis of the IR spectra was performed with the PeaKFit program (SeaSolve Software Inc.). Obtained data, in the amide I region (1700 to 1600 cm^{-1}), were first pre-smoothed using non-parametric smoother Loess ((locally weighted smoothing) procedure at 5% level, and the lineal baseline was subtracted. IR spectra were fitted employing the residuals method for finding hidden peaks, which consists on finding the difference in y-value between a data point and the sum of component peaks evaluated at the data point's x-value. From the second derivative plot for each sample absorbance spectrum, peaks and local minima were identified and the number and positions of them manually placed to deconvolute the absorbance spectra. Afterwards, automated peak fitting was done using the "AutoFit Peaks I Residuals" option with the "vary widths" condition for the autoscan procedure, until reaching iteration 7 and a r^2>0.997. The resulting area, amplitude and center values of the fitted bands were exported as a table and plotted.

Aggregation and Seeding Kinetics. Sup35-FF solutions were prepared at 50 μM in PBS pH 7.4 containing 20 μM Th-T and transferred into 96 well plates in the absence and presence of 2% (v/v) Sup35-FF amyloid seeds. Each sample was prepared by triplicate. The Th-T fluorescence at 485 nm was recorded every 3 min at 37 °C during 1000 min, using a Victor III Multilabel Plate Reader (Perkin Elmer,USA), equipped with P450 CW-lamp filter and 486/10nm emission filter. The obtained data were analyzed and plotted by following a Boltzmann function in Origin 8.5 (OriginLab Corporation).

Fluorescence Microscopy. 50 μL aggregated Sup35-GFP was centrifuged at 12000g for 30 min. Supernatant was removed and fibrils were resuspended in PBS pH 7.4. 5 μL of the resuspended fibres were mixed with 5 μL mounting medium solution and dropped on a clean glass slide (Deltalab, 26×76 mm) and covered by a cover slide (Deltalab, 22×22mm). Fluorescence imaging of nanofibers was carried out on an Eclipse 90i epifluorescence optical microscopy equipped with a Nikon DXM1200F (Nikon,Japan) camera and ACT-1 software. Images were acquired with an excitation filter of 465-495 nm and detecting fluorescence emission in a range of 515-555 nm.

Catalytic activity of Sup35-CA fibrils. Catalytic activity of Carbonic Anhydrase (CA) was assayed by using a modified Wilbur-Anderson method.[55] Specifically, 2 mL of 50 mM Tris-HCl,100 mM NaCl, pH 8.1 buffer were mixed with 200 μL 0.05% BTB (Bromothymol blue) on ice. 10 μL of 20 μM protein samples were added and the reaction was initiated by addition of 1.0 mL of ice-cold CO_2 saturated water. To determine the catalytic activity, the time required for the pH to drop from 8.1 to 6.3 was recorded. BTB was used as a pH indicator. The activity of the tested sample was calculated using the equation: WAU= $(t_0-t)/t$ where t is the time required for the pH change when the test sample is present and t_0 is the time required for the pH change when the buffer is substituted for the test sample. CA wt was used as positive control, to evaluate the effect of Sup35 fusion on the CA activity. Sup35-SAC fibrils and lysozyme were used as negative controls to show the specificity of CA activity in the fibrils.

Molecular modelling of fibril with a pseudo-helical packing of GFPs. The left-handed twisted fibrils were built with 2° angle twist per axial 4.8 Å translation between the monomers. The twisting operator was imposed around a Z-axis located at the center of mass of the main-chain atoms from the two β-strands of the Sup35 fragment. To build models of hybrid protein we used the crystal structure of GFP protein (pdb entry

1GFL).[64] To position the GFP structures around the amyloid core and evaluate their packing, we built a model of two successive hexameric layers of the subunits.[57] The first GFP subunit was manually docked on one side of the fibril stack while keeping the axis of the GFP β-barrel perpendicular to the Z-axis. The five other subunits were positioned by applying the successive rotation of ~51° around an axis that intersects the X-Y plane at the Cα atom of the last Sup35 residue of the first monomer and their successive translations of 4.8 Å along the Z-axis. This geometry was chosen so that it provides the most optimal close packing of the GFP molecules within the hexamer (**Figure 8**). To evaluate the packing of GFP domains between the hexameric units, an adjacent hexameric layer was generated by a 28.8 Å translation along the Z-axis. The 28.8 Å distance agrees well with the corresponding distance between the GFPs from *Aequorea victoria* in the crystal packing (pdb entry 1GFL).[64] CHIMERA[65] program was used for the symmetrical positioning of GFPs and molecular modelling procedures. The linkers connecting Sup35 peptides to GFPs were generated using MODLOOP.[66] The quality and consistency of the generated models were checked with the PROCHECK program.[61]

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

We thank Drs. Anita Carija and Ricardo Sant'Anna for experimental help. This work was funded by the Spanish Ministry of Economy and Competitiveness BIO2016-78310-R to S.V and by ICREA, ICREA-Academia 2015 to S.V. Weiqiang Wang acknowledges financial support from the China Scholarship Council (CSC): NO. 201606500007.

References

1. Chiti, F.; Dobson, C. M., Protein misfolding, amyloid formation, and human disease: a summary of progress over the last decade. *Annual review of biochemistry* **2017,** *86*, 27-68.

2. Si, K., Prions: what are they good for? *Annual review of cell and developmental biology* **2015,** *31*, 149-169.

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

4. Sabaté, R.; Ventura, S., Cross-β-sheet supersecondary structure in amyloid folds: techniques for detection and characterization. In *Protein Supersecondary Structures*, Springer: 2012; pp 237-257.

5. Diaz Caballero, M.; Navarro, S.; Fuentes, I.; Teixidor, F.; Ventura, S., Minimalist Prion-Inspired Polar Self-Assembling Peptides. *ACS nano* **2018**.

6. Li, C.; Bolisetty, S.; Mezzenga, R., Hybrid nanocomposites of gold single-crystal platelets and amyloid fibrils with tunable fluorescence, conductivity, and sensing properties. *Advanced Materials* **2013,** *25* (27), 3694-3700.

7. Scheibel, T.; Parthasarathy, R.; Sawicki, G.; Lin, X.-M.; Jaeger, H.; Lindquist, S. L., Conducting nanowires built by controlled self-assembly of amyloid fibers and selective metal deposition. *Proceedings of the National Academy of Sciences* **2003,** *100* (8), 4527-4532.

8. Holmes, T. C.; de Lacalle, S.; Su, X.; Liu, G.; Rich, A.; Zhang, S., Extensive neurite outgrowth and active synapse formation on self-assembling peptide scaffolds. *Proceedings of the National Academy of Sciences* **2000,** *97* (12), 6728-6733.

9. Bolisetty, S.; Boddupalli, C. S.; Handschin, S.; Chaitanya, K.; Adamcik, J.; Saito, Y.; Manz, M. G.; Mezzenga, R., Amyloid fibrils enhance transport of metal nanoparticles in living cells and induced cytotoxicity. *Biomacromolecules* **2014,** *15* (7), 2793-2799.

10. Rufo, C. M.; Moroz, Y. S.; Moroz, O. V.; Stöhr, J.; Smith, T. A.; Hu, X.; DeGrado, W. F.; Korendovych, I. V., Short peptides self-assemble to produce catalytic amyloids. *Nature chemistry* **2014,** *6* (4), 303.

11. Al-Garawi, Z.; McIntosh, B.; Neill-Hall, D.; Hatimy, A.; Sweet, S.; Bagley, M.; Serpell, L., The amyloid architecture provides a scaffold for enzyme-like catalysts. *Nanoscale* **2017,** *9* (30), 10773-10783.

12. Marinelli, P.; Navarro, S.; Baño-Polo, M.; Morel, B.; Graña-Montes, R.; Sabe, A.; Canals, F.; Fernandez, M. R.; Conejero-Lara, F.; Ventura, S., Global Protein Stabilization Does Not Suffice to Prevent Amyloid Fibril Formation. *ACS chemical biology* **2018,** *13* (8), 2094-2105.

13. Wakabayashi, R.; Suehiro, A.; Goto, M.; Kamiya, N., Designer aromatic peptide amphiphiles for self-assembly and enzymatic display of proteins with morphology control. *Chemical Communications* **2019,** *55* (5), 640-643.

14. Chien, P.; Weissman, J. S.; DePace, A. H., Emerging principles of conformation-based prion inheritance. *Annual review of biochemistry* **2004,** *73* (1), 617-656.

15. Uptain, S. M.; Lindquist, S., Prions as protein-based genetic elements. *Annual Reviews in Microbiology* **2002,** *56* (1), 703-741.

16. Ross, E. D.; Minton, A.; Wickner, R. B., Prion domains: sequences, structures and interactions. *Nature cell biology* **2005,** *7* (11), 1039.

17. Hafner-Bratkovič, I.; Bester, R.; Pristovšek, P.; Gaedtke, L.; Veranič, P.; Gašperšič, J.; Manček-Keber, M.; Avbelj, M.; Polymenidou, M.; Julius, C., Globular domain of the prion protein needs to be unlocked by domain swapping to support prion protein conversion. *Journal of Biological Chemistry* **2011,** *286* (14), 12149-12156.

18. Baral, P. K.; Swayampakula, M.; Aguzzi, A.; James, M. N., X-ray structural and molecular dynamical studies of the globular domains of cow, deer, elk and Syrian hamster prion proteins. *Journal of structural biology* **2015,** *192* (1), 37-47.

19. Zhou, X. M.; Entwistle, A.; Zhang, H.; Jackson, A. P.; Mason, T. O.; Shimanovich, U.; Knowles, T. P.; Smith, A. T.; Sawyer, E. B.; Perrett, S., Self-assembly of amyloid fibrils that display active enzymes. *ChemCatChem* **2014,** *6* (7), 1961-1968.

20. Baxa, U.; Speransky, V.; Steven, A. C.; Wickner, R. B., Mechanism of inactivation on prion conversion of the Saccharomyces cerevisiae Ure2 protein. *Proceedings of the National Academy of Sciences* **2002,** *99* (8), 5253-5260.

21. Men, D.; Guo, Y.-C.; Zhang, Z.-P.; Wei, H.-p.; Zhou, Y.-F.; Cui, Z.-Q.; Liang, X.-S.; Li, K.; Leng, Y.; You, X.-Y., Seeding-induced self-assembling protein nanowires dramatically increase the sensitivity of immunoassays. *Nano letters* **2009,** *9* (6), 2246-2250.

22. Toombs, J. A.; Petri, M.; Paul, K. R.; Kan, G. Y.; Ben-Hur, A.; Ross, E. D., De novo design of synthetic prion domains. *Proc. Natl. Acad. Sci. U. S. A.* **2012,** *109* (17), 6519-24.

23. Ross, E. D.; Edskes, H. K.; Terry, M. J.; Wickner, R. B., Primary sequence independence for prion formation. *Proc. Natl. Acad. Sci. U. S. A.* **2005,** *102* (36), 12825-30.

24. Wei, G.; Su, Z.; Reynolds, N. P.; Arosio, P.; Hamley, I. W.; Gazit, E.; Mezzenga, R., Self-assembling peptide and protein amyloids: from structure to tailored function in nanotechnology. *Chem. Soc. Rev.* **2017,** *46* (15), 4661-4708.

25. Alberti, S.; Halfmann, R.; King, O.; Kapila, A.; Lindquist, S., A systematic survey identifies prions and illuminates sequence features of prionogenic proteins. *Cell* **2009,** *137* (1), 146-58.

26. Sabate, R.; Rousseau, F.; Schymkowitz, J.; Batlle, C.; Ventura, S., Amyloids or prions? That is the question. *Prion* **2015,** *9* (3), 200-206.

27. Sant'Anna, R.; Fernández, M. R.; Batlle, C.; Navarro, S.; De Groot, N. S.; Serpell, L.; Ventura, S., Characterization of amyloid cores in prion domains. *Scientific reports* **2016,** *6*, 34274.

28. Batlle, C.; de Groot, N. S.; Iglesias, V.; Navarro, S.; Ventura, S., Characterization of Soft Amyloid Cores in Human Prion-Like Proteins. *Scientific reports* **2017,** *7* (1), 12134.

29. Zambrano, R.; Conchillo-Sole, O.; Iglesias, V.; Illa, R.; Rousseau, F.; Schymkowitz, J.; Sabate, R.; Daura, X.; Ventura, S., PrionW: a server to identify proteins containing glutamine/asparagine rich prion-like domains and their amyloid cores. *Nucleic acids research* **2015,** *43* (W1), W331-W337.

30. Toombs, J. A.; Petri, M.; Paul, K. R.; Kan, G. Y.; Ben-Hur, A.; Ross, E. D., De novo design of synthetic prion domains. *Proceedings of the National Academy of Sciences* **2012,** *109* (17), 6519-6524.

31. Alberti, S.; Halfmann, R.; King, O.; Kapila, A.; Lindquist, S., A systematic survey identifies prions and illuminates sequence features of prionogenic proteins. *Cell* **2009,** *137* (1), 146-158.

32. Kawai-Noma, S.; Pack, C.-G.; Kojidani, T.; Asakawa, H.; Hiraoka, Y.; Kinjo, M.; Haraguchi, T.; Taguchi, H.; Hirata, A., In vivo evidence for the fibrillar structures of Sup35 prions in yeast cells. *The Journal of cell biology* **2010,** *190* (2), 223-231.

33. Duernberger, Y.; Liu, S.; Riemschoss, K.; Paulsen, L.; Bester, R.; Kuhn, P.-H.; Schölling, M.; Lichtenthaler, S. F.; Vorberg, I., Prion replication in the mammalian cytosol: functional regions within a prion domain driving induction, propagation, and inheritance. *Molecular and cellular biology* **2018,** *38* (15), e00111-18.

34. Bonet, R.; Ramirez-Espain, X.; Macias, M. J., 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: Structure, Function, and Bioinformatics* **2008,** *73* (4), 1001-1009.

35. Allen, M.; Friedler, A.; Schon, O.; Bycroft, M., The structure of an FF domain from human HYPA/FBP11. *Journal of molecular biology* **2002,** *323* (3), 411-416.

36. Castillo, V.; Chiti, F.; Ventura, S., The N-terminal helix controls the transition between the soluble and amyloid states of an FF domain. *PLoS One* **2013,** *8* (3), e58297.

37. Jemth, P.; Day, R.; Gianni, S.; Khan, F.; Allen, M.; Daggett, V.; Fersht, A. R., The structure of the major transition state for folding of an FF domain from experiment and simulation. *Journal of molecular biology* **2005,** *350* (2), 363-378.

38. Kim, W.; Kim, Y.; Min, J.; Kim, D. J.; Chang, Y.-T.; Hecht, M. H., A high-throughput screen for compounds that inhibit aggregation of the Alzheimer's peptide. *ACS chemical biology* **2006,** *1* (7), 461-469.

39. Wang, X.; Zhou, B.; Hu, W.; Zhao, Q.; Lin, Z., Formation of active inclusion bodies induced by hydrophobic self-assembling peptide GFIL8. *Microbial cell factories* **2015,** *14* (1), 88.

40. García-Fruitós, E.; González-Montalbán, N.; Morell, M.; Vera, A.; Ferraz, R. M.; Arís, A.; Ventura, S.; Villaverde, A., Aggregation as bacterial inclusion bodies does not imply inactivation of enzymes and fluorescent proteins. *Microbial cell factories* **2005,** *4* (1), 27.

41. Conchillo-Solé, O.; de Groot, N. S.; Avilés, F. X.; Vendrell, J.; Daura, X.; Ventura, S., AGGRESCAN: a server for the prediction and evaluation of" hot spots" of aggregation in polypeptides. *BMC bioinformatics* **2007,** *8* (1), 65.

42. Garbuzynskiy, S. O.; Lobanov, M. Y.; Galzitskaya, O. V., FoldAmyloid: a method of prediction of amyloidogenic regions from protein sequence. *Bioinformatics* **2009,** *26* (3), 326-332.

43. Trovato, A.; Seno, F.; Tosatto, S. C. E., The PASTA server for protein aggregation prediction. *Protein Engineering, Design and Selection* **2007,** *20* (10), 521-523.

44. Hudalla, G. A.; Sun, T.; Gasiorowski, J. Z.; Han, H.; Tian, Y. F.; Chong, A. S.; Collier, J. H., Gradated assembly of multiple proteins into supramolecular nanomaterials. *Nature materials* **2014,** *13* (8), 829.

45. Peng, K.; Radivojac, P.; Vucetic, S.; Dunker, A. K.; Obradovic, Z., Length-dependent prediction of protein intrinsic disorder. *BMC Bioinformatics* **2006,** *7* (1), 208.

46. Linding, R.; Russell, R. B.; Neduva, V.; Gibson, T. J., GlobPlot: Exploring protein sequences for globularity and disorder. *Nucleic acids research* **2003,** *31* (13), 3701-3708.

47. Dosztányi, Z.; Csizmok, V.; Tompa, P.; Simon, I., IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* **2005,** *21* (16), 3433-3434.

48. Graña-Montes, R.; Marinelli, P.; Reverter, D.; Ventura, S., N-terminal protein tails act as aggregation protective entropic bristles: the SUMO case. *Biomacromolecules* **2014,** *15* (4), 1194-1203.

49. Levine Iii, H.; Scholten, J. D., [29] Screening for pharmacologic inhibitors of amyloid fibril formation. In *Methods in enzymology*, Elsevier: 1999; Vol. 309, pp 467-476.

50. Klunk, W. E.; Pettegrew, J.; Abraham, D. J., Quantitative evaluation of congo red binding to amyloid-like proteins with a beta-pleated sheet conformation. *Journal of Histochemistry & Cytochemistry* **1989,** *37* (8), 1273-1281.

51. Pujols, J.; Peña-Díaz, S.; Conde-Giménez, M.; Pinheiro, F.; Navarro, S.; Sancho, J.; Ventura, S., High-throughput screening methodology to identify alpha-synuclein aggregation inhibitors. *International journal of molecular sciences* **2017,** *18* (3), 478.

52. Xue, W.-F.; Homans, S. W.; Radford, S. E., Systematic analysis of nucleation-dependent polymerization reveals new insights into the mechanism of amyloid self-assembly. *Proceedings of the National Academy of Sciences* **2008,** *105* (26), 8926-8931.

53. Wickner, R. B.; Shewmaker, F. P.; Bateman, D. A.; Edskes, H. K.; Gorkovskiy, A.; Dayani, Y.; Bezsonov, E. E., Yeast prions: structure, biology, and prion-handling systems. *Microbiology and Molecular Biology Reviews* **2015,** *79* (1), 1-17.

54. Goormaghtigh, E.; Cabiaux, V.; RUYSSCHAERT, J. M., Secondary structure and dosage of soluble and membrane proteins by attenuated total reflection Fourier-transform infrared spectroscopy on hydrated films. *European Journal of Biochemistry* **1990,** *193* (2), 409-420.

55. Wilbur, K. M.; Anderson, N. G., Electrometric and colorimetric determination of carbonic anhydrase. *Journal of biological chemistry* **1948,** *176* (1), 147-154.

56. Jönsson, B.; Wennerström, H., Diffusion control in reversible enzyme reactions. Applications to carbonic anhydrase. *Biophysical chemistry* **1978,** *7* (4), 285-292.

57. Azizyan, R. A.; Garro, A.; Radkova, Z.; Anikeenko, A.; Bakulina, A.; Dumas, C.; Kajava, A. V., Establishment of constraints on amyloid formation imposed by steric exclusion of globular domains. *Journal of molecular biology* **2018,** *430* (20), 3835-3846.

58. (a) Baxa, U.; Cassese, T.; Kajava, A. V.; Steven, A. C., Structure, function, and amyloidogenesis of fungal prions: filament polymorphism and prion variants. *Advances in protein chemistry* **2006,** *73*, 125-180; (b) Shewmaker, F.; Kryndushkin, D.; Chen, B.; Tycko, R.; Wickner, R. B., Two prion variants of Sup35p have in-register parallel β-sheet structures, independent of hydration. *Biochemistry* **2009,** *48* (23), 5074-5082; (c) Luckgei, N.; Schütz, A. K.; Bousset, L.; Habenstein, B.; Sourigues, Y.; Gardiennet, C.; Meier, B. H.; Melki, R.; Böckmann, A., The Conformation of the Prion Domain of Sup35 p in Isolation and in the Full-Length Protein. *Angewandte Chemie International Edition* **2013,** *52* (48), 12741-12744.

59. Ahmed, A. B.; Znassi, N.; Château, M.-T.; Kajava, A. V., A structure-based approach to predict predisposition to amyloidosis. *Alzheimer's & Dementia* **2015,** *11* (6), 681-690.

60. Hess, B.; Kutzner, C.; Van Der Spoel, D.; Lindahl, E., GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of chemical theory and computation* **2008,** *4* (3), 435-447.

61. Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M., PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of applied crystallography* **1993,** *26* (2), 283-291.

62. DeLano, W. L., Pymol: An open-source molecular graphics tool. *CCP4 Newsletter On Protein Crystallography* **2002,** *40* (1), 82-92.

63. Waldo, G. S.; Standish, B. M.; Berendzen, J.; Terwilliger, T. C., Rapid protein-folding assay using green fluorescent protein. *Nature biotechnology* **1999,** *17* (7), 691.

64. Yang, F.; Moss, L. G.; Phillips, G. N., The molecular structure of green fluorescent protein. *Nature biotechnology* **1996,** *14* (10), 1246.

65. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E., UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry* **2004,** *25* (13), 1605-1612.

66. Fiser, A.; Sali, A., ModLoop: automated modeling of loops in protein structures. *Bioinformatics* **2003,** *19* (18), 2500-2501.

Supporting information for:

Prion soft amyloid core driven self-assembly of globular proteins into bioactive nanofibrils

*Weiqiang Wang[1], Susanna Navarro[1], Rafayel A. Azizyan[2], Manuel Baño-Polo[1], Sebastian A. Esperante[1] Andrey V. Kajava[2] and Salvador Ventura[1]**

[1]Institut de Biotecnologia i de Biomedicina and Departament de Bioquímica i Biologia Molecular; Universitat Autònoma de Barcelona; 08193 Bellaterra (Barcelona), Spain.

[2]Centre de Recherche en Biologie cellulaire de Montpellier, UMR 5237 CNRS, Université Montpellier, 1919 Route de Mende, 34293 Montpellier, Cedex 5, France

E-mail: salvador.ventura@uab.es

This PDF file includes Supporting Figures S1 to S6 and Tables S1 and S4.

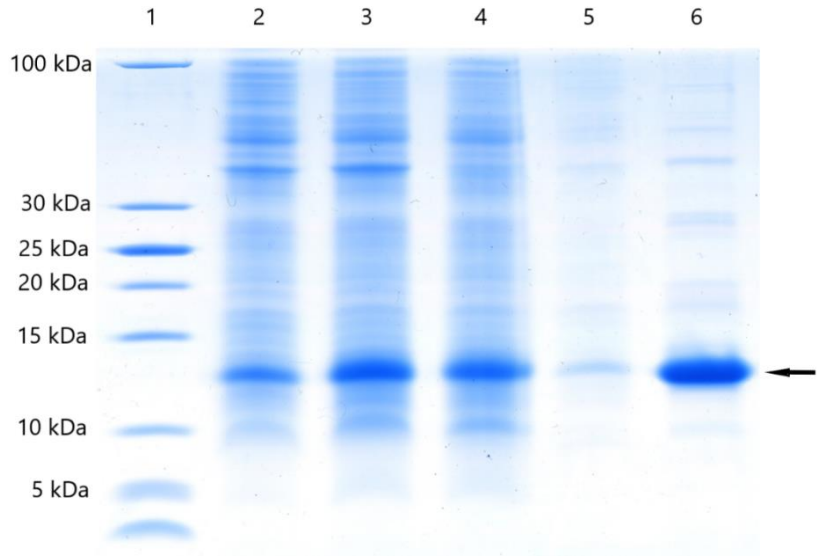

Figure S1. Analysis on SDS-PAGE of the expression and purification of Sup35-FF fusion protein. *Lane 1*, corresponds to molecular weight marker, *lane 2*, non-induced culture, *lane 3*, induced culture, *lane 4* and *5* are soluble (supernatant) and insoluble (pellet) fractions from total cell extract, and *lane 6* shows the purified fraction of Sup35-FF protein upon elution with 250 mM imidazole from a His-trap column. Black arrow indicates the band corresponding to Sup35-FF fusion protein.

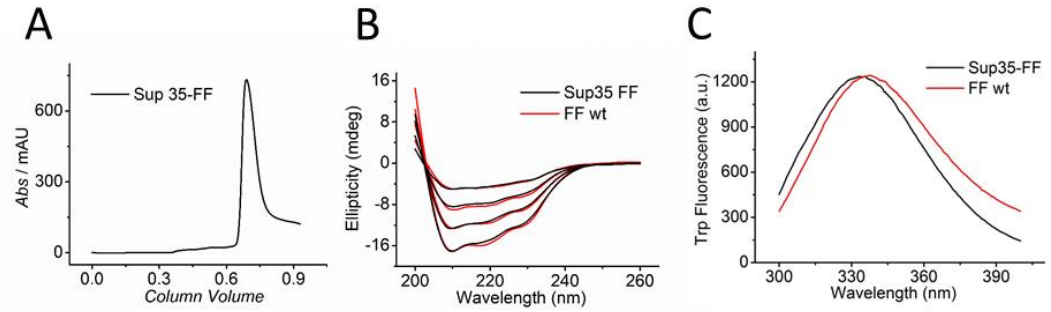

Figure S2. Conformational characterization of Sup35-FF protein. (A) Size-Exclusion Chromatography elution profile of Sup35-FF, (B) Far-UV CD spectra at different concentrations (5, 10, 15 and 20 μM) and, (C) Tryptophan intrinsic fluorescence spectra. Sup35-FF and FF-wt are shown in black and red, respectively.

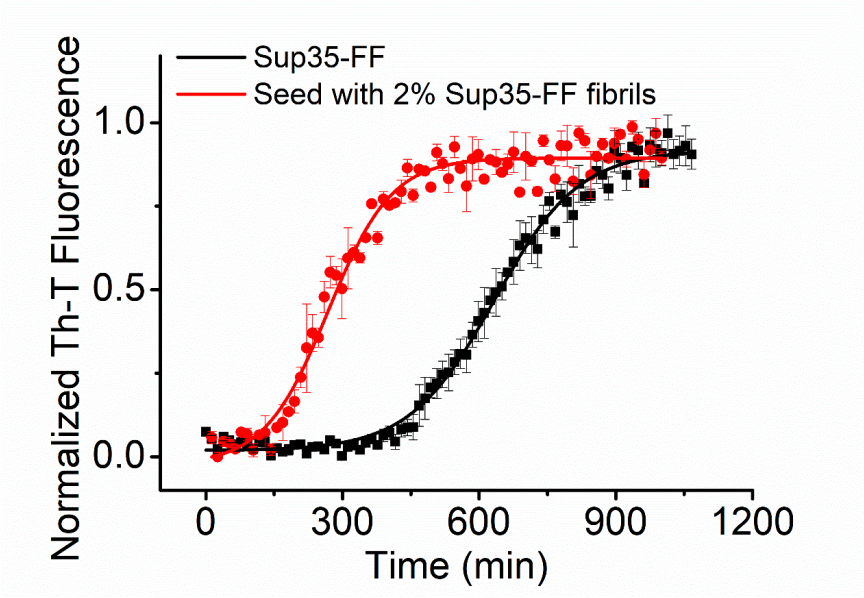

Figure S3. Aggregation kinetics and seeding reaction of Sup35-FF. Sup35-FF was dissolved at 50 μM in PBS containing 25 μM Th-T and the Th-T fluorescence emission was recorded along time in the absence (black squares) and, in the presence of 2% of pre-formed Sup35-FF fibrils as seeds (red dots).

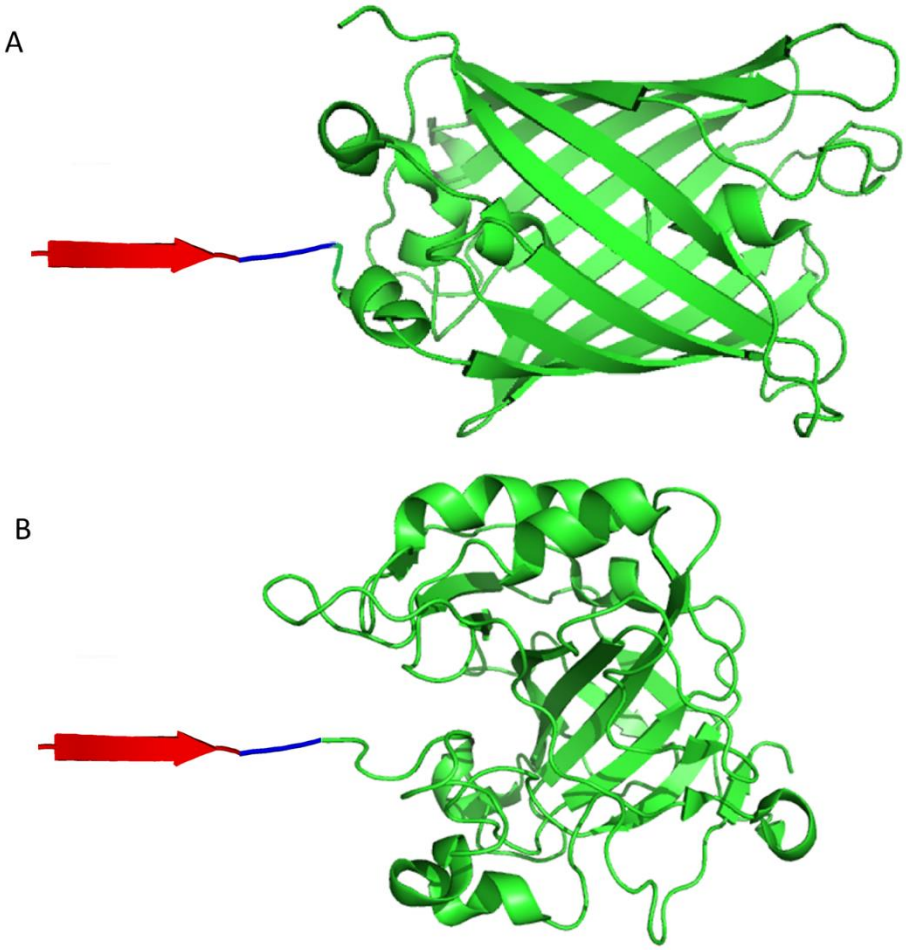

Figure S4. 3D representation of Green Fluorescent Protein and Carbonic anhydrase fused to the Sup35 soft amyloid core. Ribbon representation of (A) Sup35-GFP and (B) Sup35-CA using the PDB accession code 2B3Q and 1V9E for the GFP and CA structures, respectively. Sup35 soft amyloid core, spacer linker, GFP or CA are shown in red, blue and green, respectively.

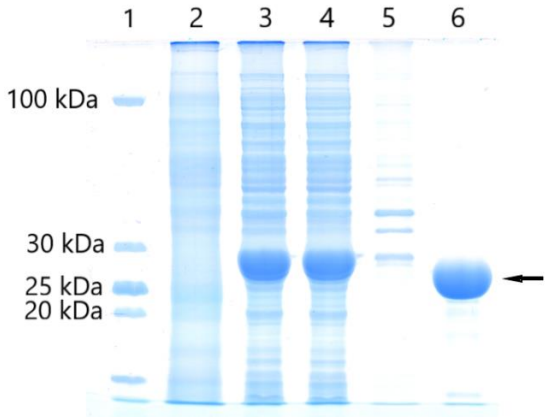

Figure S5. Analysis on SDS-PAGE of the expression and purification of Sup35-GFP fusion protein. *Lane 1*, corresponds to molecular weight marker, *lane 2*, non-induced culture, *lane 3*, total extract induced, *lane 4*, soluble fraction (supernatant), *lane 5*, insoluble fraction (pellet) and, *lane 6*, shows the purified fraction of Sup35-GFP by gel filtration. Black arrow indicates the band corresponding to Sup35-GFP fusion protein.

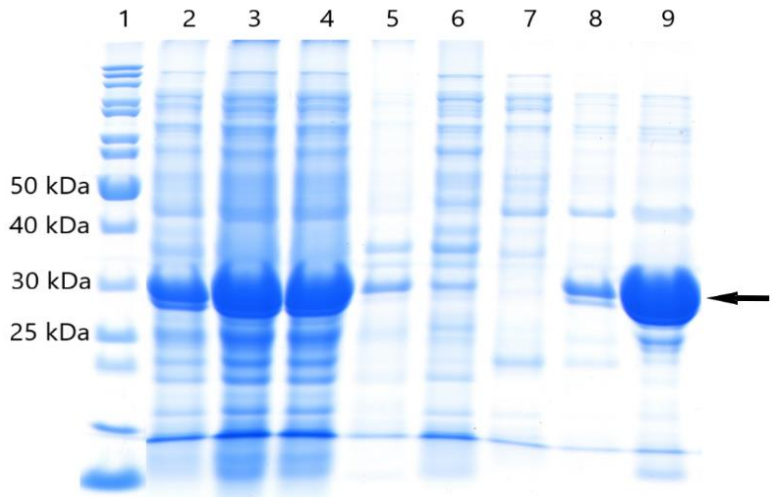

Figure S6. Analysis on SDS-PAGE of the expression and purification of Sup35-CA fusion protein. *Lane 1*, corresponds to molecular weight marker, *lane 2*, non-induced culture, *lane 3*, total extract induced, *lane 4*, soluble fraction (supernatant), *lane 5*, insoluble fraction (pellet) and *lanes 6-7*, flow through the His-trap column and, *lanes 8-9*, eluted protein at 250 mM imidazole. Black arrow indicates the band corresponding to Sup35-CA fusion protein.

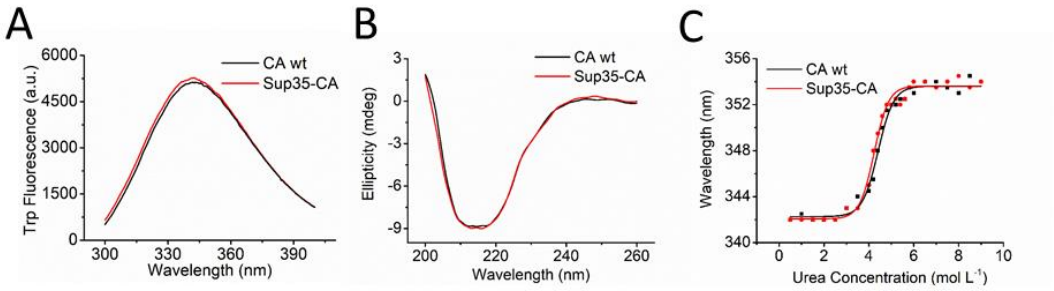

Figure S7. Conformational characterization and stability of Sup35-CA protein. (A) Tryptophan intrinsic fluorescence spectra, (B) Far-UV CD spectra, (C) Chemical equilibrium curves with urea were followed at 25 ºC by wavelength of maximum Trp fluorescence (Urea 50% of CA-wt and Sup35-CA are 4.24 ± 0.03 and 4.40 ± 0.04). CA-wt and Sup35-CA are shown in black and red, respectively.

Table S1. Prediction of aggregation propensity and disorder. (A) Positive aggregation-prone predictions are shown in bold. Positive scores, values higher than 21.4 and values lower than -4.0, correspond to aggregation-prone proteins/regions according to Aggrescan, FoldAmyloid, and PASTA 2.0, respectively. The analysis of a classical amyloid core belonging to the Aβ-peptide is shown for comparison (B) For disorder prediction, the values indicate the percentage of polypeptide predicted to be disorder.

A

Protein	**Aggrescan**	**FoldAmyloid**	**PASTA** 2.0
KLVFFA (Aβ)	**81.00**	**24.02**	**-4.53937**
Sup35-SAC	-32.60	20.84	-1.9920
Sup35-SAC-linker	-35.50	19.99	-1.9920
FF domain-His tag	-26.30	21.22	**-4.2112**
Sup35-FF	-27.60	20.81	**-4.2112**

B

Protein	**PONDR (VSL2)**	**GlobPlot**	**PASTA** 2.0	**IUPRED**
Sup35-SAC	Too short	100.00	100.00	56.52
Sup35-SAC-linker	100.00	100.00	100.00	54.84
FF domain-His tag	32.81	9.38	23.43	10.94
Sup35-FF	52.63	33.68	25.26	12.63

Table S2. Thermodynamic characterization of soluble Sup35-FF and FF-wt proteins.

	[a] $\Delta G_{U-F}^{H_2O}$ (Kcal mol^{-1})		[b] m (Kcal mol^{-1} M^{-1})		[c] [Urea]$_{50\%}$ (mol·L^{-1})	
	Intrinsic Fluorescence	**CD**	**Intrinsic Fluorescence**	**CD**	**Intrinsic Fluorescence**	**CD**
Sup35-FF	4.04±0.18	4.02±0.28	0.72±0.03	0.76±0.05	5.62±0.04	5.32±0.08
FF-wt	4.11±0.11	4.21±0.20	0.72±0.02	077±0.04	5.68±0.04	5.44±0.04

[a] Gibbs energy of unfolding with urea determined from the equilibrium parameters.

[b] Dependence of the Gibbs energy of unfolding with urea.

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

Table S3. Melting temperatures of soluble Sup35-FF and FF-wt proteins.

	T_m (℃)	
	Intrinsic Fluorescence	**CD**
Sup35-FF	62.72 ± 0.17	64.7 ± 0.4
FF-wt	64.61 ± 0.19	66.1 ± 0.5

Table S4. Assignment and area of the secondary structure components of Sup35-FF fibrils in the amide I region of the FTIR spectra. FF-wt was shown as control.

Assignments (%)	Sup35-FF fibrils	FF-wt
Inter β-sheet	27.25 (1624 cm^{-1})	-
α-helix	59.50 (1649 cm^{-1})	84.08 (1650 cm^{-1})
Turns	13.25 (1674 cm^{-1})	15.92 (1673 cm^{-1})

Chapter II

Amyloidogenicity as a driving force for the formation of functional oligomers

*Rafayel A. Azizyan[1,2], Weiqiang Wang[3] Alexey Anikeenko[4,] Zinaida Radkova[4], Anastasia Bakulina[4], Adriana Garro[5], Landry Charlier[6], Christian Dumas[7], Salvador Ventura[3] and Andrey V. Kajava[1,2]**

1 Centre de Recherche en Biologie cellulaire de Montpellier, UMR 5237 CNRS, Université Montpellier, Montpellier, France.

2 Institut de Biologie Computationnelle, Université Montpellier, Montpellier, France.

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

4 Novosibirsk State University, Novosibirsk, Russia.

5 Universidad Nacional de San Luis IMASL-CONICET, San Luis, Argentina.

6 Institut des Biomolécules Max Mousseron, Montpellier, France.

7 Centre de Biochimie Structurale, CNRS, UMR5048, INSERM U1054, Université de Montpellier, Montpellier, France.

*Correspondence and requests should be addressed to A.V.K. (email: andrey.kajava@crbm.cnrs.fr)

Abstract

Insoluble amyloid fibrils formed by self-assembly of amyloidogenic regions of proteins have a cross-β-structure. In this work, by using targeted molecular dynamics and rigid body simulation, we demonstrate that if a protein consists of an amyloidogenic region and a globular domain(s) and if the linker between them is short enough, such molecules cannot assemble into amyloid fibrils, forming instead oligomers with a certain limited number of β-strands in the cross-β core. We show that this blockage of the amyloid growth is due to the steric repulsion of globular structures linked to amyloidogenic regions. Furthermore, we establish the relationship between the linker length and the number of monomers in such nanoparticles. We hypothesize that such oligomerization can be a yet unrecognised way to form natural protein complexes involved in biological processes. Our results can also be used in protein engineering for designing soluble nanoparticles carrying different functional domains.

Keywords: Amyloids, targeted molecular dynamics, rigid body simulation, functional nanoparticles.

Introduction

Amyloid fibrils are the subject of particular interest due to their association with a number of human diseases. Despite considerable importance, until recently, the structural arrangement of amyloid fibrils was poorly understood compared with soluble proteins. This was due to the fact that conventional experimental methods (X-ray crystallography and NMR in solution) capable of determining high-resolution structure, cannot be used because of the insolubility of fibrils. The recent progress in understanding the amyloid structure has stemmed mostly from the application of experimental techniques such as solid-state nuclear magnetic resonance, cryo-electron microscopy, and scanning transmission electron microscopy mass measurements (Nelson and Eisenberg 2006; Steven et al. 2016). The majority of the naturally-occurring amyloid structures have been shown to have a cross-β structure with parallel and in-register β-strands (Colvin et al. 2016; Gorkovskiy et al. 2014; Groveman et al. 2014; Helmus et al. 2011; Kajava et al. 2004; Kajava, Baxa, and Steven 2010; Luckgei et al. 2013; Murray et al. 2017; Rodriguez et al. 2015; Vilar et al. 2008; Wasmer et al. 2008; Weirich et al. 2016) (**Figure 1a**). The formation of the parallel and in-register structures from the amyloidogenic polypeptides is spontaneous and leads to the stable cross-β fibrils (Colvin et al. 2016; Gorkovskiy et al. 2014; Groveman et al. 2014; Helmus et al. 2011; Kajava et al. 2004, 2010; Luckgei et al. 2013; Luhrs et al. 2005; Vilar et al. 2008; Weirich et al. 2016), suggesting that the interaction between the β-structural subunits of amyloids is energetically favourable. These interactions are strong enough to bring together large proteins containing, in addition to the amyloidogenic region (AR), globular domains and long unfolded linkers, as observed, for example, in Rip1/Rip3, huntingtin, and TAR-DNA binding protein, Het-s, Sup35p, Ure2p and Prp (Baxa et al. 2006, 2007; Chen et al. 2010; Groveman et al. 2014; Helmus et al. 2011; Li et al. 2012; Pfefferkorn, McGlinchey, and Lee 2010; Shen et al. 2016; Shewmaker et al. 2009; Wasmer et al. 2009) (**Figure 1b**). Usually, in these fibrils, the unfolded linker connecting the amyloidogenic regions (AR) and globular domain is long (more than 50 residues), since AR location in close proximity to the globular domain along the chain would cause steric repulsion and prevent the formation of amyloid structures (Kajava et al. 2004) (**Figure 1c**). Recently, by using molecular and mesoscopic modelling, we were able to establish a relationship between the size of the

globular domains and the length of the shortest possible linker that still allows the formation of the infinite amyloid fibrils having a cross-β structure with parallel and in-register β-strands (Azizyan et al., 2018). This relationship was confirmed experimentally (Wang et al. 2019). The objective of the present work was to study the self-assembly of ARs when the linker between the AR and the globular domain is shorter than the one allowing the formation of the infinite fibrils.

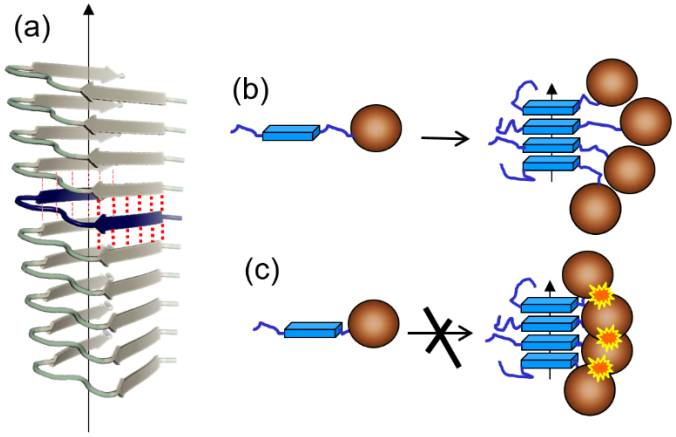

Figure 1. Structural arrangement of amyloid fibril and two possible scenarios of amyloid structure formation in the presence of flanking globular domains. (**a**) A typical structural arrangement of amyloid fibrils. The β-arch is stacked in a parallel and in-register manner into fibrillar 'β-arcade' structure. Two strands of the β-arch are integrated into two different β-sheets. (**b**) Stack of ARs (as boxes of blue colour) connected with globular domains (spheres of brown colour) by long linkers. These molecules can form infinite amyloid fibrils. The fibre axis is indicated by a thin vertical arrow. (**c**) Stack of ARs connected with globular domains by short linkers. The stacking of ARs is hampered by the steric repulsion of globular domains.

Materials and methods

Coarse-grained targeted molecular dynamics. We used the CG-TMD approach implemented with the GROMACS package (version 4.6.7) (Hess et al. 2008; Van Der Spoel et al. 2005) installed on a multi-processor Xeon based workstation with an SSE4.1 CPU acceleration set. To evaluate the steric tensions and overall stereochemistry of the oligomers obtained during the coarse-grained TMD, we converted them back to the all-atom structures by using the "Going backward" tool (Wassenaar et al. 2014). Then, we applied an energy minimisation procedure using the steepest descent algorithm of the GROMACS package version 4.6.7 (Hess et al. 2008; Van Der Spoel et al. 2005) and evaluated steric tension and overall stereochemistry of

the structures by using the PROCHECK package (Laskowski et al. 1993). For further details, see Supplementary Data.

Rigid Body Simulation. As a simplified mesoscopic modelling strategy, we used the Low Poly 3D models and Rigid Body Simulation implemented in Blender software package[37]. For additional information, see https://cg3dartist.wixsite.com/amyloid. The final structures were checked for the absence of intersections between the elements of models with Mesh Analysis tool of Blender. All other geometrical parameters and Rigid Body World setting were the same as in the previous work[36]. For further details, see Supplementary Data.

Results

Modelling of self-assembly of a hybrid protein containing AR and typical globular domain by using full-atom and coarse-grained targeted molecular dynamics. To analyse the constraints on amyloid formation imposed by steric exclusion of globular domains, we used a hybrid protein containing an amyloid-forming fragment of Aβ peptide (17-42) and Green Fluorescent Protein (GFP) as a globular domain. This choice is explained by the fact that on the one hand, the Aβ-fragment is the most studied amyloid-forming peptide and, on the other hand, the GFP globular domain is the most frequently fused to different amyloid-forming domains (Fox et al. 2010; Ochiishi et al. 2016; Waldo et al. 1999; Wurth, Guimard, and Hecht 2002). Besides, GFP, with its 238 residues, represents a typical globular domain being close to the average size of the globular domain (190 residues in MODBASE) (Pieper et al. 2014). The structure of the Aβ protofibril was taken from the PDB entry 2BEG (Luhrs et al. 2005). In this amyloid structure, each polypeptide chain is folded into β-arches and stacked in parallel and in-register manner (Fig. 1a). The 3D structure of GFP was taken from PDB entry 1GFL (Yang, Moss, and Phillips 1996). In our model, the linker between these domains has a Ser-Pro-rich sequence, which has a potential of flexible and intrinsically unfolded conformation (**Figure 2a**). In particular, 1, 2, 3, 4, 5, and 6-residue linkers consist of Ser, Ser-Ser, Pro-Ser-Ser, Ser-Pro-Ser-Ser, Ser-Ser-Pro-Ser-Ser, Pro-Ser-Ser-Pro-Ser-Ser, respectively.

Molecular modelling showed that Aβ-GFP with linkers shorter than 7 residues can still form the amyloid cross-β structures; however, these complexes represent

oligomers with a limited number of the monomers because of the steric repulsion of the globular domains. We were able to establish the relationship between the number of subunits in the oligomers and the length of the linker. For this analysis, we modelled the self-assembly of Aβ(17-42)-linker-GFP molecules by using targeted molecular dynamics (TMD) (Schlitter et al. 1993). If initial and target conformations are known, then the TMD simulation is a suitable method to predict the transition pathways by continuously diminishing the RMS distance value between these conformations. In our case, the initial state of AR is an unfolded conformation, and the final structure is a β-arch within the β-arcade (Luhrs et al. 2005) (**Figure 1a**). Concerning the other parts of the hybrid molecule, the GFP globular domain was kept in its initial crystal structure and linkers of a given length were free of constraints and flexible. The monomers were added to the growing complexes until they started to have steric tensions. Our TMD simulations, made by using the full-atom system with and without explicit water, show that this system is calculation expensive. Therefore, we turned our attention to the TMD applied to coarse-grained (CG) molecular models (Monticelli et al. 2008; Tozzini, Rocchia, and McCammon 2006). In CG models, the reduced number of degrees of freedom and the use of smoother interaction potentials allow for longer time steps, resulting in a significant increase in the calculation speed (Dror et al. 2010). Our tests showed that the CG-TMD approach implemented with the GROMACS package fits our system well. An addition of a monomer to the β-arcade target requires about 18 hours of CPU time on a multi-processor Xeon based workstation with an SSE4.1 CPU acceleration set. To evaluate the steric tensions and overall stereochemistry of the oligomers obtained during the coarse-grained TMD, we converted them back to the all-atom structures by using the "Going backward" tool (Wassenaar et al. 2014). The oligomer structures with disallowed atomic contacts or covalent structure geometry were rejected. (For details see Supplementary Data).

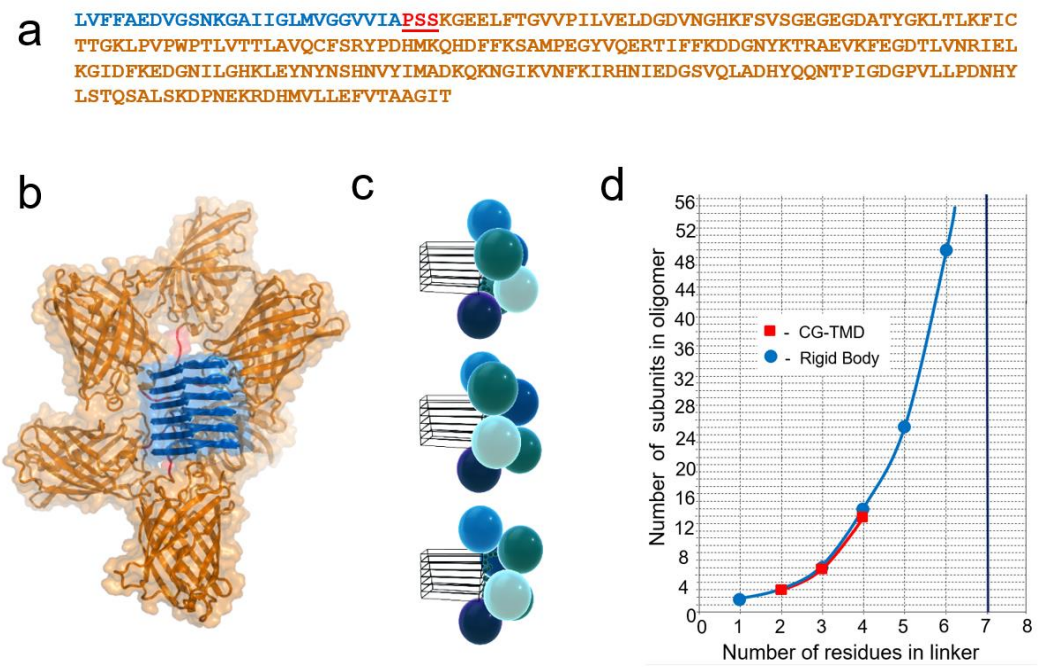

Figure 2. Models of an amyloid-like oligomer formed by hybrid molecule Aβ(17-42)-linker-GFP with 3 residue linker. (**a)** Sequence of Aβ(17-42)-linker-GFP monomer with Aβ(17-42) peptide (PDB entry 2BEG) (in blue), a linker (red and underlined) composed of Pro-Ser-Ser sequence (C-terminal Ser comes from the N-terminus of GFP), and GFP domain from Aequorea victoria (PDB entry 2Y0G) (in light brown). (**b)** The all-atom 3D structure of a hexamer containing monomers with 3 residues in the linker obtained by CG MD simulations (picture was generated using VMD package (Humphrey, Dalke, and Schulten 1996)) (regions of the hybrid molecule are coloured similarly on **a** and **b**). (**c)** Several arrangements of the hexamers obtained in the different runs of the Blender Rigid Body simulation. (**d)** Dependence of the number of subunits in oligomer on the linker length. The vertical line at the 7-residue value denotes the minimal number of residues in the linker allowing the growth of infinite amyloid fibril. Red squares and blue points correspond to data obtained in CG-TMD and rigid body simulations, correspondingly.

The results of our analysis are shown in **Figure 2b** and **Figure 2d**. When the linker consists of 1, 2, 3 or 4 residues, the oligomers can have maximum 2, 3, 6 or 13 subunits, correspondingly. The longer the linker, the higher stoichiometry of the oligomer structure, and more calculation time is required to reach the steric limits. Given that Aβ(17-42)-linker-GFP molecules with a 7-residue linker, can form the infinite fibril (Azizyan et al. 2018), we expected a sharp increase in the number of subunits for the molecules with 5 or 6 residue linkers. In this case, the precise estimation of the number of subunits in the oligomer was becoming computationally expensive and would

require the other modelling approaches. Thus, our study revealed some limitations of CG-TMD simulations when this method was applied to the larger oligomers.

Generalisation of the conclusions by using rigid body simulation. The results of CG-TMD described in the previous section can be used to validate the other simplified modelling approaches. At the same time, the simplified approaches can allow extending our conclusion to a more general case, with different linker lengths, shapes, and sizes of the AR and globular domains. As a simplified mesoscopic modelling strategy, we used the Rigid Body Simulation implemented in Blender software package (Anon 2015) (for details see Supplementary Data). The globular part was represented as a sphere with a diameter of 30 Å. All other geometrical parameters and Rigid Body World setting were the same as in the previous work (Azizyan et al. 2018). The obtained dependence between the number of subunits in oligomers and the number of residues in the linkers is shown in **Figure 2d**, **Figure 3** and **Table S1**. The comparison of the results derived from the rigid body simulation and CG-TMD showed a good agreement for the linkers with 2, 3, and 4 residues. This comparison allowed us to validate the rigid body simulation results on the CG TMD results. The agreement between the results suggested that the simplified approach could be successfully used for rapid estimations of the larger systems. For example, for the linkers of 5 and 6 residues, the rigid body simulation gave 29 and 69 monomers in the oligomers, respectively (**Figure 3**). The simplified method was rapid enough to run several simulations for the same system. It was shown that depending on the initial positions of monomers during the simulation, the stoichiometry of the obtained oligomers can vary (**Figure 3** and **Table S1**). Moreover, even when the number of monomers in the oligomers is the same, the arrangement of the globular parts around the amyloid core can be slightly different (**Figure 2c**). This variation can be explained by the local "kinetic traps" presented in such large molecular complexes.

Our modelling was done for GFP, which size is close to the average one of the globular domains (Pieper et al. 2014). At the same time, if the linker length is fixed, the number of subunits in the oligomer should vary depending on the size of globular structure. To demonstrate this, we extended the rigid body simulations to the oligomers formed by subunits with globular domains of 40 Å or 50 Å in diameter (**Figure 3**). For example, in accordance with this estimation, molecules with 6 residue linker and

globular domain of 30, 40 or 50 Å can form oligomers of 69, 18 or 12 subunits, correspondingly.

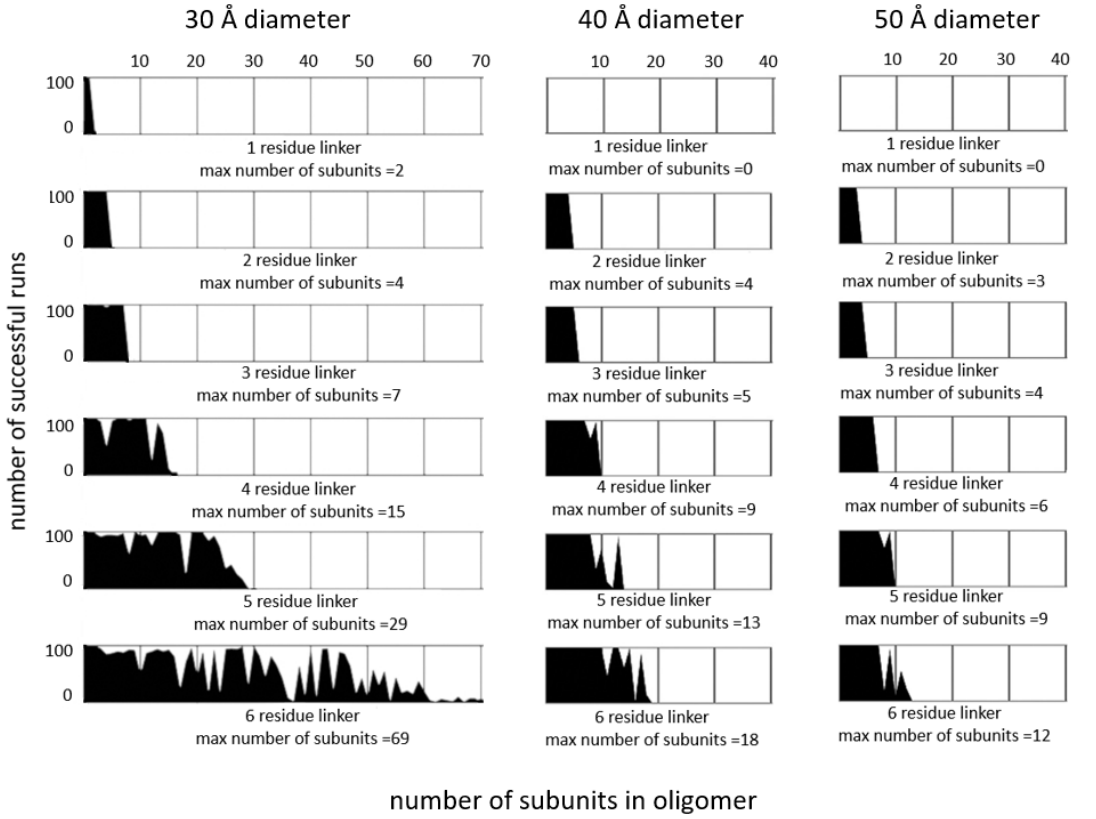

Figure 3. Dependence between the number of subunits in the cross-β oligomers and the number of residues in the linkers obtained by rigid body simulation. Globular domain was represented by a sphere of either 30, 40 or 50 Å in diameter. Linkers vary from 1 to 6 residues. For a given linker size and number of subunits, we performed 100 rigid body simulations and counted the number of minimized oligomers without steric repulsion.

The rigid body simulation does not take into consideration neither a specific amino acid sequence of the linker nor its direction in terms of N- and C-terminus. The agreement between the results of the CG TMD and rigid body simulations suggests that our results can be extended to a wide range of linker sequences independent if the globular domain is N- or C-terminal to the amyloidogenic region. At the same time, the linker sequences should be non-amyloidogenic and have flexible and intrinsically unfolded conformation.

A methodological result of this work is the implementation of CG TMD and Rigid Body simulations as suitable tools to predict the assembly of such large systems containing molecules with amyloid-forming regions linked to globular structures. The

modelling protocol established here can be used in future studies. Here, we analysed the most common case of naturally occurring amyloids with the parallel and in-register cross-β structure, but some of the cross-β amyloid fibrils are built by antiparallel stacking of β-arches (Qiang et al. 2012) or short β-solenoids (Wasmer et al. 2008). These fibrils can better tolerate the short linkers of the globular domains and can be analysed by using the approaches presented here. ARs that are flanked by two globular domains on both sides can also be tested.

Discussion

Flanking domains of the amyloidogenic regions can affect the formation and the stability of cross-β amyloids. In the most well-studied amyloids, the unfolded linker connecting the amyloidogenic region and globular domain is long (more than 50 residues) (Baxa et al. 2006, 2007; Chen et al. 2010; Groveman et al. 2014; Helmus et al. 2011; Li et al. 2012; Pfefferkorn, McGlinchey, and Lee 2010; Shen et al. 2016; Shewmaker et al. 2009; Wasmer et al. 2009). In this case, the thermal motion of the globular domain and long unfolded flanking region should inhibit the amyloidogenesis by pushing away the amyloidogenic regions due to an excluded volume effect (Rubinshtein and Colby, 2003). Here, we considered proteins with globular structures linked to amyloidogenic regions by short (less than 7 residues) linkers. We showed that the amyloid growth of these proteins is blocked due to the steric repulsion of their globular domains. In comparison with the relatively weak inhibition of the long linkers, which can only shorten the fibrils, the steric effect of proteins with the short linkers is abrupt and strong. As a result, proteins with short linkers form oligomers with a cross-β core of a certain limited size instead of the fibrils.

In most cases, amyloid fibril formation is nucleation-dependent, displaying a lag phase followed by rapid fibrillogenesis (Ferrone 1999; Xue, Homans, and Radford 2008). Knowledge about the size of the amyloid nucleus may inform us of the size in which the oligomers described here are becoming stable. Nowadays, the question about the exact number of monomers in the nuclei of amyloids is still debated. In the β-arcade, the formation of the first intermolecular H-bonds between two stacked β-arches is the most important step (Kajava et al. 2010), though it is not clear how many additional β-arches are needed to make this β-arcade stable. In other works, it was suggested that

the nucleus of amyloid fibrils might consists of 4, 6, and even up to 30 monomers (Nelson et al. 2005, Xue et al. 2008, Sorci et al. 2011). If we consider the highest number of monomers suggested for the nuclei structures, we can conclude that our hybrid molecules form stable oligomers when the linkers are 5 or more residues long (**Figure 3**).

Our results suggest that nature can use this type of amyloid-like oligomerisation to form soluble complexes carrying functional domains. A number of protein oligomers, including α-helical coiled coils, triple-helical collagen structures and complexes made from globular domains have been characterized (Steven et al. 2016). The amyloid-like oligomers described in this work have not been observed by the conventional methods of structural biology (such as X-ray crystallography or NMR spectroscopy) yet. This can be explained by the heterogeneity of these superstructures, differing in the number of monomers and in their spatial arrangement around the amyloid-like core (**Figure 2c**, **Table S1**), which hampers their atomic-level structural determination. The oligomerisation observed here may also provide structural insight into a growing body of evidence on the phase separation behavior of proteins with amyloidogenic regions (Alberti et al. 2018) Depending on conditions, several proteins with amyloid-forming regions, such as Sup35p, TDP-43, p53, CPEB3, FUS and others, also have potential to form assemblies with spherical morphologies (Mészáros et al., 2020). These observations put forth questions about the relationship between phase separation and amyloid fibrils. Our results suggest that in some cases, the aggregation state of the amyloidogenic region may be controlled by folding of the neighbouring domains.

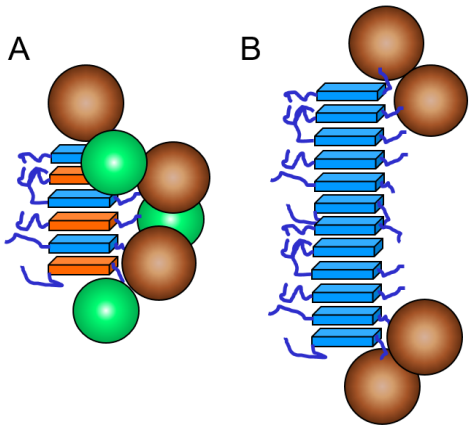

Figure 4. Schematic representation of some potential applications of amyloid-like oligomeric structures. (**A)** Soluble oligomers carrying several different functional domains. (**B)** Inhibition of the amyloid fibril growth by addition of hybrid molecules with globular domains.

This proof-of-concept study opens up further opportunities for the fabrication of nanostructures of defined size carrying multiple functional domains (**Figure 4A**). These engineered proteins can also be exploited to inhibit amyloidosis, by adding to the amyloidogenic proteins containing a given ARs a hybrid molecule consisting of the same AR linked to a globular domain (**Figure 4B**). The amyloid fibril scaffolding is already used in protein engineering to build up functional assemblies (Giraldo 2010). However, in comparison with the insoluble amyloid fibrils, the oligomers may be soluble, suggesting new opportunities in nanobiotechnology.

Acknowledgements

The authors thank Drs Nicolas Floquet and Andrey Frolov for assistance with CG-TMD approach and Ekaterina Davidova for critical reading of the manuscript. They also acknowledge the Armenian Communities Department of Calouste Gulbenkian Foundation for providing “Global Excellence Scholarship” to R.A., Erasmus Mundus program for providing travel grant to A.G. This work was also supported by the Russian Ministry of Science and Education under 5-100 Excellence Program, by EU COST Action BM1405 NGP-net, by Ministerio de Economía y Competitividad (MINECO, Spain) (BIO2016-78310-R) and by ICREA (ICREA-Academia 2015).

Additional information

Detailed methods, an additional table and figures are presented in Supporting Information.

References

Alberti, S, Gladfelter, A. and Mittag T. 2018. “Considerations and Challengesin Studying Liquid-Liquid Phase Separation.” *Cell*, 176**,** 419-34.

Anon. 2015. “Blender.Org - Home of the Blender Project - Free and Open 3D Creation Software.” *Blender.Org*.

Azizyan, Rafayel A., Adriana Garro, Zinaida Radkova, Alexey Anikeenko, Anastasia Bakulina, Christian Dumas, and Andrey V. Kajava. 2018. “Establishment of Constraints on Amyloid Formation Imposed by Steric Exclusion of Globular Domains.” *Journal of Molecular Biology* 430(20):3835–46.

Baxa, U., R. B. Wickner, A. C. Steven, D. E. Anderson, L. N. Marekov, W. M. Yau, and R. Tycko. 2007. “Characterization of Beta-Sheet Structure in Ure2p1-89 Yeast Prion Fibrils by Solid-State Nuclear Magnetic Resonance.” *Biochemistry* 46(45):13149–62.

Baxa, Ulrich, Todd Cassese, Andrey V. Kajava, and Alasdair C. Steven. 2006. “Structure, Function, and Amyloidogenesis of Fungal Prions: Filament Polymorphism and Prion Variants.” *Advances in Protein Chemistry* 73:125–80.

Berendsen, H. J. C., J. P. M. Postma, W. F. Van Gunsteren, A. Dinola, and J. R. Haak. 1984. “Molecular Dynamics with Coupling to an External Bath.” *The Journal of Chemical Physics* 81(8):3684–90.

Chen, Allan K. H., Ryan Y. Y. Lin, Eva Z. J. Hsieh, Pang Hsien Tu, Rita P. Y. Chen, Tai Yan Liao, Wenlung Chen, Chih Hsien Wang, and Joseph J. T. Huang. 2010. “Induction of Amyloid Fibrils by the C-Terminal Fragments of TDP-43 in Amyotrophic Lateral Sclerosis.” *Journal of the American Chemical Society* 132(4):1186–87.

Colvin, Michael T., Robert Silvers, Qing Zhe Ni, Thach V. Can, Ivan Sergeyev, Melanie Rosay, Kevin J. Donovan, Brian Michael, Joseph S. Wall, Sara Linse, and Robert G. Griffin. 2016. “Atomic Resolution Structure of Monomorphic Aβ(42) Amyloid Fibrils.” *Journal of the American Chemical Society* 138(30):9663–74.

Dror, Ron O., Morten Ø. Jensen, David W. Borhani, and David E. Shaw. 2010. “Exploring Atomic Resolution Physiology on a Femtosecond to Millisecond Timescale Using Molecular Dynamics Simulations.” *The Journal of General Physiology* 135(6):555–62.

Ferrone, Frank. 1999. “Analysis of Protein Aggregation Kinetics.” *Methods in Enzymology* 309:256–74.

Fiser, A., R. K. Do, and A. Sali. 2000. “Modeling of Loops in Protein Structures.” *Protein Sci* 9(9):1753–73.

Fiser, András and Andrej Sali. 2003. “ModLoop: Automated Modeling of Loops in Protein Structures.” *Bioinformatics* 19(18):2500–2501.

Fox, Ayano, Thibaut Snollaerts, Camille Errecart Casanova, Anastasia Calciano, Luiza A. Nogaj, and David A. Moffet. 2010. “Selection for Nonamyloidogenic Mutants of Islet Amyloid Polypeptide (IAPP) Identifies an Extended Region for Amyloidogenicity.” *Biochemistry* 49(36):7783–89.

Giraldo, Rafael. 2010. “Amyloid Assemblies: Protein Legos at a Crossroads in Bottom-Up Synthetic Biology.” *ChemBioChem* 11(17):2347–57.

Gorkovskiy, Anton, Kent R. Thurber, Robert Tycko, and Reed B. Wickner. 2014. “Locating Folds of the In-Register Parallel β-Sheet of the Sup35p Prion Domain Infectious Amyloid Laboratories of a Biochemistry and Genetics And.” *PNAS* 111(43):4615–22.

Groveman, Bradley R., Michael A. Dolan, Lara M. Taubner, Allison Kraus, Reed B. Wickner, and Byron Caughey. 2014. “Parallel In-Register Intermolecular β-Sheet Architectures for Prion-Seeded Prion Protein (PrP) Amyloids.” *Journal of Biological Chemistry* 289(35):24129–42.

Helmus, Jonathan J., Krystyna Surewicz, Marcin I. Apostol, Witold K. Surewicz, and Christopher P. Jaroniec. 2011. “Intermolecular Alignment in Y145Stop Human Prion Protein Amyloid Fibrils Probed by Solid-State NMR Spectroscopy.” *Journal of the American Chemical Society* 133(35):13934–37.

Hess, Berk, Carsten Kutzner, David Van Der Spoel, and Erik Lindahl. 2008. “GRGMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation.” *Journal of Chemical Theory and Computation* 4(3):435–47.

Humphrey, William, Andrew Dalke, and Klaus Schulten. 1996. “VMD: Visual Molecular Dynamics.” *Journal of Molecular Graphics* 14(1):33–38.

Kajava, A. V, U. Baxa, and A. C. Steven. 2010. “Arcades: Recurring Motifs in Naturally Occurring and Disease-Related Amyloid Fibrils.” *The FASEB Journal* 24(5):1311–19.

Kajava, Andrey V, Ulrich Baxa, Reed B. Wickner, and Alasdair C. Steven. 2004. “A Model for Ure2p Prion Filaments and Other Amyloids: The Parallel Superpleated Beta-Structure.” *Proceedings of the National Academy of Sciences of the United States of America* 101(21):7885–90.

Laskowski, R. A., M. W. MacArthur, D. S. Moss, and J. M. Thornton. 1993. “PROCHECK: A

Program to Check the Stereochemical Quality of Protein Structures." *Journal of Applied Crystallography* 26(2):283–91.

Li, Jixi, Thomas McQuade, Ansgar B. Siemer, Johanna Napetschnig, Kenta Moriwaki, Yu Shan Hsiao, Ermelinda Damko, David Moquin, Thomas Walz, Ann McDermott, Francis Ka Ming Chan, and Hao Wu. 2012. "The RIP1/RIP3 Necrosome Forms a Functional Amyloid Signaling Complex Required for Programmed Necrosis." *Cell* 150(2):339–50.

Luckgei, Nina, Anne K. Schütz, Luc Bousset, Birgit Habenstein, Yannick Sourigues, Carole Gardiennet, Beat H. Meier, Ronald Melki, and Anja Böckmann. 2013. "The Conformation of the Prion Domain of Sup35 p in Isolation and in the Full-Length Protein." *Angewandte Chemie - International Edition* 52(48):12741–44.

Luckgei, Nina, Anne K. Schütz, Birgit Habenstein, Luc Bousset, Yannick Sourigues, Ronald Melki, Beat H. Meier, and Anja Böckmann. 2014. "Solid-State NMR Sequential Assignments of the Amyloid Core of Sup35pNM." *Biomolecular NMR Assignments* 8(2):365–70.

Luhrs, T., C. Ritter, M. Adrian, D. Riek-Loher, B. Bohrmann, H. Dobeli, D. Schubert, and R. Riek. 2005. "3D Structure of Alzheimer's Amyloid- (1-42) Fibrils." *Proceedings of the National Academy of Sciences* 102(48):17342–47.

Marrink, Siewert J., H. Jelger Risselada, Serge Yefimov, D. Peter Tieleman, and Alex H. De Vries. 2007. "The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations." *Journal of Physical Chemistry B* 111(27):7812–24.

Mészáros, Bálint, Gábor Erdős, Beáta Szabó, Éva Schád, Ágnes Tantos, Rawan Abukhairan, Tamás Horváth, Nikoletta Murvai, Orsolya P Kovács, Márton Kovács, Silvio C E Tosatto, Péter Tompa, Zsuzsanna Dosztányi, Rita Pancsa. 2020. PhaSePro: the database of proteins driving liquid–liquid phase separation, Nucleic Acids Research, gkz848, https://doi.org/10.1093/nar/gkz848

Monticelli, L., S. K. Kandasamy, X. Periole, R. G. Larson, D. P. Tieleman, and S. J. Marrink. 2008. "The MARTINI Coarse Grained Force Field: Extension to Proteins." *J. Chem. Theory Comput.* 4(5):819–34.

Murray, Dylan T., Masato Kato, Yi Lin, Kent R. Thurber, Ivan Hung, Steven L. McKnight, and Robert Tycko. 2017. "Structure of FUS Protein Fibrils and Its Relevance to Self-Assembly and Phase Separation of Low-Complexity Domains." *Cell* 171(3):615-627.e16.

Nelson, Rebecca and David Eisenberg. 2006. "Structural Models of Amyloid-Like Fibrils." *Advances in Protein Chemistry*.

Nelson, Rebecca, Michael R. Sawaya, Melinda Balbirnie, Anders Madsen, Christian Riekel, Robert Grothe, and David Eisenberg. 2005. "Structure of the Cross-β Spine of Amyloid-like Fibrils." *Nature* 435(7043):773–78.

Ochiishi, Tomoyo, Motomichi Doi, Kazuhiko Yamasaki, Keiko Hirose, Akira Kitamura, Takao Urabe, Nobutaka Hattori, Masataka Kinjo, Tatsuhiko Ebihara, and Hideki Shimura. 2016. "Development of New Fusion Proteins for Visualizing Amyloid-β Oligomers in Vivo." *Scientific Reports* 6.

Periole, X., M. Cavalli, S. J. Marrink, and M. A. Ceruso. 2009. "Combining an Elastic Network With a Coarse-Grained Molecular Force Field: Structure, Dynamics, and Intermolecular Recognition." *Journal of Chemical Theory and Computation* 5(9):2531–43.

Pfefferkorn, C. M., R. P. McGlinchey, and J. C. Lee. 2010. "Effects of PH on Aggregation Kinetics of the Repeat Domain of a Functional Amyloid, Pmel17." *Proceedings of the National Academy of Sciences* 107(50):21447–52.

Pieper, Ursula, Benjamin M. Webb, Guang Qiang Dong, Dina Schneidman-Duhovny, Hao Fan, Seung Joong Kim, Natalia Khuri, Yannick G. Spill, Patrick Weinkam, Michal Hammel, John A. Tainer, Michael Nilges, and Andrej Sali. 2014. "ModBase, a Database of Annotated Comparative Protein Structure Models and Associated Resources." *Nucleic Acids Research* 42(D1).

Qiang, W., W. M. Yau, Y. Luo, M. P. Mattson, and R. Tycko. 2012. "Antiparallel β-Sheet Architecture in Iowa-Mutant -Amyloid Fibrils." *Proceedings of the National Academy of Sciences* 109(12):4443–48.

Rodriguez, Jose A., Magdalena I. Ivanova, Michael R. Sawaya, Duilio Cascio, Francis E. Reyes, Dan Shi, Smriti Sangwan, Elizabeth L. Guenther, Lisa M. Johnson, Meng Zhang, Lin Jiang, Mark A. Arbing, Brent L. Nannenga, Johan Hattne, Julian Whitelegge, Aaron S. Brewster, Marc Messerschmidt, Sebastien Boutet, Nicholas K. Sauter, Tamir Gonen, and David S. Eisenberg. 2015. "Structure of the Toxic Core of α-Synuclein from Invisible Crystals." *Nature* 525(7570):486–90.

Rubinshtein, Michael and Ralph H. Colby, Polymer Physics, Oxford University Press, 2003

Schlitter, J., M. Engels, P. KrüGer, E. Jacoby, and A. Wollmer. 1993. "Targeted Molecular Dynamics Simulation of Conformational Change - Application to the T R Transition in Insulin." *Molecular Simulation* 10(2–6):291–308.

Shen, Koning, Barbara Calamini, Jonathan A. Fauerbach, Boxue Ma, Sarah H. Shahmoradian, Ivana L. Serrano Lachapel, Wah Chiu, Donald C. Lo, and Judith Frydman. 2016. "Control of the

Structural Landscape and Neuronal Proteotoxicity of Mutant Huntingtin by Domains Flanking the PolyQ Tract." *ELife* 5(OCTOBER2016).

Shewmaker, Frank, Dmitry Kryndushkin, Bo Chen, Robert Tycko, and Reed B. Wickner. 2009. "Two Prion Variants of Sup35p Have In-Register Parallel β-Sheet Structures, Independent of Hydration." *Biochemistry* 48(23):5074–82.

Sorci, Mirco, Whitney Silkworth, Timothy Gehan, and Georges Belfort. 2011. "Evaluating Nuclei Concentration in Amyloid Fibrillation Reactions Using Back-Calculation Approach." *PLoS ONE* 6(5).

Van Der Spoel, David, Erik Lindahl, Berk Hess, Gerrit Groenhof, Alan E. Mark, and Herman J. C. Berendsen. 2005. "GROMACS: Fast, Flexible, and Free." *Journal of Computational Chemistry* 26(16):1701–18.

Steven, Alasdair, Wolfgang Baumeister, Louise N. Johnson, and Richard N. Perham. 2016. *Molecular Biology of Assemblies and Machines.* Garland Science.

Tozzini, Valentina, Walter Rocchia, and J. Andrew McCammon. 2006. "Mapping All-Atom Models onto One-Bead Coarse-Grained Models: General Properties and Applications to a Minimal Polypeptide Model." *Journal of Chemical Theory and Computation* 2(3):667–73.

Vilar, Marçal, Hui-Ting Chou, Thorsten Lührs, Samir K. Maji, Dominique Riek-Loher, Rene Verel, Gerard Manning, Henning Stahlberg, and Roland Riek. 2008. "The Fold of Alpha-Synuclein Fibrils." *Proceedings of the National Academy of Sciences* 105(25):8637–42.

Waldo, G. S., B. M. Standish, J. Berendzen, and T. C. Terwilliger. 1999. "Rapid Protein-Folding Assay Using Green Fluorescent Protein." *Nature Biotechnology* 17(7):691–95.

Wang, Weiqiang, Susanna Navarro, Rafayel A. Azizyan, Manuel Baño-Polo, Sebastian A. Esperante, Andrey V Kajava, and Salvador Ventura. 2019. "Prion Soft Amyloid Core Driven Self-Assembly of Globular Proteins into Bioactive Nanofibrils." *Nanoscale* 11(26):12680–94.

Wasmer, Christian, Adam Lange, Hélène Van Melckebeke, Ansgar B. Siemer, Roland Riek, and Beat H. Meier. 2008. "Amyloid Fibrils of the HET-s(218-289) Prion Form a β Solenoid with a Triangular Hydrophobic Core." *Science* 319(5869):1523–26.

Wasmer, Christian, Anne Schütz, Antoine Loquet, Carolin Buhtz, Jason Greenwald, Roland Riek, Anja Böckmann, and Beat H. Meier. 2009. "The Molecular Organization of the Fungal Prion HET-s in Its Amyloid Form." *Journal of Molecular Biology* 394(1):119–27.

Wassenaar, Tsjerk A., Kristyna Pluhackova, Rainer A. Böckmann, Siewert J. Marrink, and D. Peter Tieleman. 2014. "Going Backward: A Flexible Geometric Approach to Reverse

Transformation from Coarse Grained to Atomistic Models." *Journal of Chemical Theory and Computation* 10(2):676–90.

Weirich, Franziska, Lothar Gremer, Ewa A. Mirecka, Stephanie Schiefer, Wolfgang Hoyer, and Henrike Heise. 2016. "Structural Characterization of Fibrils from Recombinant Human Islet Amyloid Polypeptide by Solid-State NMR: The Central FGAILS Segment Is Part of the β-Sheet Core." *PLoS ONE* 11(9).

Wurth, Christine, Nathalie K. Guimard, and Michael H. Hecht. 2002. "Mutations That Reduce Aggregation of the Alzheimer°Fs Aβ42 Peptide: An Unbiased Search for the Sequence Determinants of Aβ Amyloidogenesis." *Journal of Molecular Biology* 319(5):1279–90.

Xue, Wei-Feng, Steve W. Homans, and Sheena E. Radford. 2008. "Systematic Analysis of Nucleation-Dependent Polymerization Reveals New Insights into the Mechanism of Amyloid Self-Assembly." *Proceedings of the National Academy of Sciences* 105(26):8926–31.

Yang, Fan, Larry G. Moss, and George N. Phillips. 1996. "The Molecular Structure of Green Fluorescent Protein." *Nature Biotechnology* 14(10):1246–51.

Supporting information for:

Amyloidogenicity as a driving force for the formation of functional oligomers

*Rafayel A. Azizyan[1,2], Weiqiang Wang[3] Alexey Anikeenko[4,] Zinaida Radkova[4], Anastasia Bakulina[4], Adriana Garro[5], Landry Charlier[6], Christian Dumas[7], Salvador Ventura[3] and Andrey V. Kajava[1,2]**

1 Centre de Recherche en Biologie cellulaire de Montpellier, UMR 5237 CNRS, Université Montpellier, Montpellier, France.

2 Institut de Biologie Computationnelle, Université Montpellier, Montpellier, France.

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

4 Novosibirsk State University, Novosibirsk, Russia.

5 Universidad Nacional de San Luis IMASL-CONICET, San Luis, Argentina.

6 Institut des Biomolécules Max Mousseron, Montpellier, France.

7 Centre de Biochimie Structurale, CNRS, UMR5048, INSERM U1054, Université de Montpellier, Montpellier, France.

*Correspondence and requests should be addressed to A.V.K. (email: andrey.kajava@crbm.cnrs.fr)

Table S1. Number of the Rigid Body simulation runs out of 100 for given linker size resulted in a given number of the monomers

number of monomers	linker_1	linker_2	linker_3	linker_4	linker_5	Linker_6
2	100	100	98	100	98	100
3	0	100	100	95	92	95
4		100	96	50	95	85
5		0	100	96	95	87
6			100	99	94	90
7			100	100	98	88
8			0	97	60	93
9				99	100	92
10				98	94	49
11				100	96	87
12				23	76	90
13				93	98	94
14				74	98	90
15				11	100	93
16				0	99	81
17					96	24
18					22	72
19					100	97
20					99	45
21					99	90
22					85	15
23					94	95
24					76	15
25					36	94
26					10	95
27					26	94
28					17	98
29					1	29
30					0	99
31						82
32						45
33						82
34						66
35						43
36						8
37						2
38						66
39						10
40						97
41						6
42						94
43						95
44						20
45						89
46						86

47						64
48						15
49						44
50						7
51						59
52						30
53						56
54						1
55						45
56						17
57						21
58						15
59						38
60						20
61						4
62						1
63						6
64						4
65						2
66						10
67						1
68						7
69						7
70						0

Methods

CG-TMD simulations

The CG-TMD simulations were performed using GROMACS version 4.6.7 package,[1,2] with the MARTINI force field.[3] We used the MARTINI CG water model with one CG bead corresponding to four water molecules. During the simulations, periodic boundary conditions were used with a 200 Å cubic periodic box. Each monomer addition took 0.2 microseconds with 20 fs integration time. At each integration time step of 20 fs, the RMS distance between the coordinates of the current and target structures was computed, and the force on each pseudo-atom is given by the gradient of the potential:

$$U_{TMD}=\frac{1}{2}\frac{k}{N}[\text{RMS (t)-RMS*(t)}]^2$$

where RMS(t) is the time-dependent instantaneous best-fit RMS distance of current coordinates from the target coordinates, and RMS*(t) evolves linearly from RMSD at the first to the last TMD steps. The spring constant k = 1000 kJ/mol/nm^2 is scaled down by the number of N targeted pseudo-atoms.

A cutoff radius of 1.2 nm was used in the calculation of nonbonded interactions with a shifted function. The Lennard-Jones potential was shifted from 0.9 to 1.2 nm, whereas electrostatic potential was shifted from 0 to 1.2 nm (distance between pseudo-atoms). Both the energy and the force were switched to zero at the cutoff distance. A leap-frog algorithm for integrating Newton's equations of motion was presented for the general case involving constraints with coupling to both a constant temperature and a constant pressure bath. Berendsen temperature coupling algorithm[4] was used to maintain a constant temperature at 300 K with a coupling time constant of 1 ps. Isotropic pressure coupling was applied using the Berendsen algorithm[4] with a reference pressure of 1 bar. A coupling constant of 5.0 ps and a compressibility of 4×10^{-5} bar^{-1} were used. All covalent bond lengths were constrained using the LINCS algorithm[2].

To model the formation of amyloids, we applied several constraints. The Aβ part, on its way to the target structure is free of any constraints, whereas in the final structure of growing oligomer these parts are positionally restrained. Linkers between Aβ region

and GFP domain have no restraints. ELNEDYN model was used as a structural scaffold to maintain the overall shape of the GFP globular domains[5].

The position of globular domains and linkers are not taken into account in the process of defining the target β-arcade structure. The only condition was to preserve the same amount of atoms or CG pseudo-atoms in the initial and target structure. Afterwards, we placed two initial monomers in the periodic box. Aβ-part of the one monomer was kept unmoving within the target structure in the center of simulation box, and the other monomer was moved by steering force to the adjacent β-arch position to stack two β-arches in the β-arcade target. This cycle of defining the target structure and moving the next monomer to target has been repeated several times until the oligomer structure started to have energetically unfavorable steric tension originated from the globular domains. It should be mentioned that the initial position of moving monomer was changed during every new run of MD simulations. Thus, we tested if the results depend on the initial position of the moving monomers.

Evaluation of the quality of oligomer structures obtained

To evaluate the quality of oligomer structures obtained we have converted the resulting CG models back to AA models. In our study, the "Going backward" [6] approach was used to convert the models. It should be mentioned that the atomic models derived from the CG models were imperfect because, after the series of MD procedures included in the protocol of Backward tool, we obtained the positional deviations in the target β–arcade structure of Aβ. To overcome this problem we superimposed the target amyloid structure of Aβ on the converted AA models, using Cα atoms of the backbone. After the superposition, the Aβ-arcade was substituted by the target structure. The same procedure has been applied also to each globular domain of GFP. Afterwards, using ModLoop tool[7,8] we have connected globular domains and corresponding Aβ molecules by the corresponding linkers. Then, we removed bad contacts between the monomers by energy minimization procedure using steepest descent algorithm of GROMACS package version 4.6.7[1,2]. Finally, by using PROCHECK package[9] we evaluated a quality score for the of overall stereochemistry of oligomer structures. The validation of models has been done based on G-factor values. The G-factors provide a measure of

how unusual, or out-of-the-ordinary, a property is, with values below -0.5 being unusual, and values below -1.0 being highly unusual[9].

Rigid body simulation

We used Blender software package[10] for the Rigid Body Simulation. Models were generated by a Python script in Blender. For additional information see https://cg3dartist.wixsite.com/amyloid. It sets the number of monomers in the oligomer, the number of residues represented by small spheres of 3.5 units in the linker, and the number of independent runs. The globular part of the molecule was represented as a sphere with a diameter of 30 units. The geometrical parameters and Rigid Body World setting of the Aβ-part were the same as in the previous study[11]. Each run takes 0.2 second of Rigid Body Simulation for an oligomer of two subunits and 30 seconds for an oligomer of 70 subunits by using a personal desktop computer. Overlaps between elements of the model were automatically detected. The run resulted in the structure without overlaps was considered as successful. For each linker length, the number of successful runs was recorded. To vary the initial conditions of the runs, random value of the gravity force from 0 to 1 along the fibril axis was set for each run. Visual inspection of the final models demonstrated that the final structures are slightly different in the number of monomers and spatial arrangement of globular parts around the amyloid core. We analyzed oligomers with the numbers of residues in the linker from 2 to 6 and the number of monomers in the models from 2 to more than 50.

REFERENCES

(1) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, Flexible, and Free. *Journal of Computational Chemistry*. 2005, pp 1701–1718.

(2) Hess, B.; Kutzner, C.; Van Der Spoel, D.; Lindahl, E. GRGMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* 2008, *4* (3), 435–447.

(3) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; De Vries, A. H. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *J. Phys. Chem. B* 2007, *111* (27), 7812–7824.

(4) Berendsen, H. J. C.; Postma, J. P. M.; Van Gunsteren, W. F.; Dinola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* 1984, *81* (8), 3684–3690.

(5) Periole, X.; Cavalli, M.; Marrink, S. J.; Ceruso, M. A. Combining an Elastic Network With a Coarse-Grained Molecular Force Field: Structure, Dynamics, and Intermolecular Recognition. *J. Chem. Theory Comput.* 2009, *5* (9), 2531–2543.

(6) Wassenaar, T. A.; Pluhackova, K.; Böckmann, R. A.; Marrink, S. J.; Tieleman, D. P. Going Backward: A Flexible Geometric Approach to Reverse Transformation from Coarse Grained to Atomistic Models. *J. Chem. Theory Comput.* 2014, *10* (2), 676–690.

(7) Fiser, A.; Do, R. K.; Sali, A. Modeling of Loops in Protein Structures. *Protein Sci* 2000, *9* (9), 1753–1773.

(8) Fiser, A.; Sali, A. ModLoop: Automated Modeling of Loops in Protein Structures. *Bioinformatics* 2003, *19* (18), 2500–2501.

(9) Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. PROCHECK: A Program to Check the Stereochemical Quality of Protein Structures. *J. Appl. Crystallogr.* 1993, *26* (2), 283–291.

(10) blender.org - Home of the Blender project - Free and Open 3D Creation Software https://www.blender.org/.

(11) Azizyan, R. A.; Garro, A.; Radkova, Z.; Anikeenko, A.; Bakulina, A.; Dumas, C.; Kajava, A. V. Establishment of Constraints on Amyloid Formation Imposed by Steric Exclusion of Globular Domains. *J. Mol. Biol.* 2018, *430* (20), 3835–3846.

Chapter III

Multifunctional amyloid oligomeric nanoparticles for specific cell targeting and drug delivery

Weiqiang Wang[1], Rafayel A. Azizyan[2,3], Adriana Garro[4], Andrey V. Kajava[2,3] and Salvador Ventura[1,]*

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

2 Centre de Recherche en Biologie cellulaire de Montpellier, UMR 5237 CNRS, Université Montpellier, Montpellier, France.

3 Institut de Biologie Computationnelle, Université Montpellier, Montpellier, France.

4 Universidad Nacional de San Luis IMASL-CONICET, San Luis, Argentina.

Correspondence to: salvador.ventura@uab.es

Abstract

Natural selection has endorsed proteins with amazing structures and functionalities that cannot be matched by synthetic means, explaining the exponential interest in developing protein-based materials. Protein self-assembly allows fabricating complex supramolecular structures from relatively simple building blocks, a bottom-up strategy naturally employed by amyloid fibrils. However, the design of amyloid inspired materials with biological activity is inherently difficult. Here we exploit a modular procedure to generate functional amyloid nanostructures with tight control of their mesoscopic properties. This enabled us to generate, for the first time, biocompatible protein-only amyloid-like oligomeric nanoparticles with defined dimensions in which embedded globular proteins remain highly active. The modular design allowed us to obtain multifunctional nanoparticles. We show here how this property can be exploited for antibody-directed targeting of specific cell types and the localized delivery of methotrexate, resulting in the intracellular uptake of the drug by cancer cells and their death. Overall, the novel protein particles we describe in this work might find applications in areas as diverse as biocatalysis, bioimaging, or targeted therapies.

Keywords: protein self-assembly, amyloid fibrils, oligomers, protein-based nanoparticles, drug delivery, methotrexate, antibodies.

Introduction

The design and production of biologically inspired assemblies that can be exploited for the fabrication of functional materials is a rapidly growing area of research.[1] Amyloid fibrils have been traditionally associated with human diseases.[2] However, the recent discovery of amyloids assisting biological functions in a wide range of organisms, from bacteria to humans,[3] has changed this perception and inspired the use of their unique architecture to build up nanostructured materials for applications in biomedicine and biotechnology.[4]

A majority of the available amyloid-based functional nanomaterials result from the self-assembly of short synthetic peptides of natural or artificial origin.[5] In contrast, the assembly of globular proteins into amyloid-like materials with biological activity has been less successful. This is expected since the aggregation of globular domains into amyloid structures necessarily implies a process of conformational conversion in which they lose their native fold and thus their activity.[6]

The main advantage of protein-based materials is the nature of globular proteins, which allows to alter the material functionality by genetic redesign to fit the intended application. We have recently succeeded in designing highly ordered amyloid-like nanofibrils containing properly folded and highly active proteins using a modular strategy.[7] The distinctive feature of the approach consists in the use of a Soft Amyloid Core (SAC) as the assembling unit. SACs are short amyloidogenic sequence stretches initially identified in the disordered and low complexity domains of yeast prions.[8] They differ from the classical amyloid cores of pathogenic proteins in that they are slightly longer and more polar.[9,10] This results in a weaker and more diffuse amyloid propensity, but still sufficient to nucleate the self-assembly reaction. The SAC is fused through a Gly/Ser soluble and flexible linker to the globular protein of interest. The resulting fusion proteins are produced recombinantly at high yield and in a soluble manner; still, their ordered aggregation can be induced under defined conditions, rendering functional synthetic amyloid fibrils in which the SAC forms the amyloid spine and the globular

domains hang from it in an active conformation, as demonstrated for fluorescent proteins and enzymes. [7]

We use the SAC (residues 100-118) of the Sup35 yeast prion (Sup35-SAC) as the default assembling module. Its small size allows to use a connecting linker of only eight residues and obtain amyloid fibrils in which the fused protein is active without major steric impediments and with a reduced entropy penalization for the assembly reaction. This is in contrast with natural amyloid proteins displaying similar architecture, like the prion proteins Het-s,[11] Sup35p and Ure2p,[12] human Rip1/Rip3,[13] or the TAR-DNA binding proteins,[14] where the connecting linker is significantly longer (> 50 residues).

Recently, by using molecular and mesoscopic modeling, we were able to establish a theoretical relationship between the size of a globular domain and the length of the shortest linker that still allows the formation of infinite amyloid fibrils when fused to an amyloidogenic sequence.[15] The model suggested that shortening the linker below this size limit would result in the formation oligomers of defined size instead of the typical fibrils.[16] This might provide a unique opportunity to generate small protein-based nanoparticles decorated with *a la carte* globular protein functionalities.

Here we connected Sup35-SAC to the dihydrofolate reductase (DHFR) enzyme[17] (Sup35-DHFR**)** and demonstrate that we can control the mesoscopic properties of the resulting assemblies just by playing with the linker length, obtaining either catalytic fibrils or catalytic oligomeric nanoparticles. These nanoparticles are spherical assemblies of an amyloid-like nature but devoid of any toxicity. They are homogenous in size, stable, and more active that the correspondent fibrils.

We explored the potential of these novel protein-based nanoparticles to generate functionalized materials by creating a tripartite fusion in which we incorporated the Z domain,[18] an engineered analog of the B domain of *Staphylococcus aureus* protein A[19] to the C-terminus of the Sup35-DHFR protein. The new fusion protein also assembles into spherical oligomeric nanoparticles that can now be decorated with any antibody of

interest, thanks to the Z-domain's nanomolar affinity for antibodies' Fc domains, thus generating multivalent particles.

We show that antibody loaded nanoparticles can be directed specifically against cancer cells and demonstrate that they can be used as nanocarriers for cell-specific delivery of methotrexate (MTX). MTX acts as an antagonist of folic acid, which is necessary for DNA synthesis and accordingly is used as a chemotherapeutic agent for the treatment of cancer.[20,21] However, it displays a weak pharmacokinetic profile, and significant off-target toxicity[22,23,24] and MTX based nanomedicines have been designed to overcome these drawbacks.[25] In our nanoparticles, the Z-domain bound antibody acts as a cellular selector and the DHFR moiety as MTX carrier. Once released from its carrier, the drug is internalized, displaying a potent and localized cytotoxic effect. The strategy is simple and modular and can be adapted to target any cell type of interest.

Results and Discussion

Self-assembly of Sup35-8aa-DHFR into active amyloid fibrils

In order to generate functional amyloid fibrils with a previously unexplored catalytic activity, we fused Sup35-SAC to the *Escherichia coli* DHFR through an 8-residues long flexible linker consisting of SG_3SG_2S (Sup35-8aa-DHFR) (**Figure S1A** and **Figure S1B**). The size of DHFR (21.5 kDa) is in the range of GFP (27 kDa), and carbonic anhydrase (30 kDa), two proteins that assemble into active amyloid fibrils when fused to Sup35-SAC.[7]

The fusion protein, Sup35-8aa-DHFR, was expressed recombinantly at 64 mg/L and located in the soluble cellular fraction, from which we purified it (**Figure S1C**). We compared the conformational properties of soluble wild type DHFR (DHFR-wt) and Sup35-8aa-DHFR proteins by far-UV circular dichroism (CD) (**Figure 1A**). Their spectra were virtually identical, indicating that Sup35-SAC does not induce significant changes in the DHFR structure. Then, we used the amyloid-specific dyes Thioflavin-T (Th-T) and Congo Red (CR) to assess if, despite its solubility, Sup35-8aa-DHFR self-

assembles with time into amyloid-like structures. To this aim, we incubated the fusion protein in PBS buffer pH 7.4 at 37 ºC for four days. The incubated Sup35-8aa-DHFR protein promoted a substantial increase in the Th-T fluorescence emission signal, whereas DHFR-wt incubated in the same conditions had a negligible effect (**Figure 1B**). In agreement with these results, CR binding was observed for incubated Sup35-8aa-DHFR, resulting in a red-shift of the CR absorption spectrum, while DHFR-wt did not promote any CR spectral change (**Figure 1C**). Moreover, the Fourier Transform InfraRed (FTIR) absorbance spectrum of incubated Sup35-8aa-DHFR in the amide I region evidenced the existence of a band at 1621 cm^{-1}, which can be assigned to the inter-molecular β-sheet structure characteristic of amyloids (**Figure 1D and Table S1**). This band was absent in the FTIR spectrum of incubated DHFR-wt (**Figure S3 and Table S1**). The morphological analysis of the two protein solutions by negative-staining and Transmission Electron Microscopy (TEM) indicated that the incubated fusion protein assembled into a typical fibrillar amyloid-like structure, whereas DHFR-wt did not form any kind of assembly (**Figure 1E and F**).

To test if the DHFR enzyme keeps the native structure within the observed amyloid fibrils, we used a fluorescein labelled version of MTX (fMTX), a competitive inhibitor that binds to the active site of DHFR. In the presence of the inhibitor, the amyloid fibrils appeared green, as visualized by fluorescence microscopy (**Figure 1G)**, indicating that fMTX can bind to the catalytic site of the DHFR moieties embedded in these nanostructures.

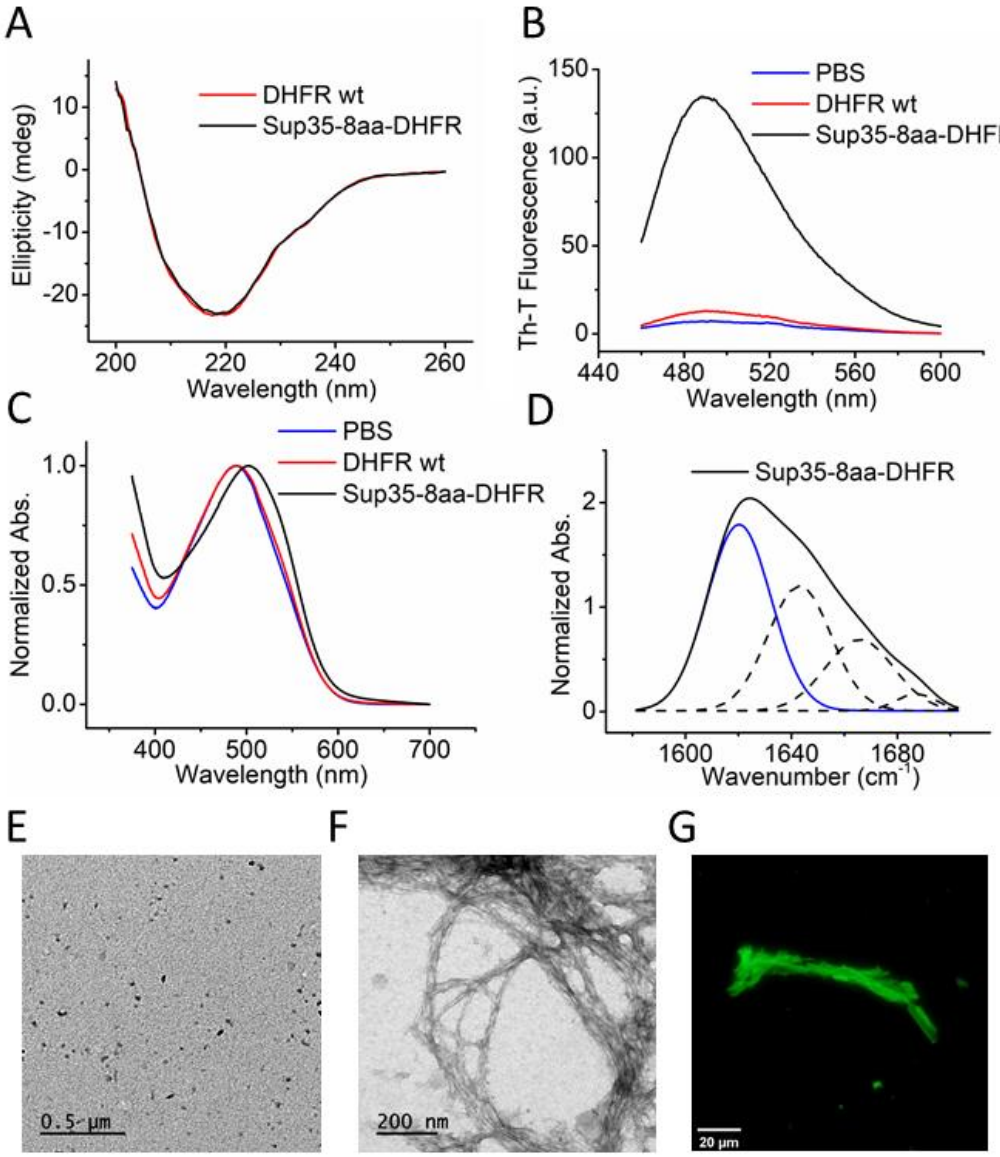

Figure 1. Characterization of soluble and aggregated Sup35-8aa-DHFR. (A) Far-UV CD spectra of soluble Sup35-8aa-DHFR and DHFR-wt, shown in black and red, respectively. Incubated Sup35-8aa-DHFR was analyzed by measuring Th-T fluorescence emission (B) and Congo red absorbance (C) DHFR-wt and Sup35-8aa-DHFR are shown in red and black, respectively. PBS without protein was included as a control (blue line). (D) FTIR absorbance spectra of incubated Sup35-8aa-DHFR in the amide I region of the specturm (solid black line) and the components bands (dashed lines), the intermolecular β-sheet component is shown in blue. (E and F) Representative TEM micrographs of incubated proteins upon negative staining: (E) DHFR-wt and (F) Sup35-8aa-

DHFR. Scale bar represents 500 and 200 nm, respectively. (G) Fluorescence microscopy image of Sup35-8aa-DHFR fibrils incubated with fMTX. Scale bar represents 20 μm.

Self-assembly of Sup35-5aa-DHFR into amyloid oligomeric nanoparticles

As shown above, a linker of 8 residues suffices to allow the formation of infinite fibrils in which DHFR remains in its native conformation (**Figure 2**). Our recently derived model predicts that introducing a shorter linker to connect the amyloidogenic and globular moieties would generate steric hindrance and might result in the formation of oligomeric structures, instead of fibrils (**Figure 2**).[16] To experimentally demonstrate this hypothesis, we constructed a Sup35-5aa-DHFR fusion, in which we connected Sup35-SAC and DHFR with a 5-residues linker consisting of SGSGS (**Figure S2A and B**).

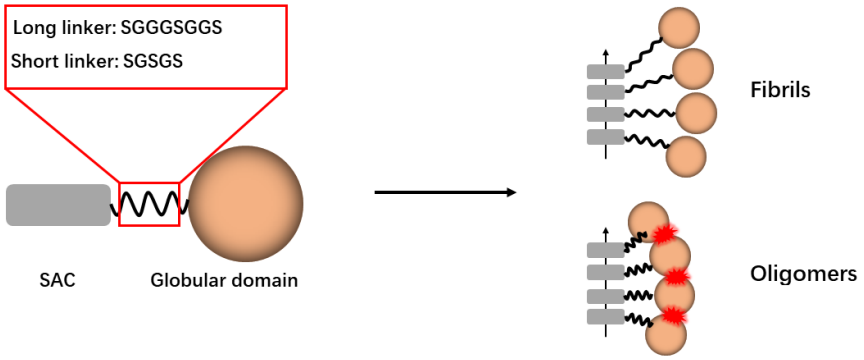

Figure 2. Schematic illustration of the rational for the formation of stable amyloid oligomers. Scheme of SAC (grey box) connected with a globular domain (orange ball) by linkers of different lengths. The molecules with long linkers can form infinite amyloid fibrils, whereas those with short linkers would form preferably oligomers, owing to the steric restriction imposed by adjacent globular domains.

Again, the protein was well-expressed (71 mg/L) and soluble (**Figure S2C**). Purified Sup35-5aa-DHFR, displayed a conformation very similar to that of DHFR-wt in solution, as assessed by far-UV CD (**Figure 3A**). The protein was incubated at 37 ºC for four days, rendering Th-T (**Figure 3B**) and CR (**Figure 3C**) positive assemblies,

with detectable inter-molecular β-sheet content (**Figure 3D and Table S1**). TEM imaging evidenced that, despite these amyloid-like properties, and in contrast to Sup35-8aa-DHFR, Sup35-5aa-DHFR assembles into spherical nanoparticles, which is consistent with an amyloid oligomeric nature (**Figure 3E and 3F**). The presence of spherical structures in the incubated Sup35-5aa-DHFR solution was further confirmed by scanning electron microscopy (SEM) (**Figure 3G**). The particles were very homogenous in size, with a diameter of 147±20 nm and 159±24 nm as measured by TEM and SEM, respectively. These sizes are in good agreement with the 150±50 nm measured for the particles in solution by dynamic light scattering (DLS). Incubation of this sample with fMTX rendered green fluorescent aggregates (**Figure 3H**), indicating that, as in the amyloid fibrils, the inhibitor can bind to folded DHFR domains in the structure of the oligomeric protein nanoparticles. Therefore, as predicted, playing with the linker length allowed us to tune the assemblies' mesoscopic properties and produce, for the first time, amyloid-like active spherical nanoparticles.

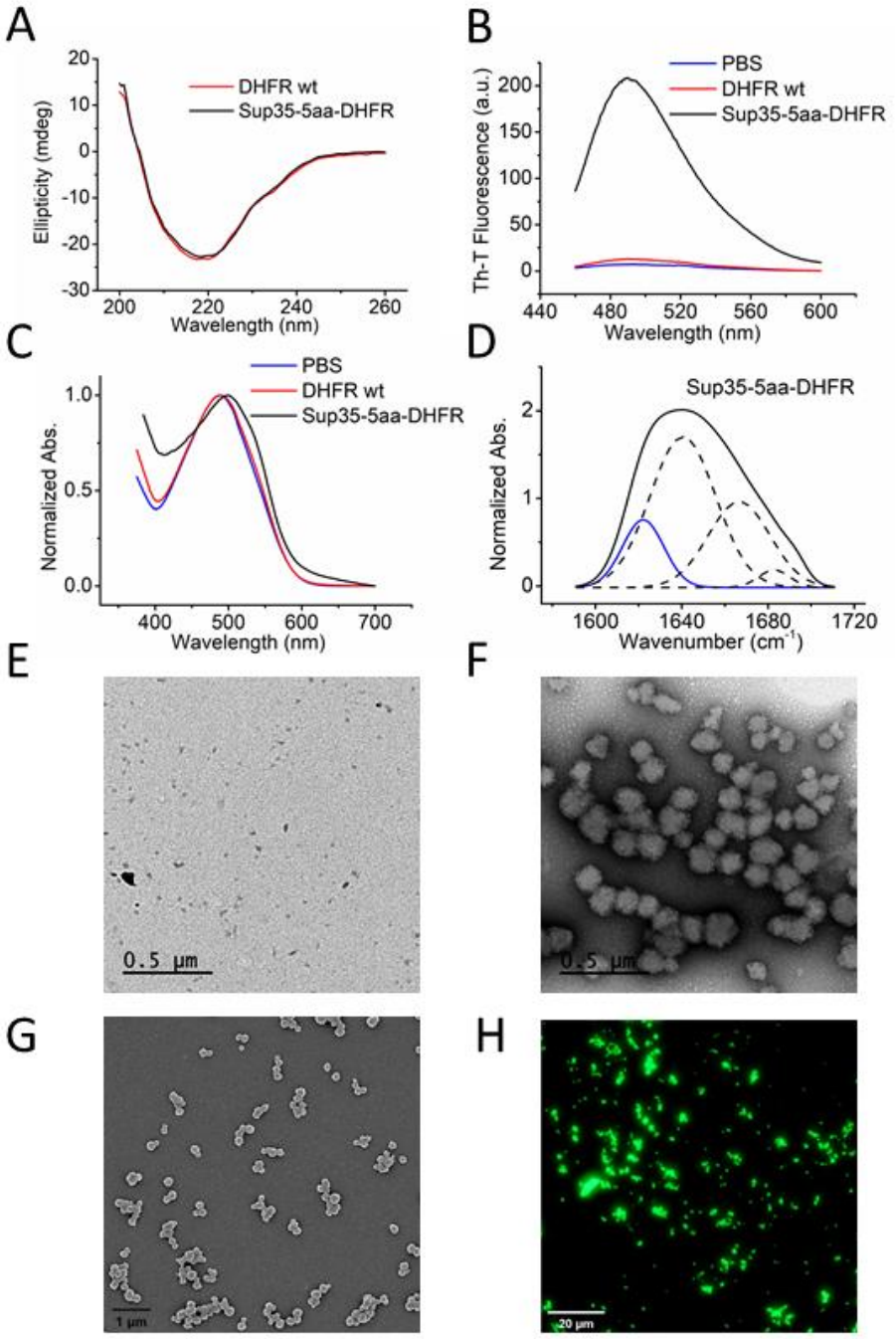

Figure 3. Characterization of soluble and aggregated Sup35-5aa-DHFR. (A) Far-UV CD spectra of soluble Sup35-5aa-DHFR and DHFR-wt, shown in black and red, respectively. Incubated Sup35-5aa-DHFR was analyzed by measuring Th-T fluorescence emission (B) and CR absorbance (C) DHFR-wt and Sup35-5aa-DHFR are shown in red and black, respectively. PBS without protein was included as a control (blue line). (D) FTIR absorbance spectra of incubated Sup35-8aa-DHFR in the amide I region of the spectrum (solid black line) and the components bands (dashed lines),

the inter-molecular β-sheet component is shown in blue. (E and F) Representative TEM micrographs of incubated proteins upon negative staining: (E) DHFR-wt and (F), Sup35-5aa-DHFR. Scale bar represents 500 nm. (G) Representative SEM micrograph of Sup35-5aa-DHFR particles. Scale bar represents 1 μm. (H) Fluorescence microscopy image of Sup35-5aa-DHFR particles incubated with fMTX. Scale bar represents 20 μm.

Catalytic activity of DHFR in fibrils and nanoparticles

DHFR catalyzes the reduction of dihydrofolate (DHF) to tetrahydrofolate (THF) in the presence of reduced nicotinamide adenine dinucleotide phosphate (NADPH).[26] The binding of fMTX to fibrils and nanoparticles suggested that DHFR might keep its enzymatic activity in both kinds of assemblies. To further confirm this extent and evaluate the effect of each superstructure on the activity of the enzyme, we measured the catalytic activity of fibrils and nanoparticles at 100 nM in the presence of the DHF and NADPH. We recorded the absorbance change at 340 nm for 20 min and plotted the resulting traces (**Figure 4A**). For both fibrils and nanoparticles, the decrease of NADPH absorbance with time indicates that they are catalytically active, whereas the fibrils formed by the Sup35-SAC peptide alone or lysozyme, used as negative controls, do not affect the NADPH absorbance at 340 nm. Importantly, the signal decreased faster for the nanoparticles than for the fibrils, which suggested that the smaller assemblies were more active.

Next, we sought to determine the apparent kinetic constants for both amyloid-like assemblies. We measured the initial velocities of the reactions in the presence of NADPH in a range of concentrations from 10-200 μM, while the concentration of DHF was fixed at 20 μM. Then, we calculated the kinetic constants from the resulting Lineweaver-Burk plots (**Figure 4B and Table S2**). The nanoparticles' V_{max} was 1712±25 nM min^{-1}, which does not differ significantly from that of the fibrils (V_{max} = 1745±21 nM min^{-1}). This was expected since the NADPH binding site and the catalytic centre of DHFR should keep the same conformation in both assemblies. However, the nanoparticles exhibited a significantly lower K_m (6.23±0.32 μM) than that of the fibrils

(9.44±0.41 μM). This translates into a higher specificity constant for the nanoparticles (K_{cat}/K_m = 0.23±0.01 μM^{-1} s^{-1}) (**Table S2)**.

The thermodynamic dissociation constants (K_d) for binding of NADPH to both assemblies were measured by monitoring the quenching of the DHFR intrinsic fluorescence (**Figure 4C**). The obtained values indicated that the binding affinity of nanoparticles for NADPH (K_d = 1.58±0.07 μM) was 4-fold higher than that of the fibrils, with a K_d value of 6.26±0.44 μM (**Table S2**). Thus, both the K_m and K_d constants indicate that the higher catalytic activity of the nanoparticles results from a higher NADPH binding capability of DHFR when embedded in the spherical amyloid-like structures, likely because the cofactor binding site is more accessible in these assemblies or, alternatively, because the proportion of conformationally active DHFR molecules is higher in the particles than in the fibrils.

Overall, the activity, homogeneity, stability, spherical structure and moderate size of the novel oligomeric amyloid-like nanoparticles we describe here seem optimal for their exploitation as functionalized nanomaterials.

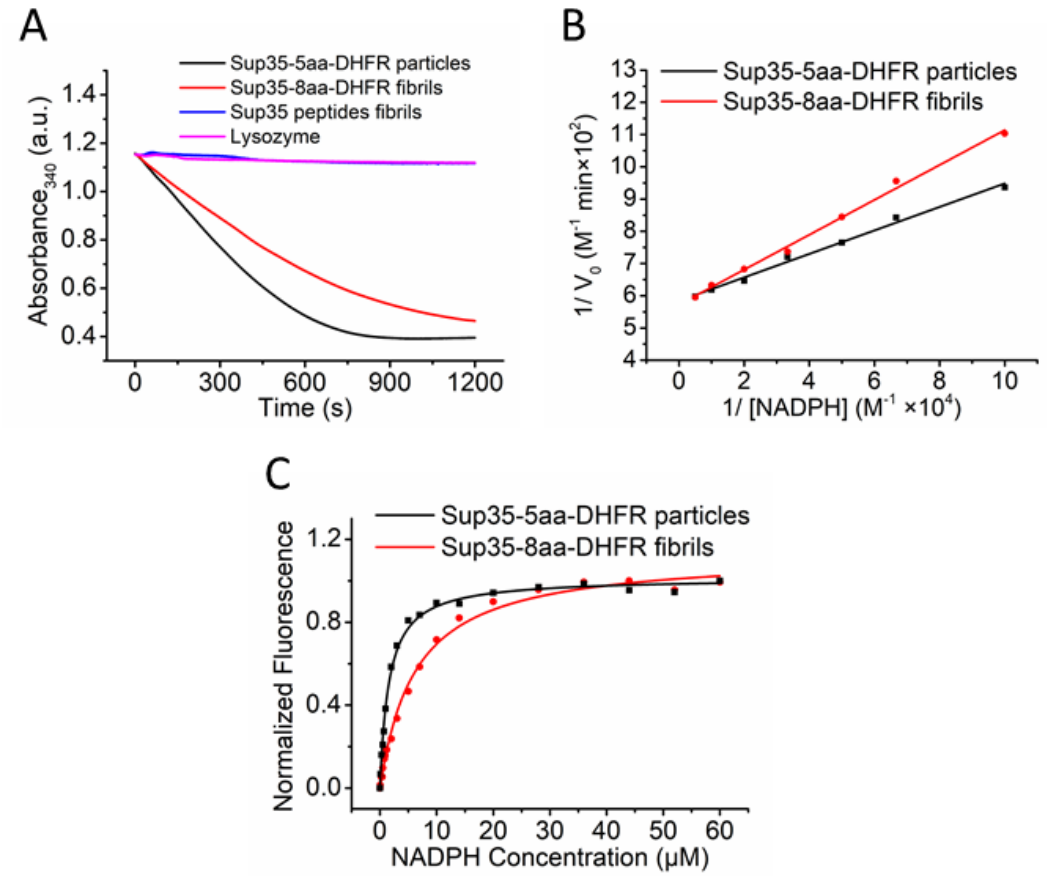

Figure 4. Catalytic activity and NADPH binding affinity of Sup35-5aa-DHFR particles and Sup35-8aa-DHFR fibrils. (A) The change of the absorbance at 340 nm was monitored for each sample for 20 min in the presence of DHF and NADPH. Both assemblies were prepared at 100 nM. Sup35-SAC peptide fibrils and lysozyme are negative controls. (B) Lineweaver-Burk plot obtained by plotting the reciprocal of initial velocities of reactions against the reciprocal of the concentration of NADPH. (C) Measurement of K_d for NADPH binding to both assemblies by monitoring the quenching of DHFR intrinsic fluorescence. Fluorescence titrations were performed using 100 nM of assemblies and the indicated concentrations of NADPH.

Building up Sup35-5aa-DHFR-Z bifunctional nanoparticles

We sought to exploit the ability of Sup35-5aa-DHFR to form spherical structures of defined size to build up bifunctional nanoparticles with potential applications in targeted delivery. To this aim, we incorporated the Z domain of *Staphylococcus aureus*

protein A (SpA) at the C-terminus of Sup35-5aa-DHFR through a flexible SG_2SG linker, to obtain the Sup35-5aa-DHFR-Z tripartite fusion protein (**Figure S4A and B**). The Z-domain is an engineered analog of the B-domain, one of the five homologous IgG-binding domains of SpA[27]. It consists of 58-residues (6.5 kDa) and folds into a bundle-like composed of three α-helices. The Z-domain binds with high affinity to the Fc region of antibodies from different species and subclasses. The idea was that if Sup35-5aa-DHFR-Z assembles into amyloid nanoparticles, we might decorate them with any antibody of interest.

Once more, the protein was well expressed (80 mg/L) and soluble (**Figure S4C**). Once purified, we incubated the fusion protein at 37 ºC for four days and assessed whether it binds to Th-T and CR. Again, the increase in Th-T fluorescence emission (**Figure 5A**) and the red-shift of CR absorbance (**Figure 5B**) were indicative of an amyloid-like structure. The band at 1620 cm^{-1} in the deconvoluted FTIR absorbance spectra confirmed the presence of intermolecular β-sheet structure (**Figure 5E and Table S1**). Finally, TEM images showed that Sup35-5aa-DHFR-Z self-assembled into spherical particles (**Figure 5D**), whereas the soluble, non-incubated, tandem fusion protein did not form any assembly (**Figure 5C**). The size of such oligomers, as measured by TEM, was 43.6±2.8 nm, in good agreement with the 40.1±5.8 nm measured by DLS. Higher magnification TEM images showed the high homogeneity and regular round shape of these nanoparticles (**Figure S5**). The size of these nanostructures is ~ 1/3 of that formed by Sup35-5aa-DHFR. This reduction in size can be univocally attributed to the presence of the Z-domain. Sup35-5aa-DHFR-Z is ~ 6 kDa larger than Sup35-5aa-DHFR; thus, if we assume that the two nanoparticles display similar compactness, this necessarily implies that the number of molecules in Sup35-5aa-DHFR-Z particles is significantly smaller. This reduction in size upon increasing the dimension of the adjacent domain was already predicted by the above-described relationship between AR and adjacent globular domains.[15] The larger the molecule appended to the AR, the lower the number of units that can be incorporated into the amyloid-like oligomers and the smaller the size of the resulting particle.

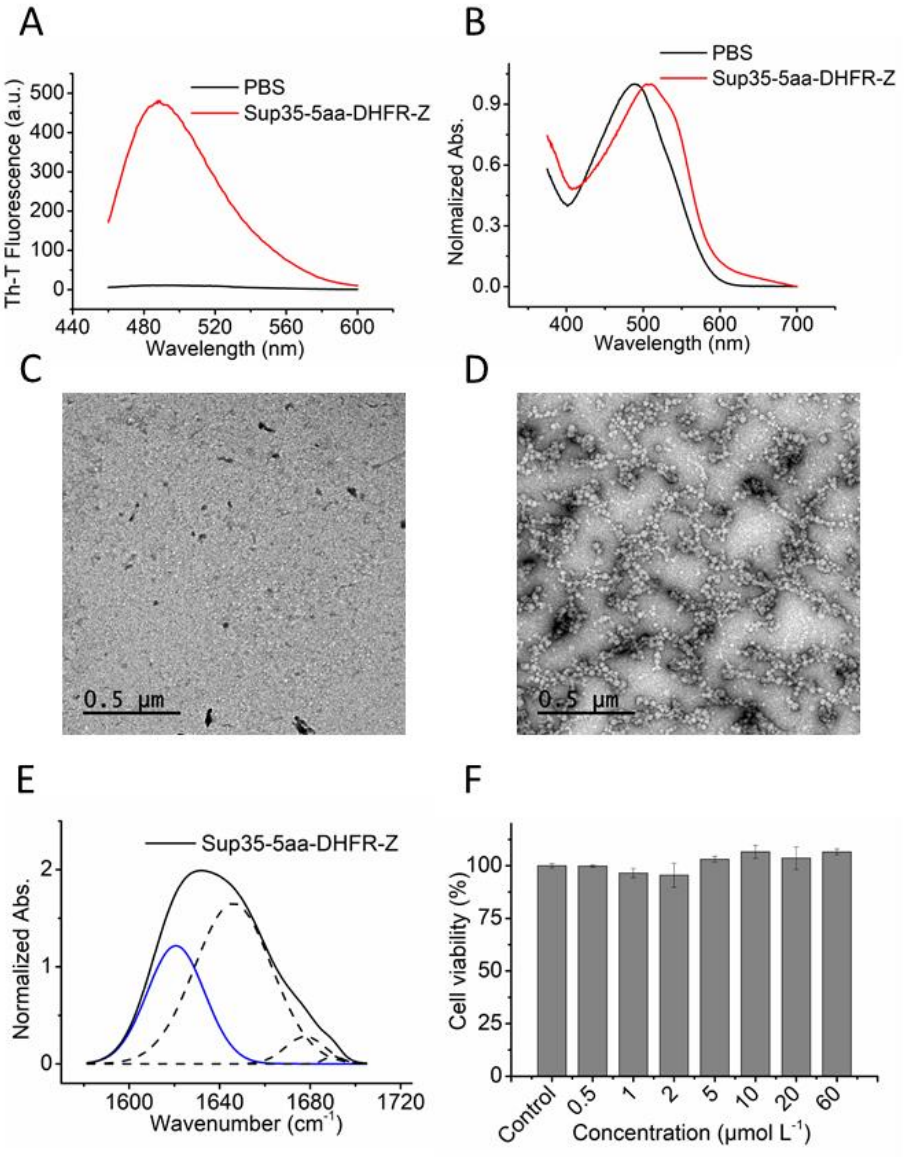

Figure 5. Characterization of aggregated Sup35-5aa-DHFR-Z. Incubated Sup35-5aa-DHFR-Z was analyzed by measuring Th-T fluorescence emission (A) and Congo red absorbance (B). Sup35-5aa-DHFR-Z are shown in red, while PBS without protein was included as a control (black line). Representative TEM micrographs of soluble Sup35-5aa-DHFR-Z **(**C**)** and incubated Sup35-5aa-DHFR-Z (D). Scale bar represents 500 nm. (E) FTIR absorbance spectra of incubated Sup35-5aa-DHFR-Z in the amide I region of the spectra (solid black line) and the components bands (dashed lines), the intermolecular β-sheet component is shown in blue. (F) Cytotoxicity of incubated Sup35-5aa-DHFR-Z. Results are expressed as means ± SD, n=3, and analyzed using a one-way ANOVA test. The statistical differences between the control group and the test group were established at $P < 0.05$.

Sup35-5aa-DHFR-Z bifunctional nanoparticles are biocompatible

One of the main limitations of using amyloid-like materials in biomedical applications is that they might possess cytotoxic activity.[28] To discard this possibility, we tested the cytotoxicity of the Sup35-5aa-DHFR-Z nanoparticles at different concentrations, ranging from 0.5 μM to 60 μM, using the PrestoBlue assay (**Figure 5F**). The statistical analysis using a one-way ANOVA test indicated that the particles did not exhibit significant toxicity for human HeLa cells, suggesting that they would have excellent biocompatibility.

Dual-binding activity of Sup35-5aa-DHFR-Z nanoparticles

To confirm that both DHFR and the Z domain keep their native structure and functionality in Sup35-5aa-DHFR-Z nanoparticles, we incubated them with fMTX (green fluorescence) and an anti-EGFR antibody labeled with Alexa fluor 555 (red fluorescence). Then they were precipitated and washed three times to remove any unbound molecule and resuspended in PBS buffer. When imaged using fluorescence microscopy, the particles appeared green and red in the respective channels, and the two signals overlapped when the channels were merged. (**Figure 6**). This indicated that the DHFR and the Z-domain embedded in the spherical nanoparticles were bound to fMTX and the antibody, respectively. In contrast, the fibrils formed by the SAC-Sup35 peptide alone did not bind any of the two molecules, confirming that the observed binding to Sup35-5aa-DHFR-Z nanoparticles does not result from an unspecific interaction of the reagents with the common amyloid-like structure.

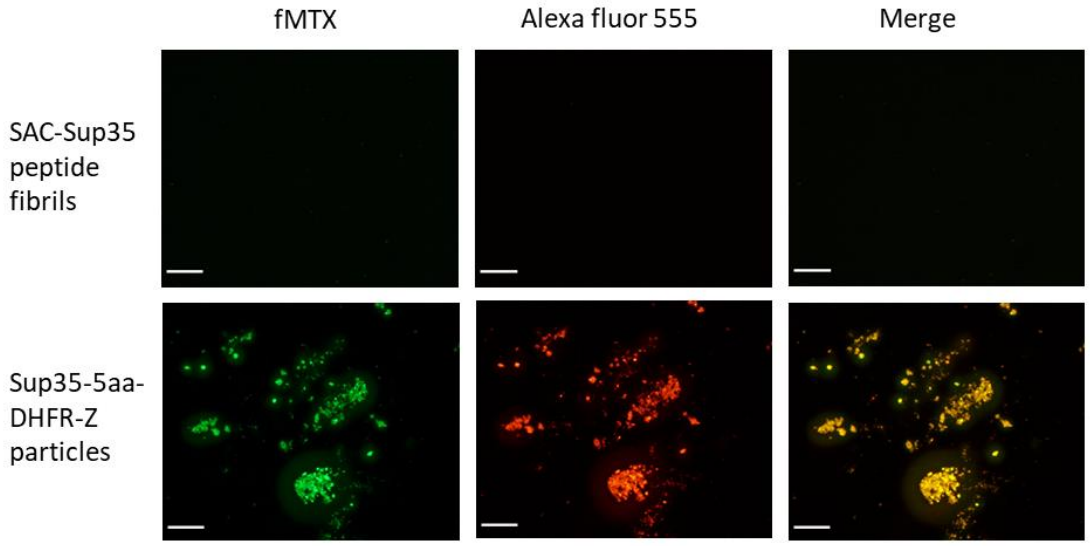

Figure 6. Dual-binding activity of Sup35-5aa-DHFR-Z nanoparticles. Representative fluorescence microscopy images of SAC-Sup35 peptide fibrils (upper panel) and Sup35-5aa-DHFR-Z nanoparticles (lower panel) incubated with fMTX (fluorescein, green channel) and an anti-EGFR antibody (Alexa fluor 555, red channel). Scale bar represents 50 μm.

Sup35-5aa-DHFR-Z functionalized nanoparticles target cancer cells specifically

We assessed if Sup35-5aa-DHFR-Z nanoparticles can target specific antigens in living cells, once they have been loaded with antibodies through their multiple Z-domains. Nanoparticles loaded with red-labeled anti-EGFR antibody (NPs-anti-EGFR) were incubated with HeLa cells, which are known to overexpress EFGR at their membranes.[29] As can be observed in **Figure 7A**, confocal microscopy fluorescence images indicated that the vast majority of the cells were red fluorescent and therefore, that the multivalent NPs-anti-EGFR had recognized them. In contrast, when the Sup35-5aa-DHFR-Z nanoparticles were loaded with a green-labeled anti-CD3 antibody and incubated with HeLa cells, no cellular labeling was detected, consistent with the fact that this cell type does not express the CD3 complex. Thus, the data indicated that the recognition of HeLa cells by NPs-anti-EGFR was antibody-driven and specific.

Next, we loaded Sup35-5aa-DHFR-Z nanoparticles, with both fMTX and anti-EGFR antibody (NPs-fMTX-anti-EGFR), with fMTX and a secondary goat anti-rabbit antibody (NPs-fMTX-anti-rabbit) or only with fMTX (NPs-fMTX) and we added the three functionalized protein nanoparticles to HeLa cell cultures. After washing the cells with PBS, they were immediately analyzed by flow cytometry, monitoring the green fluorescence of fMTX, present in all three nanoparticles, with a FITC emission detector. Only HeLa cells treated with NPs-fMTX-Anti-EGFR exhibited fMTX fluorescence (**Figure 7B**), whereas no significant fluorescence was detected for HeLa cells treated with NPs-fMTX-anti-rabbit or NPs-fMTX. Flow cytometry is quantitative, and the analysis indicated that > 70 % of the cells were bound to NPs-fMTX-anti-EGFR, which shows a high binding affinity of the functionalized nanoparticles for EGFR expressing cells. Indeed, confocal images of cells incubated with NPs-fMTX-anti-EGFR indicated that they target the cell membrane (**Figure S6**).

Overall, it appears that Sup35-5aa-DHFR-Z nanoparticles provide a modular strategy to target any cell type of interest, just by incorporating the adequate antibody thanks to its strong interaction with the properly folded Z-domains and the multivalency of the assembly.

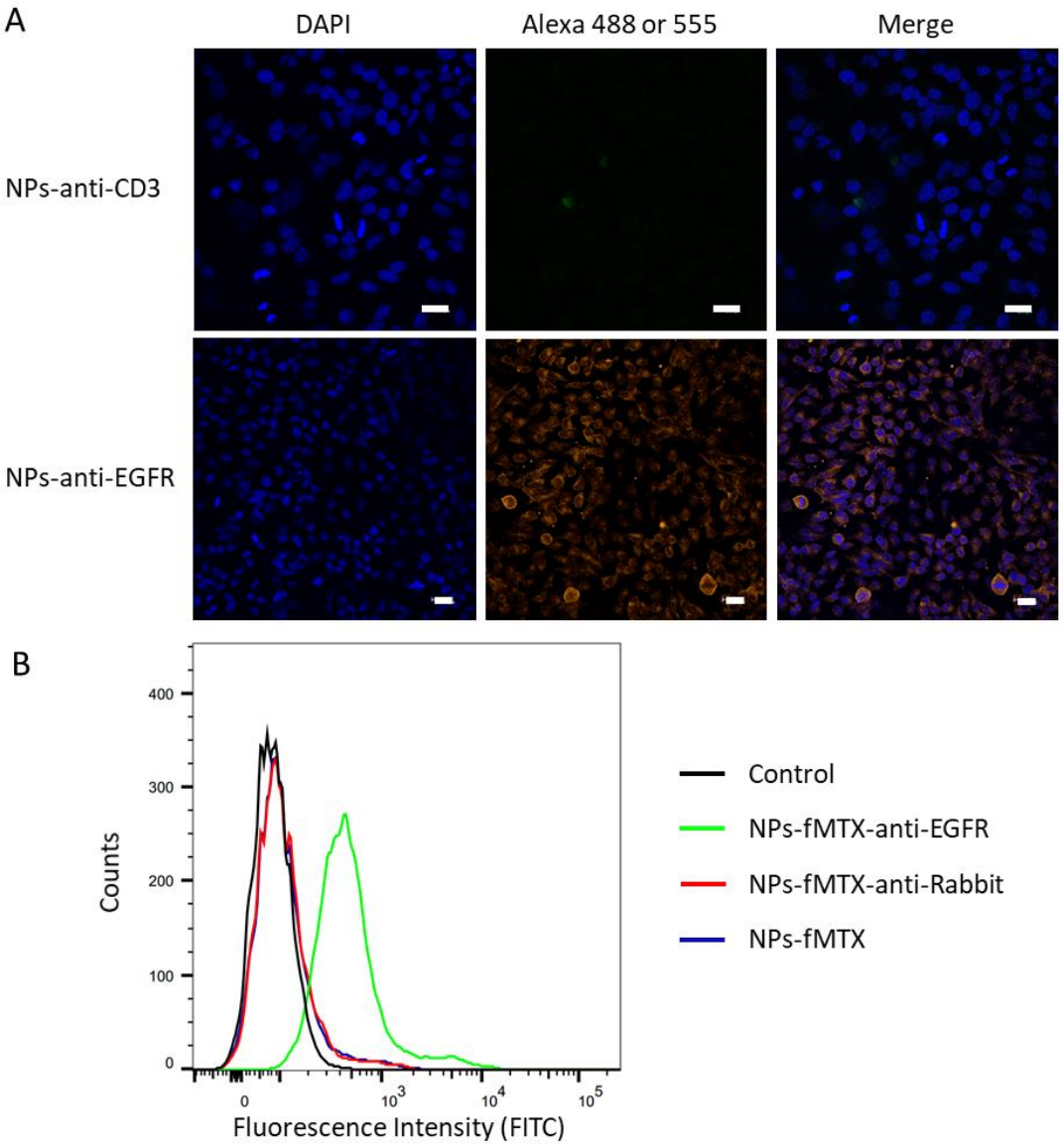

Figure 7. Binding specificity of functionalized Sup35-5aa-DHFR-Z nanoparticles to HeLa cells. (A) Representative confocal microscopy images of HeLa cells incubated with nanoparticles conjugated with an anti-CD3 antibody (NPs-anti-CD3, Alexa 488) (upper panel) or an anti-EGFR antibody (NPs-anti-EGFR, Alexa 555) (lower panel). Scale bar represents 20 μm. (B) Quantitative analysis of fluorescein fluorescence on HeLa cells by flow cytometry. HeLa cells were incubated with fMTX, and anti-EGFR loaded nanoparticles (NPs-fMTX-anti-EGFR, green line), fMTX and goat anti-rabbit antibody loaded nanoparticles (NPs-fMTX-anti-rabbit, red line), and fMTX loaded nanoparticles (NPs-fMTX, blue line). HeLa cells treated with PBS were used as a control. Fluorescein labeled MTX (fMTX) was used as a fluorescence probe to calculate the proportion of cancer cells bound to nanoparticles using an FITC detector.

Proteinase-controlled release of MTX from Sup35-5aa-DHFR-Z nanoparticles

Cancer cells secrete a significant number of proteinases, including metalloproteinases,[30] serine proteinases[31] and cathepsins.[32] This feature has been exploited to develop proteinase-responsive nanocarriers[33] and polymer nanoshells.[34]

It is known that amyloid fibrils are highly resistant to proteolysis, but we hypothesized that the globular domains hanging from the amyloid core in Sup35-5aa-DHFR-Z nanoparticles would be protease-sensitive. This will open an opportunity to use the DHFR moiety as a carrier for MTX, under the assumption that the inhibitor will be released from the oligomers once they arrive at a protease-rich environment and the accessible DHFR domains would be proteolytically attacked.

We incubated fMTX loaded nanoparticles with the broad-spectrum proteinase K (PK) at a final concentration of 1 μg/mL, and measured the kinetics of fluorescein fluorescence apparition in the supernatant, upon centrifugation (**Figure 8A**). The nanoparticles were incubated also in PBS alone or treated with lysozyme at the same concentration to ensure that any observed release is not spontaneous or unspecific. In contrast to control samples, PK treated nanoparticles showed a continuous delivery of fMTX to the solution, with a cumulative release efficiency > 90% after 4 hours of reaction.

An SDS-PAGE of the nanoparticles indicated that they were SDS-sensitive since, despite high molecular species were observed, the vast majority of the fusion protein ran as a monomer. This sensitivity to the detergent is typical of amyloid-oligomers.[35] This allowed us to follow the PK digestion by SDS-PAGE, which confirmed that the nanoparticles are proteinase-sensitive, and that not intact Sup35-5aa-DHFR-Z fusion remains after 45 min; neither any fragment compatible with an intact DHFR domain (**Figure S7**), explaining why fMTX is released from the protein nanoparticles.

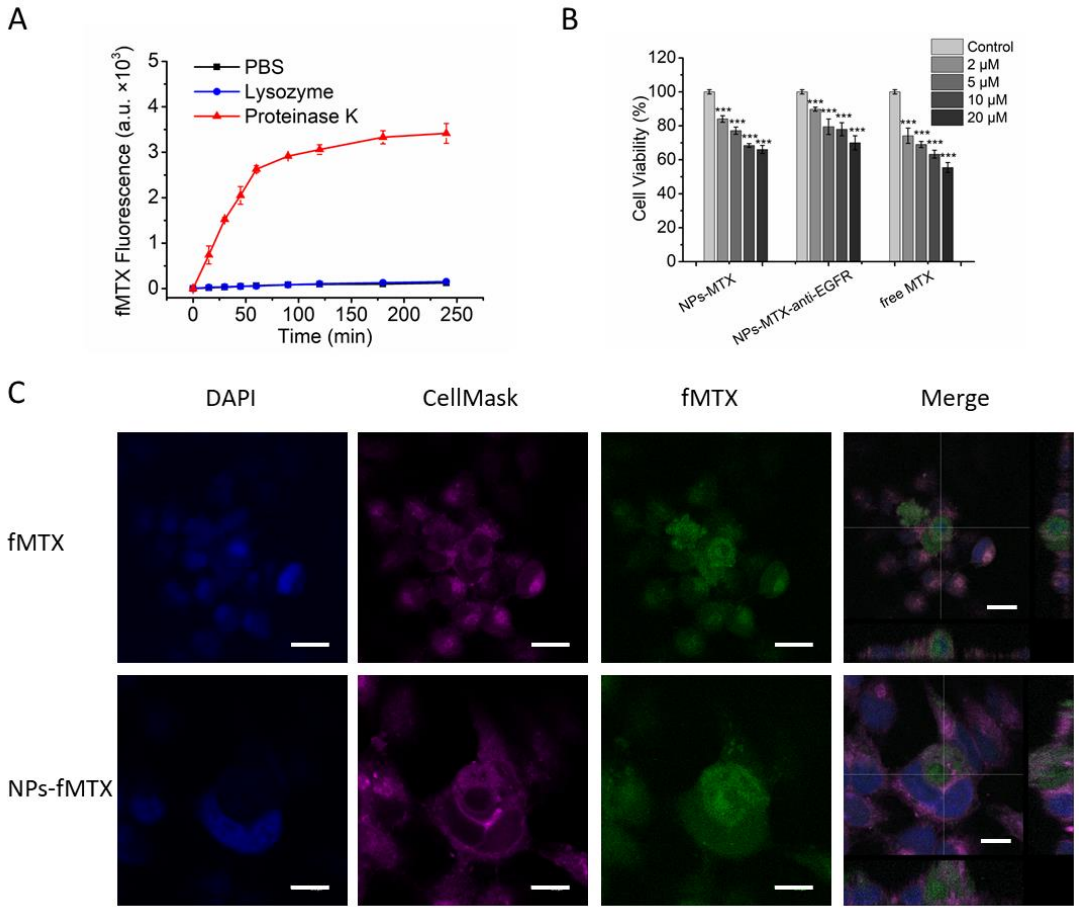

Figure 8. Release, uptake, and cytotoxicity of fMTX in Sup35-5aa-DHFR-Z nanoparticles. (A) *In vitro* release of fMTX from Sup35-5aa-DHFR-Z nanoparticles in PBS buffer in the presence of 1 μg/mL proteinase K (red line) at 37 °C (n=2). The fluorescence emission of fluorescein-labeled MTX at 515 nm in the supernatant was monitored at the indicated time points. PBS (black line) and lysozyme (blue line) treated nanoparticles were used as a negative controls (B) Viability of HeLa cells in the presence of free MTX, MTX loaded nanoparticles (NPs-MTX) or MTX and anti-EGFR loaded nanoparticles (NPs-MTX-anti-EGFR). PBS buffer was used as a control. A final concentration of 5 ng/mL of proteinase K was added to the medium to mimic the proteinase-enriched microenvironment of tumours. Results are expressed as means ± SD (n=3), and analyzed using a one-way ANOVA test. The statistical differences between the control group and the test group were established at $P < 0.001$. (C) Representative confocal images of HeLa cells incubated with free fMTX or fMTX loaded nanoparticles (NPs-fMTX) preincubated with proteinase K. Nuclei were stained with DAPI and membranes were stained with CellMask Deep Red. Fluorescein labeled MTX (fMTX) was detected in the FITC channel. Scale bar represents 10 μm.

Internalization of MTX released from loaded nanoparticles

We investigated if the DHFR inhibitor released from NPs-fMTX can be uptaken by HeLa cells. To this aim, we incubated the cells with 10 μM of free fMTX or fMTX loaded nanoparticles, both preincubated with 50 ng/mL PK for 5 min, followed by protease inactivation with EDTA. After 4 hours of incubation, the medium was removed, and the cells were washed with medium for three times. Then cells were stained with DAPI and CellMask Deep Red for nuclei and membrane visualization, respectively. Intracellular green fluorescent fMTX was observed by confocal microscopy both for free added fMTX and NPs-fMTX (**Figure 8C**), indicating that the nanoparticle released inhibitor is efficiently internalized.

MTX loaded nanoparticles induce death of cancer cells

We have demonstrated that the MTX loaded nanoparticles can liberate DHFR bound MTX and that the inhibitor can internalize into cancer cells. This immediately suggested that MTX loaded nanoparticles could induce proteolysis-mediated cell death. Therefore, we measured the viability of HeLa cells in the presence of NPs-MTX, NPs-MTX-anti-EGFR, or free MTX. A final concentration of 5 ng/mL PK was added to the cell culture medium to mimic the proteinase-enriched microenvironment of tumor tissues. In all cases, a significant dose-dependent decrease in HeLa cell death was observed, compared with control cells incubated with PK only (**Figure 8B**). The concentration of proteinase in the assay is lower than the one described in tumors' extracellular environment;[36] still it is sufficient to promote MTX release from the nanoparticles, internalization of the inhibitor, and cancer cell death.

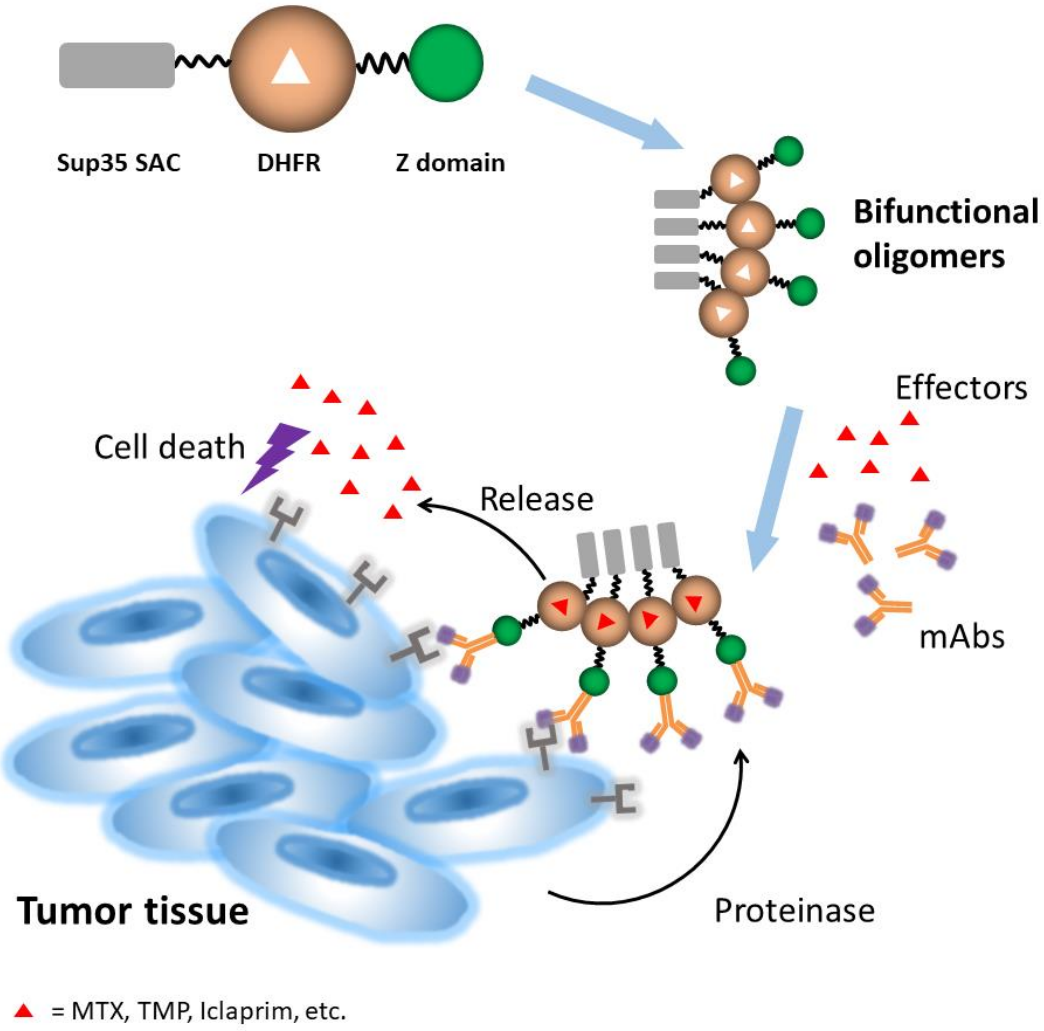

Figure 9. Schematic illustration of bifunctional amyloid oligomeric nanoparticles and their potential application for targeted drug delivery. The construct consists of Sup35 soft amyloid core (Sup35-SAC, grey square), dihydrofolate reductase (DHFR, orange circle) and the Z domain derived from protein A (green circle). The short linker between Sup35-SAC and DHFR enable the fusion protein to self-assemble into stable bifunctional oligomeric nanoparticles. Drugs (red triangles) and monoclonal antibodies (mAbs) loaded into the nanoparticles target the tumor cells and release the drugs in the presence of the proteinase enriched environment. As a consequence, the released drugs are internalized and induce the death of the targeted tumor cells.

Conclusion

In the present work, we first demonstrated that by modulating inter-domain linker length, one could attain a tight control of the mesoscopic properties of the resulting amyloid-like nanostructures. This strategy does not require intricate structural engineering since it relies on fundamental biophysical principles. The approach allowed us to generate oligomeric amyloid-like nanoparticles. These spherical nanoparticles are homogenous in size, stable, and biocompatible. Besides, the multiple globular domains they contain are highly active in the assembled state, as shown here for DHFR and the Z-domain. Therefore, they constitute a novel kind of functional and functionalizable nanomaterial.

We provide a proof-of-concept of the utility of these de novo designed nanostructures by showing how they can be decorated with an antibody of interest, which act as an antenna, directing the multivalent nanoparticles to the specific cell types expressing the selected antigen at their surfaces, allowing to discriminate between diseased and functional cells. When this ability is combined with the capability of DHFR to carry and shield MTX, avoiding non-specific toxic effects, it appears as a very appealing strategy for targeted delivery of the drug in the proteinase enriched microenvironment of tumors, which is followed by MTX internalization into the cytosol and localized killing of cancer cells (**Figure 9**). Overall, the multifunctional, self-assembled amyloid nanoparticles we present here constitute a new and safe nanotechnological modular scaffold with the potential of facilitating the specific delivery of agents to specific sites in the body, overpassing the major barrier for bioimaging and tissue-targeted therapies. Importantly, the building blocks of these nanostructures can be produced at high yield and purified at homogeneity from the soluble cell fraction, which results in reduced costs when compared with alternative nanomedicines.

Materials and Methods

Reagents and Enzymes. Reagents and enzymes were purchased from Sigma-Aldrich (UK) unless otherwise stated. Carbon grids (400 square mesh copper) were purchased from Micro to Nano (Netherlands), and the uranyl acetate solution was provided by the microscopy service at Universitat Autònoma de Barcelona. Sup35-SAC peptides were purchased from CASLO ApS (Scion Denmark Technical University).

Expression and Purification of Proteins. The cDNA of Sup35-5aa-DHFR, cloned in the plasmid pET28(a) with a His6 tag, was acquired from GenScript (USA). The constructs: pET28(a)/Sup35-8aa-DHFR, pET28(a)/DHFR-wt, pET28(a)/Sup35-5aa-DHFR-Z were obtained by mutagenesis on top of the plasmid pET28(a)/Sup35-5aa-DHFR. *E.coli* BL21 (DE3) competent cells were transformed with the corresponding plasmids. Then, transformed cells were grown in 10 mL LB medium containing 50 μg/mL kanamycin, overnight at 37 °C and transferred into 1 L fresh LB media containing 50 μg/mL kanamycin. After reaching an OD_{600} of 0.6, the culture was induced with 0.4 mM IPTG and grown at 20 °C for 16 h. Cells were collected by centrifugation at 5000 rpm for 15 min at 4 °C. The collected pellet was resuspended into 20 mL PBS pH 7.4 containing 20 mM imidazole, 1 mg/mL lysozyme, and 1 mM PMSF. The solution was incubated on ice, followed by sonication for 20 min. The supernatant was collected by centrifugation at 15000 rpm for 30 min at 4 °C and purified using a nickel-charged IMAC column, followed by a gel filtration onto a HiLoad™ Superdex™ 75 prep grade column (GE Healthcare, USA). The purified proteins were frozen with liquid nitrogen and stored at -80 °C. SDS-PAGE confirmed the purity of the samples. The concentration of the proteins: DHFR wt, Sup35-5aa-DHFR, Sup35-8aa-DHFR, and Sup35-5aa-DHFR-Z was determined by UV absorption using an ε value of 33585 $L \cdot mol^{-1} \cdot cm^{-1}$, 38055 $L \cdot mol^{-1} \cdot cm^{-1}$, 38055 $L \cdot mol^{-1} \cdot cm^{-1}$, and 39420 $L \cdot mol^{-1} \cdot cm^{-1}$ respectively.

Conformational Characterisation. Proteins were dissolved at a final concentration of 20 μM in PBS pH 7.4 buffer, then samples were filtered through a 0.22 μm Millipore filter and immediately analyzed. Far-UV CD spectra were recorded from 260 nm to 200 nm at 1 nm bandwidth, with a response time of 1 second, and a scan speed of 100 nm/min in a Jasco-710 spectropolarimeter (Jasco Corporation, Japan), thermostated at 25 °C. Ten accumulations were averaged for each spectrum.

Aggregation Assay. Proteins were prepared at 200 μM for Sup35-5aa-DHFR and Sup35-8aa-DHFR, 600 μM for Sup35-5aa-DHFR-Z in PBS pH 7.4, and filtered through a 0.22 μm filter. The samples were incubated at 25 °C, with agitation for four days. DHFR-wt was incubated at the same concentrations and conditions.

Amyloid Dyes Binding. Thioflavin T (Th-T) and Congo red (CR) were used to monitor the formation of amyloid assemblies. For the Th-T binding assay, incubated proteins were diluted to a final concentration of 20 μM in PBS pH 7.4, in the presence of 25 μM Th-T. Emission fluorescence was recorded in the 460-600 nm range, using an excitation wavelength of 445 nm and emission bandwidth of 5 nm on a Jasco FP-8200 Spectrofluorometer (Jasco Corporation, Japan). For the CR binding assay, incubated proteins were prepared at a final concentration of 20 μM and, CR was mixed to a final concentration of 20 μM. Absorption spectra were recorded in the range from 375 to 700 nm in a Specord 200 Plus spectrophotometer (Analytik Jena, Germany). Spectra of protein alone and buffer were acquired to subtract their contribution to the signal.

Transmission Electron Microscopy (TEM). For TEM samples preparation, 10 μL of the incubated proteins were deposited on a carbon-coated copper grid for 10 min, and the excess liquid was removed with filter paper, followed by negative staining with 10 μL of 2 % (w/v) uranyl acetate for 1 min. Grids were scanned using a JEM 1400 transmission electron microscope (JEOL Ltd, Japan) operating at 80 kV, and images were acquired with a CCD GATAN ES1000W Erlangshen camera (Gatan Inc., USA). The particles' diameter was analyzed with the Image J software (National Health Institute), averaging the measures of 100 individual particles.

Scanning Electron Microscopy (SEM). A sample of 50 µL of incubated protein was centrifuged, and the resulting supernatant removed. The precipitate was resuspended in 50 µL water and washed twice to remove salt traces. 5 µL of resuspension was deposited on a silicon slice (0.5 cm × 0.5 cm) and dried with nitrogen flow. Silicon slices were scanned using a Merlin field-emission scanning electron microscopy (Zeiss ltd, Germany) at 2 kV, and images acquired with an in-lens SE detector. The diameter of the particles was calculated with Image J (National Health Institute), averaging the measures of 100 individual particles.

Fourier Transform Infrared Spectroscopy (FTIR). DHFR-wt was dissolved at 100 µM in PBS, pH 7.4, and filtered through a 0.22 µm filter. 30 µL of the incubated proteins were centrifuged at 12000×g for 30 min and resuspended in 10 µL of water. Samples were placed on the ATR crystal and dried out under N_2 flow. The experiments were carried out in a Bruker Tensor 27 FTIR (Bruker Optics, USA) supplied with a Specac Golden Gate MKII ATR accessory. Each spectrum consists of 32 acquisitions measured at a resolution of 2 cm^{-1} using the three-term Blackman-Harris Window apodization function. Data was acquired and normalized using the OPUS MIR Tensor 27 software (Bruker Optics, USA), with the Min/Max normalization method, which scales spectrum intensities to the effect that the minimum absorbance unit will be 0 and the maximum 2. IR spectra were fitted employing a nonlinear peak-fitting equation using Origin 8.5 (OriginLab Corporation).

Fluorescence Microscopy. 20 µL of incubated fusion proteins were centrifuged at 12000×g for 30 min, and the supernatant was removed. The precipitate was resuspended in PBS pH 7.4 containing 100 µM fMTX (methotrexate labeled with fluorescein, Thermo Fisher Scientific, USA) and incubated for 30 min. For the bifunctional fusion protein Sup35-5aa-DHFR-Z, the precipitate was resuspended in PBS pH 7.4 containing 100 µM fMTX and 1 µg anti-EGFR antibody labeled with Alexa 555 (Thermo Fisher Scientific, USA) and incubated for 30 min. Samples were washed three times to remove any unbound fMTX and antibody, then resuspended in PBS to final volume 50 µL. 5 µL of the resuspension were dropped onto a clean glass slide

(Deltalab, 26×76 mm) and covered by a cover slide (Deltalab, 22×22mm). Fluorescence imaging of particles and fibrils was carried out on an Eclipse 90i epifluorescence optical microscope equipped with a Nikon DXM1200F (Nikon, Japan) camera and ACT-1 software. Images were acquired with an excitation filter of 465-495 nm or 540-580 nm and detecting fluorescence emission in a range of 515-555 nm or 590-630 nm.

Dynamic Light Scattering (DLS). The size of the incubated proteins was determined using a Malvern Zetasizer Nano ZS90 (ATA Scientific, Australia) in PBS buffer, pH 7.4, at 25 °C.

Catalytic Activity of Dihydrofolate Reductase Embedded in Amyloid Assemblies. The Sup35-5aa-DHFR particles and Sup35-8aa-DHFR fibrils were prepared at 2 μM in activity buffer (0.1 M potassium phosphate, pH 7.4, 1 mM DTT, 0.5 M KCl, 1 mM EDTA, 20 mM sodium ascorbate). 50 μL of the sample were mixed with 850 μL of activity buffer in a cuvette and preincubated at room temperature for 10 min. The reaction was then initiated by adding 50 μL of 2 mM 7,8-dihydrofolate (DHF) solution and 50 μL of a 2 mM NADPH solution. Sup35-SAC fibrils and lysozyme were used as negative controls. For the determination of the kinetic parameters, a final concentration of 20 nM assemblies and 20 μM of DHF were preincubated in activity buffer at room temperature. The reaction was initiated by adding NADPH to a final concentration in the range from 10-200 μM. The absorbance change at 340 nm was monitored on a Specord 200 Plus spectrophotometer (Analytik Jena, Germany). The initial velocity at each concentration was determined upon fitting the reaction with Origin 8.5 (OriginLab Corporation). The reciprocal of the velocity of the reaction was plotted against the reciprocal of the concentration of NADPH. The catalytic constants V_{max}, K_m, K_{cat}, and K_{cat}/K_m were calculated.

Thermodynamic Dissociation Constants. The equilibrium dissociation constants (K_d) for enzyme-NADPH complexes were determined by fluorescence titration at 25 °C. In particular, the Sup35-5aa-DHFR and Sup35-8aa-DHFR assemblies were prepared at 100 nM in activity buffer in a quartz cuvette. The titrations were carried out by serial

additions of aliquots (1-4 μL) of NADPH into the cuvette. The intrinsic fluorescence spectra in the range from 300 nm to 400 nm were recorded 2 min after the addition of NADPH, using an excitation wavelength of 290 nm and emission bandwidth of 5 nm in a Jasco FP-8200 Spectrofluorometer (Jasco Corporation, Japan). A standard tryptophan solution was titrated, and the emission used to correct for the inner filter effect caused by NADPH. The change in intrinsic fluorescence emission at 340 nm was plotted against NADPH's concentration using Origin 8.5 (OriginLab Corporation) and the K_d calculated from the binding equation.

Preparation of fMTX or/and Antibody Loaded Nanoparticles. The incubated protein Sup35-DHFR-Z was precipitated by centrifugation at 13000 rpm for 30 min. The precipitate was resuspended in PBS buffer pH 7.4 containing 1 μg antibody or/and 100 μM fMTX (fluorescein-labeled MTX) and then incubated for 30 min. The concentration of nanoparticles was determined by the reduction of absorbance at 280 nm in the supernatant fraction. Three labeling antibodies (Thermo Fisher Scientific, USA): anti-EGFR antibody labeled with Alexa fluor 555, anti-CD3 antibody labeled with Alexa fluor 488, and goat anti-rabbit antibody labeled with Alexa fluor 555 were used in this study. Then the antibody or/and fMTX conjugated particles were washed three times with PBS buffer to remove any unbound antibody or/and fMTX and resuspended in PBS buffer. The antibody loaded nanoparticles (NPs-anti-EGFR and NPs-anti-CD3), fMTX loaded nanoparticles (NPs-fMTX) and fMTX and antibody loaded nanoparticles (NPs-fMTX-anti-EGFR and NPs-fMTX-anti-rabbit) were used for experiments immediately.

Cell Culture. The human HeLa cell line was obtained from American Type Culture Collection (ATCC). HeLa cells were maintained in Minimum Essential Medium Alpha media (MEM-α), supplemented with 10% fetal bovine serum (FBS), and incubated at 37 ℃ with 5% CO_2.

Cytotoxicity Assay. HeLa cells were seeded on a 96-well plate at a concentration of 3 $\times 10^3$/well and incubated at 37 ℃ for 24 h. The medium was replaced with 100 μL fresh

medium containing the treatments. Unloaded Sup35-5aa-DHFR-Z particles and free anti-EGFR antibody were assayed in a range of concentrations 0.5~60 μM and 2~100 μg/mL, respectively. MTX loaded nanoparticles (NPs-MTX), MTX and anti-EGFR antibody loaded nanoparticles (NPs-MTX-anti-EGFR) and free MTX were assayed in a concentration range of 2-20 μM. For NPs-MTX-anti-EGFR samples, the medium was removed, and cells were rinsed three times with DPBS buffer after incubation of 20 min. Then fresh medium was added. PBS alone was used as vehicle control and the medium without cells as a blank. A final concentration of 5 ng/mL proteinase K was added to the medium to mimic the proteinase-enriched microenvironment of tumor tissues. Each sample was measured in triplicate, and the plate was incubated at 37 ℃ with 5% CO_2 for 48 h. 10 μL of PrestoBlue® cell viability reagent (reagent (ThermoFisher Scientific, USA) was added to each well and incubated for another 1 h. The fluorescence emission was analyzed on a Victor III Multilabel Plate Reader (Perkin Elmer, USA), equipped with a 530/10 nm CW-lamp filter and 590/20 nm emission filter. The viability of cells was calculated as follows:

$$\text{Viability (\%)} = (I_{test}-I_{blank}) / (I_{control}-I_{blank}) \times 100\%$$

Where the I_{test}, I_{blank}, and $I_{control}$ are the fluorescence intensity of test, blank and control group, respectively. The significance of the differences between the test groups and the control were analyzed by one-way Analysis of Variance (ANOVA) using the Origin 8.5 program (OriginLab Corporation).

Confocal Microscopy. HeLa cells were cultured on an 8-well Millicell® EZ slide (Millipore, Germany) to a final confluence of 70-80%. Then the medium was replaced with fresh medium containing 20 μM of NPs-anti-EGFR and NPs-fMTX-anti-EGFR. The slide was incubated at 37 ℃, 5% CO_2 for 20 min. The anti-CD3 IgG loaded nanoparticles (NPs-anti-CD3) were used as a control. For the internalization assay, the medium was replaced with fresh medium containing 10 μM Free fMTX and fMTX loaded nanoparticles (NPs-fMTX) preincubated with 50 ng/mL proteinase for 5 min followed by addition of EDTA. The slide was incubated at 37 ℃, 5% CO_2 for 4 h. Then

the medium was removed. The adherent cells on the slide were rinsed three times with fresh medium. Cells were stained with CellMask Deep Red® (Invitrogen, ThermoFisher Scientific) at a final dilution of 1:1000 for 10 min. The medium was removed and washed three times with PBS buffer. Cells were fixed with 4% PFA at room temperature for 20 min, followed by a washing step with PBS buffer. 10 μL of mounting medium containing DAPI was dropped onto each well of the slide and covered with a coverslip. The slide was imaged on a Leica TCS SP5 confocal microscope (Leica Biosystems, Germany). Images were acquired using 405 nm, 488 nm or 561 nm, and 633 nm excitation laser for DAPI, Alexa fluor 488 or 555nm, and CellMask Deep Red, respectively.

Flow Cytometry Assay. The NPs-fMTX-anti-EGFR, NPs-fMTX, and NPs-fMTX-anti-rabbit were prepared as described above. HeLa cells were prepared in PBS buffer pH 7.4 at a final concentration of 1×10^6/mL. 200 μL of HeLa cells were precipitated and then resuspended in 200 μL PBS buffer containing 20 μM NPs-fMTX-anti-EGFR, NPs-fMTX, and NPs-fMTX-anti-rabbit, respectively. After 30 min of incubation, the cells were pelleted and washed three times. Then 200 μL of cell suspension were analyzed using a FACSCalibur cytometry (BD Biosciences, Becton Dickinson, USA), equipped with a FITC laser. Fluorescence intensities of cell-bound nanoparticles were analyzed and quantitated using FlowJo™ (BD Biosciences, USA). Cells treated with PBS were used as control.

Proteinase Digestion of fMTX Loaded Sup35-DHFR-Z Particles. The fMTX loaded particles (NPs-fMTX) were resuspended in PBS buffer containing proteinase K at a final concentration of 1 μg/mL. SDS-PAGE was used to analyze the proteolytic progress at 15, 30, and 45 min. The supernatant's fluorescence spectrum was recorded in the 500-600 nm range at different time points, using an excitation wavelength of 488 nm and emission bandwidth of 5 nm on a Jasco FP-8200 Spectrofluorometer (Jasco Corporation, Japan). The emission maximum was plotted as a function of time. The

experiment was performed in duplicate. Lysozyme and PBS, instead of proteinase K, were used as negative controls.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This work was funded by the Spanish Ministry of Economy and Competitiveness BIO2016-78310-R to S.V and by ICREA, ICREA-Academia 2015 to S.V. Weiqiang Wang acknowledges financial support from the China Scholarship Council (CSC): NO. 201606500007.

References

1. Knowles, T. P.; Mezzenga, R., Amyloid fibrils as building blocks for natural and artificial functional materials. *Advanced Materials* **2016,** *28* (31), 6546-6561.

2. Ross, C. A.; Poirier, M. A., Protein aggregation and neurodegenerative disease. *Nature medicine* **2004,** *10* (7), S10-S17.

3. Fowler, D. M.; Koulov, A. V.; Balch, W. E.; Kelly, J. W., Functional amyloid–from bacteria to humans. *Trends in biochemical sciences* **2007,** *32* (5), 217-224.

4. Wei, G.; Su, Z.; Reynolds, N. P.; Arosio, P.; Hamley, I. W.; Gazit, E.; Mezzenga, R., Self-assembling peptide and protein amyloids: from structure to tailored function in nanotechnology. *Chemical Society Reviews* **2017,** *46* (15), 4661-4708.

5. Rufo, C. M.; Moroz, Y. S.; Moroz, O. V.; Stöhr, J.; Smith, T. A.; Hu, X.; DeGrado, W. F.; Korendovych, I. V., Short peptides self-assemble to produce catalytic amyloids. *Nature chemistry* **2014,** *6* (4), 303-309.

6. Chiti, F.; Dobson, C. M., Protein misfolding, amyloid formation, and human disease: a summary of progress over the last decade. *Annual review of biochemistry* **2017,** *86*, 27-68.

7. Wang, W.; Navarro, S.; Azizyan, R. A.; Baño-Polo, M.; Esperante, S. A.; Kajava, A. V.; Ventura, S., Prion soft amyloid core driven self-assembly of globular proteins into bioactive nanofibrils. *Nanoscale* **2019,** *11* (26), 12680-12694.

8. Sant'Anna, R.; Fernández, M. R.; Batlle, C.; Navarro, S.; De Groot, N. S.; Serpell, L.; Ventura, S., Characterization of amyloid cores in prion domains. *Scientific reports* **2016,** *6* (1), 1-10.

9. Sabate, R.; Rousseau, F.; Schymkowitz, J.; Ventura, S., What makes a protein sequence a prion? *PLoS Comput Biol* **2015,** *11* (1), e1004013.

10. Sabate, R.; Rousseau, F.; Schymkowitz, J.; Batlle, C.; Ventura, S., Amyloids or prions? That is the question. *Prion* **2015,** *9* (3), 200-206.

11. Wasmer, C.; Schütz, A.; Loquet, A.; Buhtz, C.; Greenwald, J.; Riek, R.; Böckmann, A.; Meier, B. H., The molecular organization of the fungal prion HET-s in its amyloid form. *Journal of molecular biology* **2009,** *394* (1), 119-127.

12. Baxa, U.; Cassese, T.; Kajava, A. V.; Steven, A. C., Structure, function, and amyloidogenesis of fungal prions: filament polymorphism and prion variants. *Advances in protein chemistry* **2006,** *73*, 125-180.

13. Li, J.; McQuade, T.; Siemer, A. B.; Napetschnig, J.; Moriwaki, K.; Hsiao, Y.-S.; Damko, E.; Moquin, D.; Walz, T.; McDermott, A., The RIP1/RIP3 necrosome forms a functional amyloid signaling complex required for programmed necrosis. *Cell* **2012,** *150* (2), 339-350.

14. Chen, A. K.-H.; Lin, R. Y.-Y.; Hsieh, E. Z.-J.; Tu, P.-H.; Chen, R. P.-Y.; Liao, T.-Y.; Chen, W.; Wang, C.-H.; Huang, J. J.-T., Induction of amyloid fibrils by the C-terminal fragments of TDP-43 in amyotrophic lateral sclerosis. *Journal of the American Chemical Society* **2010,** *132* (4), 1186-

1187.

15. Azizyan, R. A.; Garro, A.; Radkova, Z.; Anikeenko, A.; Bakulina, A.; Dumas, C.; Kajava, A. V., Establishment of constraints on amyloid formation imposed by steric exclusion of globular domains. *Journal of molecular biology* **2018,** *430* (20), 3835-3846.

16. Kajava, A. V.; Baxa, U.; Wickner, R. B.; Steven, A. C., A model for Ure2p prion filaments and other amyloids: the parallel superpleated β-structure. *Proceedings of the National Academy of Sciences* **2004,** *101* (21), 7885-7890.

17. Schnell, J. R.; Dyson, H. J.; Wright, P. E., Structure, dynamics, and catalytic function of dihydrofolate reductase. *Annu. Rev. Biophys. Biomol. Struct.* **2004,** *33*, 119-140.

18. Tashiro, M.; Tejero, R.; Zimmerman, D. E.; Celda, B.; Nilsson, B.; Montelione, G. T., High-resolution solution NMR structure of the Z domain of staphylococcal protein A. *Journal of molecular biology* **1997,** *272* (4), 573-590.

19. Forsgren, A.; Sjöquist, J., "Protein A" from S. aureus: I. Pseudo-immune reaction with human γ-globulin. *The Journal of Immunology* **1966,** *97* (6), 822-827.

20. Huffman, D. H.; Wan, S. H.; Azarnoff, D. L.; Hoogstraten, B., Pharmacokinetics of methotrexate. *Clinical Pharmacology & Therapeutics* **1973,** *14* (4part1), 572-579.

21. Blaney, J. M.; Hansch, C.; Silipo, C.; Vittoria, A., Structure-activity relationships of dihydrofolated reductase inhibitors. *Chemical Reviews* **1984,** *84* (4), 333-407.

22. Qindeel, M.; Khan, D.; Ahmed, N.; Khan, S.; Rehman, A. u., Surfactant-Free, Self-Assembled Nanomicelles-Based Transdermal Hydrogel for Safe and Targeted Delivery of Methotrexate against Rheumatoid Arthritis. *ACS nano* **2020,** *14* (4), 4662-4681.

23. Sadrjavadi, K.; Shahbazi, B.; Fattahi, A., De-esterified tragacanth-chitosan nano-hydrogel for methotrexate delivery; optimization of the formulation by Taguchi design. *Artificial cells, nanomedicine, and biotechnology* **2018,** *46* (sup2), 883-893.

24. Tawfik, M. K., Combination of coenzyme Q10 with methotrexate suppresses Freund's complete adjuvant-induced synovial inflammation with reduced hepatotoxicity in rats: effect on oxidative stress and inflammation. *International Immunopharmacology* **2015,** *24* (1), 80-87.

25. Choi, G.; Kim, T.-H.; Oh, J.-M.; Choy, J.-H., Emerging nanomaterials with advanced drug delivery functions; focused on methotrexate delivery. *Coordination Chemistry Reviews* **2018,** *359*, 32-51.

26. Bystroff, C.; Oatley, S. J.; Kraut, J., Crystal structures of Escherichia coli dihydrofolate reductase: The NADP+ holoenzyme and the folate. cntdot. NADP+ ternary complex. Substrate binding and a model for the transition state. *Biochemistry* **1990,** *29* (13), 3263-3277.

27. Moks, T.; ABRAHMSÉN, L.; NILSSON, B.; HELLMAN, U.; SJÖQUIST, J.; UHLÉN, M., Staphylococcal protein A consists of five IgG-binding domains. *European journal of biochemistry* **1986,** *156* (3), 637-643.

28. Díaz-Caballero, M.; Navarro, S.; Ventura, S., Soluble assemblies in the fibrillation pathway of

prion-inspired artificial functional amyloids are highly cytotoxic. *Biomacromolecules* **2020**.

29. Laskin, J. J.; Sandler, A. B., Epidermal growth factor receptor: a promising target in solid tumours. *Cancer treatment reviews* **2004,** *30* (1), 1-17.

30. Folgueras, A. R.; Pendas, A. M.; Sanchez, L. M.; Lopez-Otin, C., Matrix metalloproteinases in cancer: from new functions to improved inhibition strategies. *International Journal of Developmental Biology* **2004,** *48* (5-6), 411-424.

31. Koshikawa, N.; Yasumitsu, H.; Umeda, M.; Miyazaki, K., Multiple secretion of matrix serine proteinases by human gastric carcinoma cell lines. *Cancer research* **1992,** *52* (18), 5046-5053.

32. Koblinski, J. E.; Dosescu, J.; Sameni, M.; Moin, K.; Clark, K.; Sloane, B. F., Interaction of human breast fibroblasts with collagen I increases secretion of procathepsin B. *Journal of Biological Chemistry* **2002,** *277* (35), 32220-32227.

33. Zhu, L.; Kate, P.; Torchilin, V. P., Matrix metalloprotease 2-responsive multifunctional liposomal nanocarrier for enhanced tumor targeting. *ACS nano* **2012,** *6* (4), 3491-3498.

34. Yang, J.; Yang, Y.; Kawazoe, N.; Chen, G., Encapsulation of individual living cells with enzyme responsive polymer nanoshell. *Biomaterials* **2019,** *197*, 317-326.

35. Gadad, B. S.; Britton, G. B.; Rao, K., Targeting oligomers in neurodegenerative disorders: lessons from α-synuclein, tau, and amyloid-β peptide. *Journal of Alzheimer's Disease* **2011,** *24* (s2), 223-232.

36. Schmalfeldt, B.; Prechtel, D.; Härting, K.; Späthe, K.; Rutke, S.; Konik, E.; Fridman, R.; Berger, U.; Schmitt, M.; Kuhn, W., Increased expression of matrix metalloproteinases (MMP)-2, MMP-9, and the urokinase-type plasminogen activator is associated with progression from benign to advanced ovarian cancer. *Clinical Cancer Research* **2001,** *7* (8), 2396-2404.

Supporting information for:

Multifunctional amyloid oligomeric nanoparticles for specific cell targeting and drug delivery

*Weiqiang Wang[1], Rafayel A. Azizyan[2,3], Adriana Garro[4], Andrey V. Kajava[2,3] and Salvador Ventura[1, *]*

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

2 Centre de Recherche en Biologie cellulaire de Montpellier, UMR 5237 CNRS, Université Montpellier, Montpellier, France.

3 Institut de Biologie Computationnelle, Université Montpellier, Montpellier, France.

4 Universidad Nacional de San Luis IMASL-CONICET, San Luis, Argentina.

E-mail: salvador.ventura@uab.es

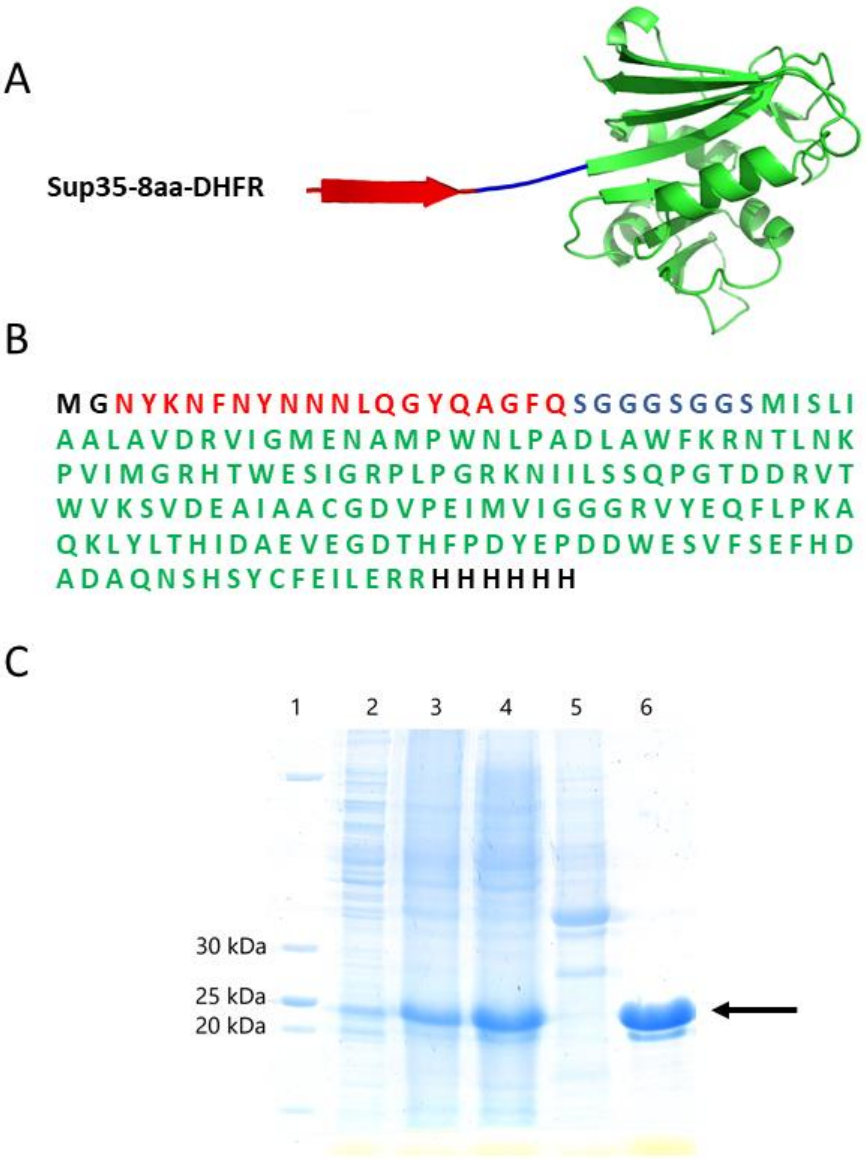

Figure S1. Schematic representation, sequence, expression and purification of the Sup35-8aa-DHFR fusion protein. (A) Sup35-8aa-DHFR with Sup35 soft amyloid core (SAC) (residues 100-118) fused to *Escherichia coli* DHFR (PDB: 7DFR) shown in cartoon representation. (B) Sequence of the Sup35-8aa-DHFR. The SAC, spacer linkers, globular structure and His6 tag are shown in red, blue, green and black, respectively. (C) SDS-PAGE analysis of the expression and purification of Sup35-8aa-DHFR. *Lane 1*, corresponds to the molecular weight marker, *lane 2,* non-induced culture, *lane 3*, total extract of induced culture, *lane 4*, soluble fraction (supernatant), *lane 5*, insoluble fraction (pellet) and, *lane 6,* shows purified Sup35-8aa-DHFR by gel filtration. A black arrow indicates the band corresponding to Sup35-8aa-DHFR.

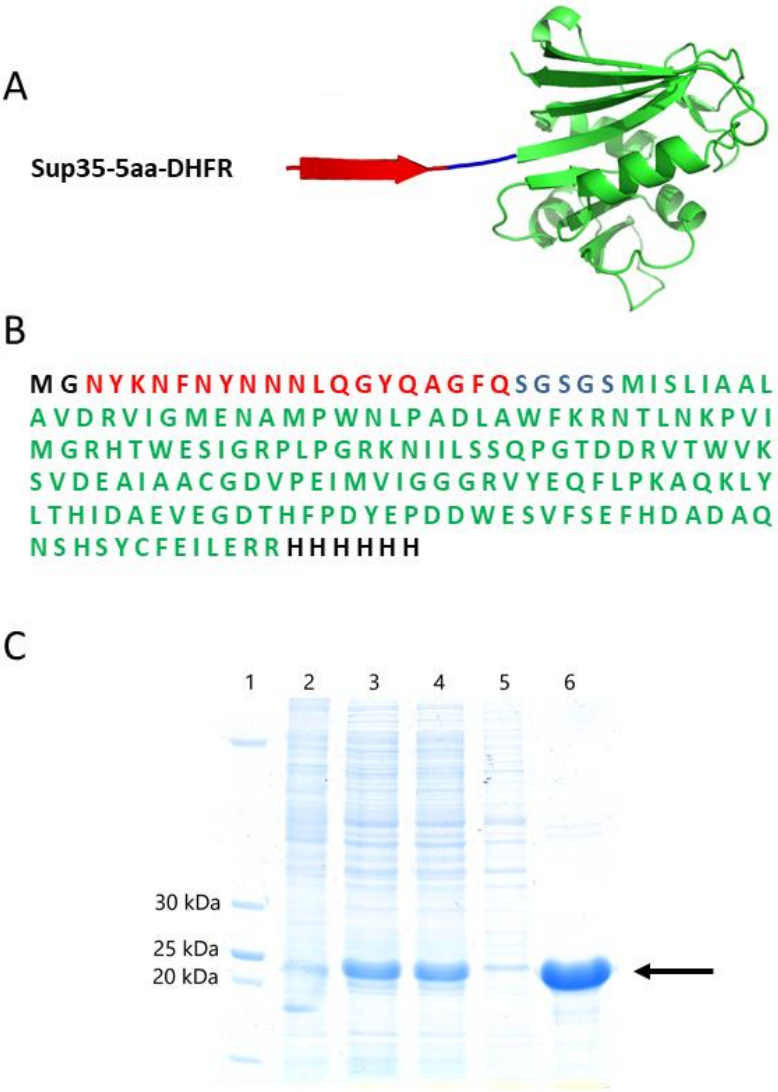

Figure S2. Schematic representation, sequence, expression and purification of the Sup35-5aa-DHFR fusion protein. (A) Sup35-5aa-DHFR with Sup35 soft amyloid core (SAC) (residues 100-118) fused to *Escherichia coli* DHFR (PDB: 7DFR) shown in cartoon representation. (B) Sequence of the Sup35-5aa-DHFR. The SAC, spacer linkers, globular structure and His6 tag are shown in red, blue, green and black, respectively. (C) SDS-PAGE analysis of the expression and purification of Sup35-5aa-DHFR. *Lane 1*, corresponds to the molecular weight marker, *lane 2,* non-induced culture, *lane 3*, total extract of induced culture, *lane 4*, soluble fraction (supernatant), *lane 5*, insoluble fraction (pellet) and, *lane 6,* shows purified Sup35-5aa-DHFR by gel filtration. A black arrow indicates the band corresponding to Sup35-5aa-DHFR.

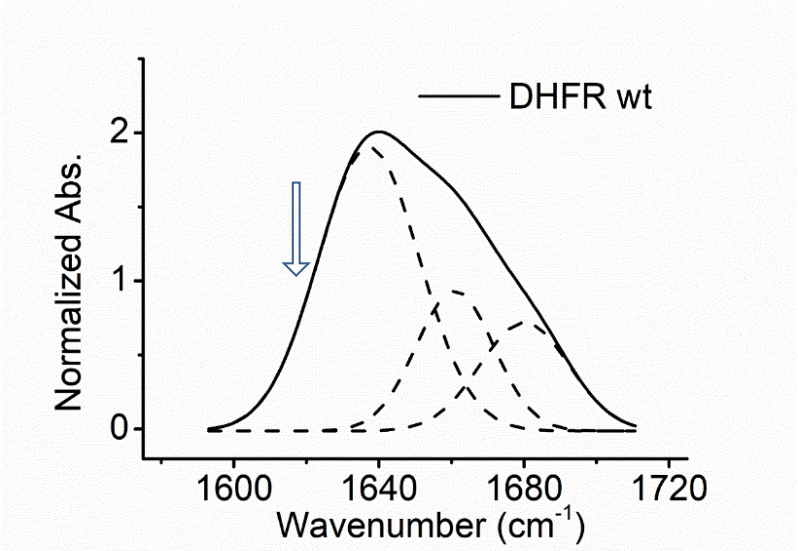

Figure S3. FTIR absorbance spectrum of incubated DHFR-wt in the amide I region (solid black line) and the components bands (dashed lines). The position of a potential intermolecular β-sheet component is indicated by a blue arrow.

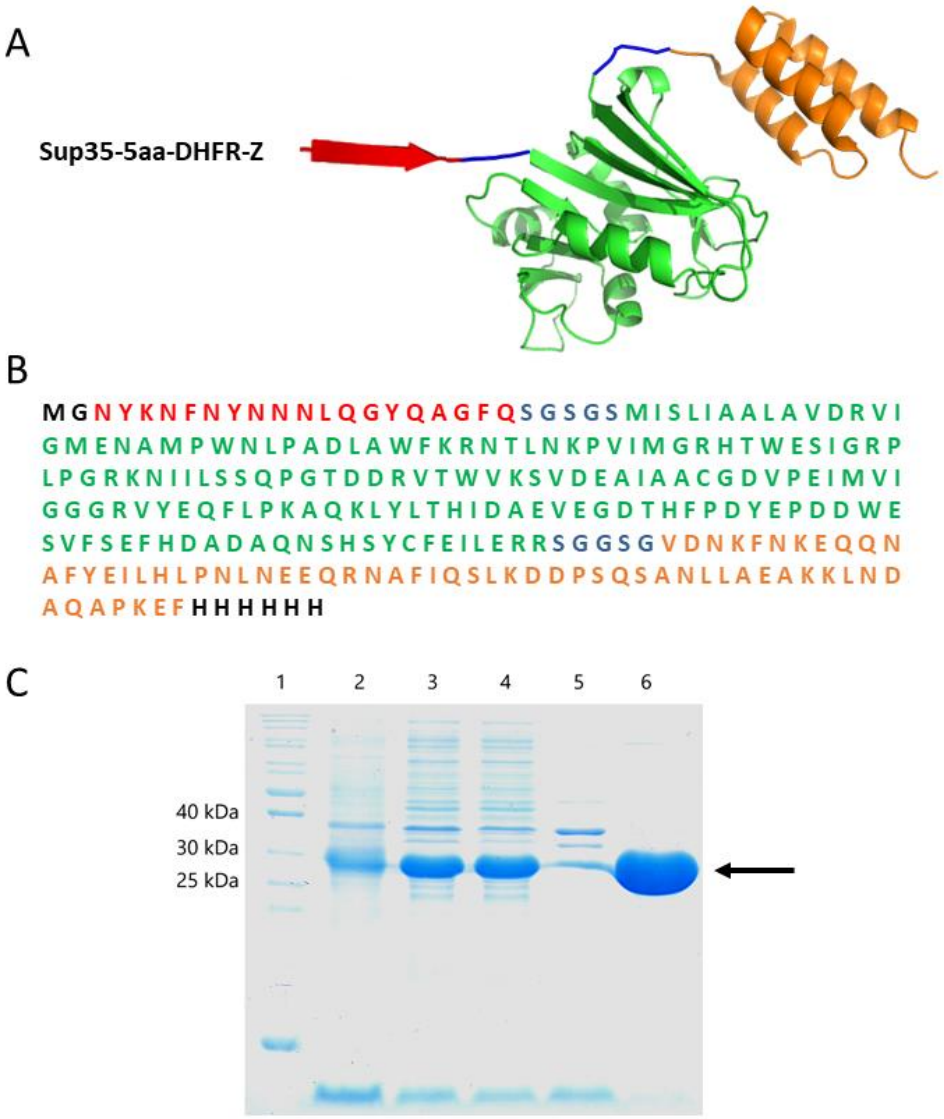

Figure S4. Schematic representation, sequence, expression and purification of the Sup35-5aa-DHFR-Z tandem fusion protein. (A) Sup35-5aa-DHFR-Z with Sup35 soft amyloid core (residues 100-118) fused to *Escherichia coli* DHFR (PDB: 7DFR) followed by a Z domain (PDB: 1Q2N) of *Staphylococcus aureus* protein A shown in cartoon representation. (B) Sequence of the Sup35-5aa-DHFR-Z. The SAC, spacer linkers, globular structures and His6 tag are shown in red, blue, green, orange and black, respectively. (C) SDS-PAGE analysis of the expression and purification of Sup35-5aa-DHFR-Z. *Lane 1*, corresponds to the molecular weight marker, *lane 2,* non-induced culture, *lane 3*, total extract of induced culture, *lane 4*, soluble fraction (supernatant), *lane 5*, insoluble fraction (pellet) and, *lane 6,* shows purified Sup35-5aa-DHFR-Z by gel filtration. A black arrow indicates the band corresponding to Sup35-5aa-DHFR-Z.

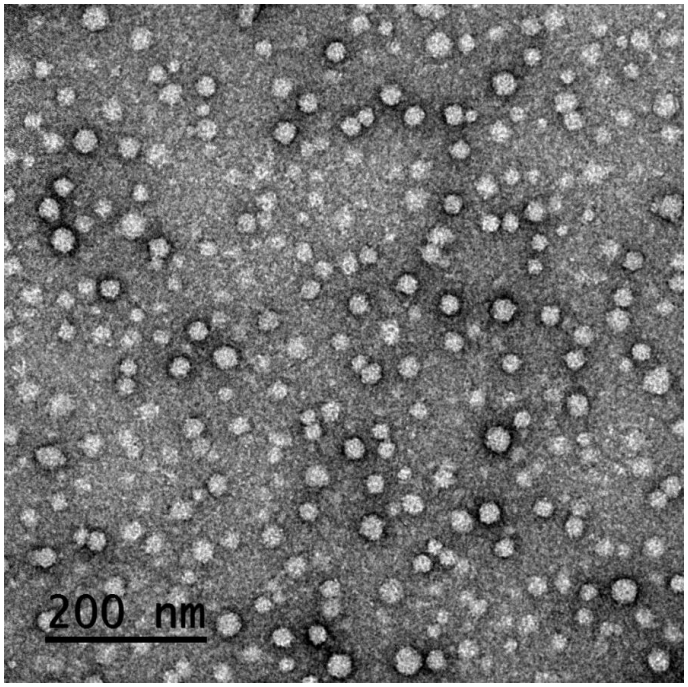

Figure S5. High magnification TEM image of Sup35-5aa-DHFR-Z nanoparticles. Scale bar represents 200 nm.

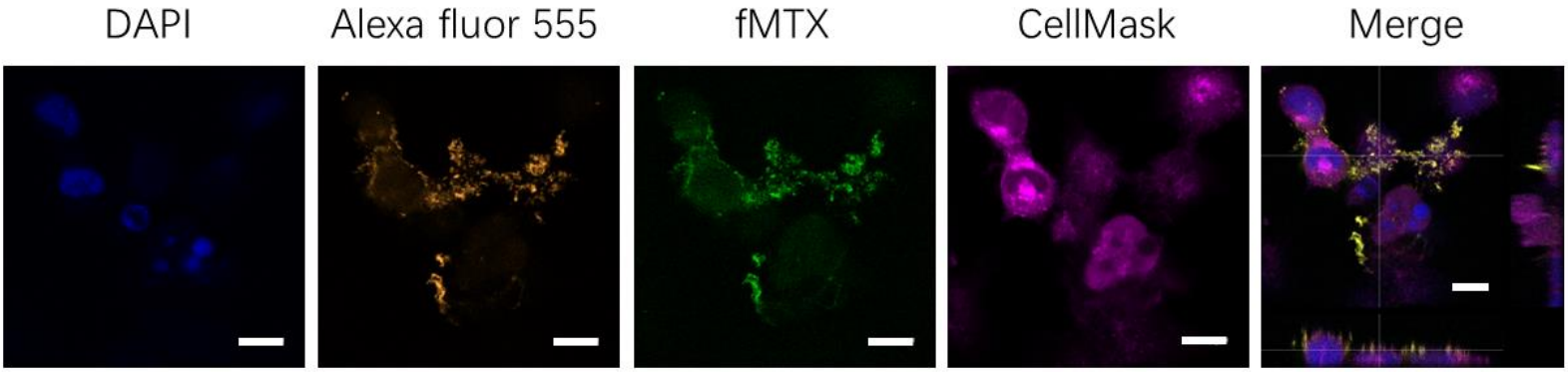

Figure S6. High magnification confocal microscopy images of HeLa cells incubated with fMTX and anti-EGFR antibody loaded nanoparticles (NPs- fMTX-anti-EGFR). The cross-sectional projection in the merged image shows the NPs- fMTX-anti-EGFR target the cell membrane. The cell membrane was stained with CellMask Deep Red and the nuclei was stained with DAPI, respectively. Scale bar represents 10 μm.

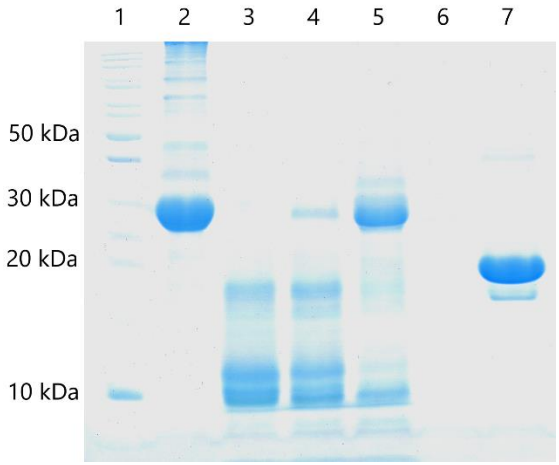

Figure S7. SDS-PAGE analysis of proteinase K digestion of Sup35-5aa-DHFR-Z nanoparticles. *Lane 1*, corresponds to molecular weight marker, *lane 2*, nanoparticles without digestion *lane 3*, 45 min digestion, *lane 4*, 30 min digestion, *lane 5*, 15 min digestion, *lane 6*, proteinase K, and *lane 7*, DHFR wt.

Table S1. Assignment and area of the secondary structure components of different assemblies in the amide I region of the FTIR spectra. DHFR-wt was used as a control.

Assignments (%)	**Sup35-8aa-DHFR fibrils**	**Sup35-5aa-DHFR nanoparticles**	**Sup35-5aa-DHFR-Z nanoparticles**	**DHFR-wt**
Inter β-sheet	47.04 (1620 cm^{-1})	14.78 (1622 cm^{-1})	33.32 (1621 cm^{-1})	-
β-sheet/α-helix	31.10 (1643 cm^{-1})	52.44 (1642 cm^{-1})	60.64 (1646 cm^{-1})	58.31 (1640 cm^{-1})
Turns	19.19 (1667 cm^{-1})	28.04 (1668 cm^{-1})	5.38 (1676 cm^{-1})	22.04 (1662 cm^{-1})
β-sheet	2.67 (1687 cm^{-1})	4.74 (1684 cm^{-1})	0.66 (1689 cm^{-1})	19.65 (1683 cm^{-1})

Table S2. Kinetic and thermodynamic dissociation constants of Sup35-5aa-DHFR nanoparticles and Sup35-8aa-DHFR fibrils.

Parameters	Sup35-5aa-DHFR nanoparticles	Sup35-8aa-DHFR fibrils
V_{max} (nM min^{-1})	1712±25	1745±21
K_m (μM)	6.23±0.32	9.44±0.41
K_{cat} (s^{-1})	1.43±0.02	1.45±0.02
K_{cat}/K_m (μM^{-1} s^{-1})	0.23±0.01	0.15±0.01
K_d (μM)	1.58±0.07	6.26±0.44

Chapter IV

Dual antibody-conjugated amyloid nanorods to promote selective interactions between different cell types

*Weiqiang Wang[1] and Salvador Ventura[1]**

[1]Institut de Biotecnologia i de Biomedicina and Departament de Bioquímica i Biologia Molecular; Universitat Autònoma de Barcelona; 08193 Bellaterra (Barcelona), Spain.

E-mail: salvador.ventura@uab.es

Abstract

The well-known enhanced permeability and retention (EPR) effect enable nanostructured materials to be promising scaffold in addressing many challenges encountered by small molecules in the treatment of diseases. The grafting biomolecules on the surface allow to improve the therapeutic and diagnosis efficacy by specific targeting to pathogenic cells. However, the monofunctional nanomaterials for a specific target by conjugating only one type of ligand are less succeeded in many diseases, which required two or more targets/receptors have to be targeted and activated with one object spontaneously. Therefore, multivalent nanomaterials for dual- or multi-targeting has been suggested as an emerging proof-of-concept in future nanomedicine. Amyloid-based functional nanomaterials have been widely used for biological and biomedical applications since their bioactive, biodegradable and biocompatible properties. The main advantage of such protein-based materials is the nature of the decorated globular proteins, which allows to alter the material functionality by genetic redesign to fit the intended application. Here we exploited a modular approach to generate functional amyloid fibrils decorated with an antibody capture moiety. We show here the high antibody binding affinity and capacity of the resulting nanofibrils. We further engineered the size of nanofibrils with a simple physical procedure and obtain amyloid nanorods. We further show such nanorods can be exploited for antibody-directed targeting of specific cells and association of different cell types. Overall, the novel antibody capture nanofibrils or nanorods exhibit a high potential in designing dual- or multi-targeting materials by decorating different antibodies of interest.

Keywords: Amyloid, dual- or multi-targeting, multivalency, antibody, nanorods, nanomaterials.

Introduction

Nanomaterials with large surface/volume ratios, tunable and diverse physical properties, and multiple surface functional groups, have emerged as an novel potential platform for diagnosis and therapy of diseases.[1] In comparison with small molecules, these nanomaterials, such as supramacromolecules, nanotubes, micelles, and protein-polymers conjugates, are able to efficiently accumulate in tumor sites with higher concentration and longer time, namely the enhanced permeability and retention (EPR) effect.[2] Therefore, such unique pharmacokinetic property of nanomaterials enable them to be promising scaffold in addressing many challenges encountered by small molecules in biomedical applications.[3] Moreover, the multiple surface functional groups usually provide a unique space for grafting biomolecules.[4] Then, the introduce of tailored ligands specifically targeting pathogenic cells onto the surface of nanomaterials will not only minimize the toxic side effects of the materials but also improve the therapeutic efficacy by selective targeting. So far, most of efforts have been focused on the synthetic monofunctional nanomaterials for a specific target by conjugating only one type of ligand, such as RGD peptides,[5] monoclonal antibody[6] and other proteins.[7] As a consequence, limited success were discovered to some cases, because there are many diseases might attribute to multiple factors or some treatment involved in a cascade of reactions, which are not able to be carried out by using monospecific conjugates.[8] Therefore, multivalency is required for some biological process and two or more targets/receptors have to be targeted and activated with one object in the treatment of such diseases.[9,10] In fact, the concept of dual targeting has been initially applied to the bispecific antibodies (BsAbs), in which the two different variable regions simultaneously addressing different antigens or epitopes, involved in the inhibition of two cell surface receptors, cross-linking two receptors, blocking of two ligands or recruitment of T cells to cancer cells,[11] result in a strongly increased targeting and therapeutic efficacy.[12] However, the issues including low yields,[13] molecular heterogeneity,[14] short half-time *in vivo*,[15] and toxic side effects[16,17] have limited the clinical applications of BsAbs. Alternatively, dual targeting nanoparticles, conjugating

two different small molecules,[18] peptides,[1] monoclonal antibodies[19] or recognizable proteins,[20] have exhibited high potential in bionanotechnology and therapeutic use. Thus, nanomaterials decorated two or more ligands with tailored ratio for dual or multi targeting have been suggested as an emerging proof-of-concept in future nanomedicine.[20]

Recently, amyloids with biological functions has inspired the building up of functionalized nanomaterials.[21] These bioactive, biodegradable and biocompatible peptide or protein-based nanomaterials, have been widely used for biological and biomedical applications, ranging from cancer therapy, bioimaging or tissue engineering to regenerative medicine.[22] Self-assembled peptide-based nanomaterials offer a high surface area versus volume ratio and hold stable superstructures as well as fascinating biological effects, such as improved blood circulation time, better targetability and so on.[23] The traditional nanomaterials are usually synthesised with series of complex process, which might result in a large number of decorated ligands losing their binding affinity and accessibility.[24] In contrast, the main advantage of protein-based materials is the nature of globular proteins, which allows to alter the material functionality by genetic redesign to fit the intended application. Therefore, the amyloid-based nanomaterials are ideal scaffolds to decorate interested protein ligands and build up a hybrid complex possessing dual- or multi-targeting ability.

We have recently succeeded in designing highly ordered amyloid-like nanofibrils containing properly folded and highly active proteins using a modular strategy. In particular, a Soft Amyloid Core (SAC) was used as driving force for self-assembling, and fused to any globular proteins of interest. The fusion protein are produced in soluble with high yield, but still can be induced into highly ordered fibrillar structure, in which the SAC forms the core of amyloid fibrils and the globular domains hang from it in a folded state.[25] Thus, the appended globular protein is highly bioactive and accessible to the targeted objectives when embedded in the fibrils. A similar approach has been applied to manufacture functionalized nanofibrils decorated with a Z-domain,[26] an engineered analog of B domain of *Staphylococcus aureus* protein A.[27,28] The resulting

amyloid fibrils exhibited a high binding affinity and capture capacity for antibody. In addition, we have recently described an amyloid oligomers containing a Z-domain, specifically targeting pathogenic cells, when conjugated with corresponding monoclonal antibody (mAb). Altogether, It suggests that such hybrid composition of amyloid fibrils containing a Z-domain exhibit a high potential in dual- or multi-targeting when decorated with different mAbs. Ideally, the decorated two or more different monospecific antibodies on nanomaterials enable it to target or activate different antigens (receptors, ligands, molecules) spontaneously, displaying a dual- or multi-targeting functionality resembling BsAbs.

In this study, we fused a Z-domain to Sup35-SAC. The fusion protein was expected to be induced into self-assembled amyloid nanofibrils, in which the Z-domain remains its folded conformation (**Scheme 1**). The resulting amyloid fibrils are stable and exhibited a high capture capacity for antibody in serum. Thus, the amyloid fibrils can be decorated with any antibody of interest. The diverse and tuneable size and shape of nanomaterials can increase their dispersion, cellular uptake and delivery efficacy.[29] To exploit the potential of such nanofibrils for biomedical applications, we further engineered the size of nanofibrils with a simple physical procedure and obtain amyloid nanorods which are biocompatible. The single mAb conjugated nanorods can be directed specifically against and activate cells expressing the relative receptors. While double mAbs (i.e. anti-EGFR and anti-CD3 antibody) coupled nanorods complex can redirect CD3 expressing T cells to EGFR expressing tumor cells. Ideally, the targeted T cells will be activated and kill the tumor cells spontaneously. Overall, the high antibody capacity, high stability, non-toxicity and high homogeneity of Z-domain containing amyloid nanorods enable it to be a good alternative scaffold for dual- or multi-targeting, which shows a high potential in immunotherapy. The simple, modular and straightforward strategy described here can be adapted to build up other materials for dual- or multi- targeting any types of objectives by decorating the relative mAbs.

Results and Discussion

Design of a fusion protein to build up antibody capturing nanofibrils

In order to generate functional amyloid fibrils with antibody capturing activity, we fused Sup35-SAC to the Z-domain,[26] an engineered analog of the B domain of *Staphylococcus aureus* protein A (**Figure S1**).[27] The Z-domain consists of 58-residues (6.5 kDa) and folds into a bundle-like composed of three α-helices. In contrast with the larger GFP and carbonic anhydrase proteins, which required a separation from Sup35 SAC of at least 8-residues to form ordered fibrils,[25] molecular modeling[30] suggested that a 5-residue flexible linker (SGSGS) should suffice to allow amyloid fibril formation without significant steric constraints for the Z-domain. This will reduce the entropic cost of immobilizing the disordered N-terminus of the protein fusion in a potential amyloid structure. The Z-domain binds with high affinity to the Fc region of antibodies from different species and subclasses. The strategy was that if Sup35-Z would assemble into amyloid fibrils, we might decorate these nanostructures with any antibody of interest.

Sup35-SAC does not affect the solubility, conformation and thermodynamic stability of the Z-domain

A requirement to use Sup35-Z for building an antibody-capturing nanomaterial is that the N-terminal Sup35-SAC does not alter the solubility, native structure, and stability of the adjacent Z-domain, and, therefore, it does not impact the potential antibody binding activity of the fusion.

We expressed the Sup35-Z fusion protein (10 kDa) in *E.coli*. The protein was localized entirely in the soluble cell fraction, from which it was purified, and it was produced at high yield (62 mg/L) (**Figure S2A**). Then we compared the secondary structure content of purified Sup35-Z and the Z-domain alone (Z-domain) at pH 7.4 and 25 °C by monitoring their far-UV CD spectra. The spectra of the two proteins resemble each other and are dominated by typical α-helical signals (**Figure S2B**). The

thermal unfolding of both proteins at pH 7.4 and 25 °C was followed by monitoring the changes in ellipticity at 222 nm, which reports on the stability of the Z-domain α-helical structure (**Figure S2C**). The obtained melting curves were similar, with a single cooperative transition being observed, consistent with a two-state unfolding reaction. In both cases, the Z-domain was highly stable and not completely denatured, even at 90 °C. Fitting of the data to a two-state reaction rendered apparent melting temperatures of 74.7 ± 0.8 °C and 74.5±1.0 °C for Sup35-Z and the Z-domain alone, respectively. All these data indicate that, as intended, the Sup35 SAC does not affect the solubility, conformation, and stability of the adjacent globular domain, consistent with our previous studies on other protein folds.[25]

However, we should also discard that the N-terminal exogenous sequence's presence causes steric impediments for antibody binding to the Z-domain or that its potential fluctuations hide the antibody binding site. To exclude these possibilities, we used soluble Sup35-Z and the Z-domain to purify IgG antibodies from a complex matrix, such as bovine blood serum. The proteins were immobilized in NI-NTA columns through their respective His6 tags, and the serum was chromatographed. The identity of the purified proteins was analyzed by SDS-PAGE, which, for both proteins, revealed the presence of three major bands at ~75 kDa, ~50 kDa and ~25 kDa, corresponding to IgGs and their heavy and light chains, respectively, without unspecific binding to highly abundant serum proteins, like serum albumin (**Figure S3**). The data suggest that soluble Sup35-Z captures IgGs from serum with an efficiency comparable to that of the Z-domain.

Sup35-SAC induces the assembly of the Sup35-Z fusion protein into amyloid fibrils

We used the amyloid-specific dyes Thioflavin-T (Th-T) and Congo Red (CR) to assess if the Sup35-Z protein fusion self-assembles into amyloid-like structures under mild conditions. To this aim, Sup35-Z and Z-domain were incubated at pH 7.4 and 37 ºC for 5 days. Th-T is a dye in which fluorescence emission maximum at 488 nm

increases in the presence of amyloid-like structures.[31] The presence of incubated Sup35-Z promoted a large increase of Th-T fluorescence emission signal, whereas the Z-domain incubated in the same conditions had a negligible effect (**Figure 1A**). In agreement with these results, CR binding was observed for Sup35-Z, resulting in a clear red-shift of CR absorption spectrum, indicative of the dye binding to an amyloid structure,[32] whereas the Z-domain did not promote any spectral shift (**Figure 1B**). The morphological analysis of the two protein solutions by negative-staining and transmission electron microscopy (TEM) confirmed the presence of typical long and unbranched amyloid fibrils of 12.7±0.7 nm in width for Sup35-Z (**Figure 1D**). In contrast, the Z-domain solution did not show any detectable ordered aggregates (**Figure 1C**).

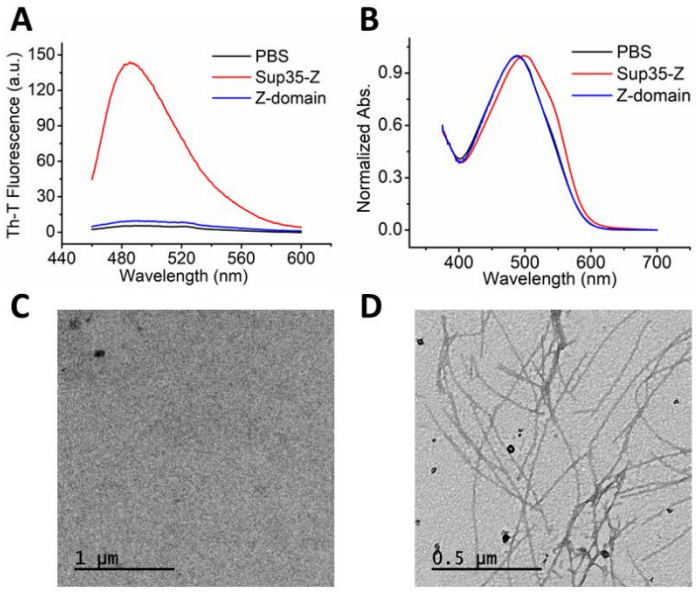

Figure 1. Characterization of Sup35-Z fibrils. Sup35-Z and Z-domain solutions were incubated for 5 days and analyzed by measuring (A) Th-T fluorescence emission and (B) Congo red absorbance. Z-domain and Sup35-Z are shown in blue and red, respectively. PBS without protein was included as a control (black line). Representative TEM micrographs of incubated proteins upon negative staining: (C) Z-domain and, (D) Sup35-Z. The scale bar represents 1 µm and 0.5 µm, respectively.

To assess if, as previously described for other protein folds,[25] the Z-domain remained folded in the Sup35-Z assembled state, we characterized the secondary structure content of Sup35-Z amyloid fibrils using Attenuated Total Reflectance Fourier Transform Infrared spectroscopy (ATR-FTIR). The all-alpha fold of the Z-domain should allow us to track its native state when embedded in the amyloid fibrillar structure known to be β-sheet enriched. We recorded the fibrils' infrared spectra in the amide I region of the spectrum (1700-1600 cm^{-1}), corresponding to the absorption of the carbonyl peptide bond group of the protein main chain. The spectra's deconvolution allowed us to assign the secondary structure elements and their relative contribution to the primary signal (**Figure S4** and **Table S1**). The spectra displayed two major signals assignable to the contribution of intermolecular *β*-sheets (1626 cm^{-1}) and α-helices (1654 cm^{-1}), which likely arise from the Sup35-SAC amyloid spine and helical Z-domains, respectively.

All these data indicated that Sup35-SAC is both necessary and sufficient to promote the self-assembly of the Sup35-Z fusion into amyloid fibrils, were the Z-domain remains in a folded conformation.

Antibody binding activity and capacity of Sup35-Z amyloid fibrils

Molecular modeling[30] suggests that folded Z-domains would be exposed in the periphery of the amyloid fibrils and thus accessible to antibodies, allowing to obtain nanostructures enriched in any antibody of interest. To confirm that this was the case, we incubated preformed Sup35-Z fibrils and fibrils formed by the Sup35-SAC peptide alone,[33] with 2 μg of a secondary IgG antibody labeled with Alexa 488 at room temperature for 30 min. Then they were precipitated and washed three times to remove any unbound IgG and resuspended in PBS buffer. When imaged using fluorescence microscopy and a FITC filter (excitation at 465-495 nm), highly fluorescent particles were observed for Sup35-Z fibrils, whereas Sup35-SAC fibrils were devoid of fluorescence (**Figure 2A** and **2B)**. To further determine the Z-domain's antibody capture capacity when embedded in the fibrils, we incubated the green-labeled

secondary antibody with fibrils in the range of 0-0.4 μM. Then we recorded the fluorescence emission spectra of the incubated fibrils after precipitation and washing. Incubated Sup35-Z fibrils exhibited a fluoresce maximum at ~518 nm, which is also observed in a labeled-antibody solution, whereas Sup35-SAC fibrils did not exhibit any significant Alexa 488 fluorescence signal (**Figure 2C**). A titration of the fluorescence of Alexa 488 as a function of incubated fibrils indicated that Sup35-Z fibrils have a binding capacity of ~2.5 μg IgG per μg of fibrils (**Figure 2D**).

To further confirm the antibody binding affinity of Sup35-Z fibrils in a complex matrix, we incubated the fibrils with bovine blood serum for 30 min, followed by precipitation and washing steps and elution of the fibril-bound protein with glycine-HCl buffer pH 3.0. The analysis by SDS-PAGE confirmed that the Sup35-Z fibrils bind preferentially to IgGs, as evidenced by the band's correspondent to the heavy and light chains in the gel with little contamination of abundant proteins such as serum albumin (**Figure S5**). To assess if Sup35-Z fibrils are stable in physiological conditions, a requirement for biomedical applications, we incubated the Sup35-Z fibrils in bovine blood serum for up to 3 days. SDS-PAGE analysis indicated that the fibrils are stable and not degraded in these conditions (**Figure S6**).

Overall the data in this section indicate that Sup35-Z fibrils display a remarkable antibody capturing activity in both defined and complex media and that this property is not due to unspecific binding to the amyloid macromolecular structure but to the folded Z-domains in the fibrils.

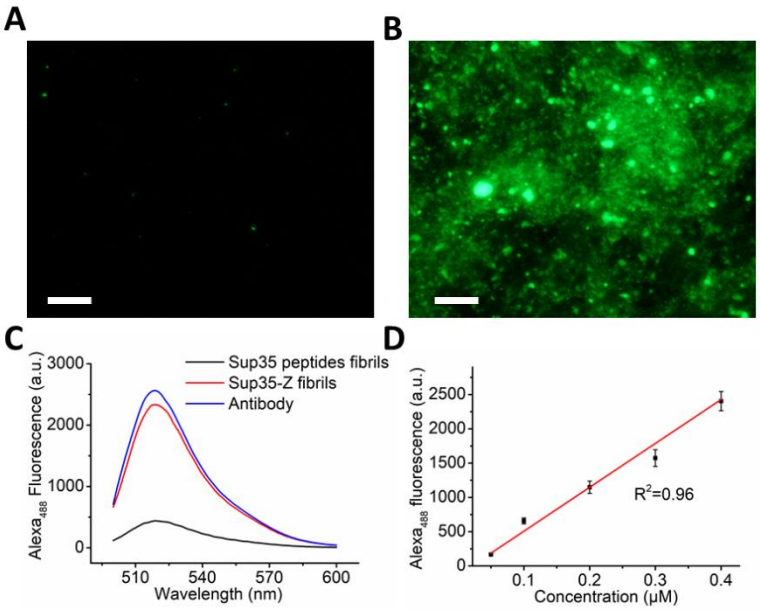

Figure 2. Antibody binding affinity of Sup35-Z fibrils. Representative fluorescence microscopy image of fibrils incubated with single IgG labeled Alexa 488: (A) Sup35 peptide fibrils and, (B) Sup35-Z fibrils. The scale bar represents 50 μm. (C) Fluorescence emission spectra of incubated fibrils at 0.4 μM. The blue line represents the fluorescence spectra of the antibody alone. (D) A linear plot of the fluorescence intensity of incubated fibrils as a function of the concentration of the fibrils.

Accessibility and functionality of the conjugated antibody on Sup35-Z fibrils

Another requirement to build up functional antibody-conjugated nanofibrils is that the antibody displayed in Sup35-Z fibrils keeps its intact structure and can target the desired antigen epitope. To assess if this is the case, we incubated Sup35-Z fibrils with a mouse anti-GFP IgG antibody. The antibody-bound fibrils were then incubated with soluble GFP for 30 min, precipitated, and washed 3 times to eliminate any unbound GFP. The presence of green fluorescent aggregates, as imaged by fluorescence microscopy, indicated that the antibody-conjugated fibrils target the intended antigen (**Figure S7B**), whereas GFP does not bind to Sup35-Z fibrils if they are not previously incubated with the antibody (**Figure S7A**).

On the other hand, we incubated the mouse anti-GFP IgG bound Sup35-Z fibrils

with a goat anti-mouse IgG labeled with Alexa 555 and measured the resulting fluorescence spectra after fibril precipitation and washing. Sup35-Z fibrils were also incubated directly with the secondary Alexa 555-labeled antibody and treated in the same way. Fibrils incubated with the primary and secondary antibodies exhibiting a much higher fluorescence maximum at ~570 nm that fibrils incubated directly with secondary IgG (**Figure S7C**).

Overall, the data indicated that the IgG bound to the fibrils, kept its intact structure, binds its antigen, and can be targeted by a specific secondary antibody, resulting in a significant amplification of the fluorescence signal.

Dual antibody binding to Sup35-Z nanofibrils

In principle, the Sup35-Z fibrils could be endorsed with multivalence by conjugating them simultaneously with different antibodies. To test if this was possible, the fibrils were incubated simultaneously with two different antibodies labeled either with Alexa 488 or Alexa 555. After precipitation and washing, the fibrils were imaged using fluorescence microscopy an a FITC filter (excitation at 465-495 nm) or a TxRed filter (excitation at 540-580 nm). The particles appeared green and red in the respective channels, and the two signals overlapped when the channels were merged (**Figure 3, upper panel**). In contrast, Sup35-SAC fibrils incubated with the two antibodies, in the same way, did not exhibit any significant fluorescence (**Figure 3, bottom panel**). Thus, the data indicate that the Sup35-Z fibrils can be multi-functionalized specifically. Controlling the proportion of each antibody in the initial mixture should allow obtaining fibrils decorated with the desired ratio of them.

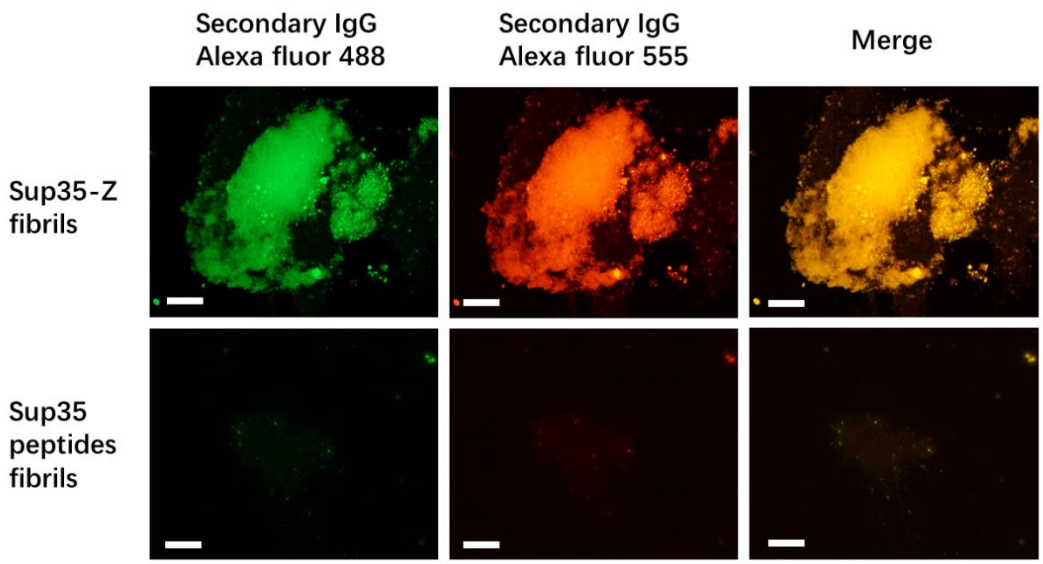

Figure 3. Double antibody binding of Sup35-Z fibrils. The representative fluorescence image of the Sup35-Z fibrils (upper panel) and Sup35 peptides fibrils (bottom panel) incubated with two secondary antibodies: rabbit anti-mouse IgG labeled Alexa 488, and goat-anti mouse IgG labeled Alexa 555. The scale bar represents 50 μm.

Sup35-Z nanorods are biocompatible

The size and shape of nanomaterials impact their dispersion, cellular uptake, and delivery efficacy.[29] We sought to generate shorter versions of our functional amyloid fibrils that can be employed as nanoparticles. To this aim, we sonicated the fibrils shortly and obtained relatively homogeneous rod-like nanostructures of 50-100 nm in length, as visualized by TEM (**Figure 4**).

One of the main limitations for the use of amyloid-like materials in biomedical applications is that they might possess and inherent cytotoxic activity[34]. The toxicity is associated with oligomeric assemblies, rather than to mature fibrils, but it is unknown if mechanical shearing of mature fibrils might render toxic particles. To discard this possibility, we tested the cytotoxicity of the Sup35-Z nanorods at different concentrations, ranging from 1 μM to 25 μM, using the PrestoBlue assay (**Figure S8**). The statistical analysis using a one-way ANOVA test indicated that the particles did

not exhibit significant toxicity for human HeLa cells, suggesting that they would have excellent biocompatibility.

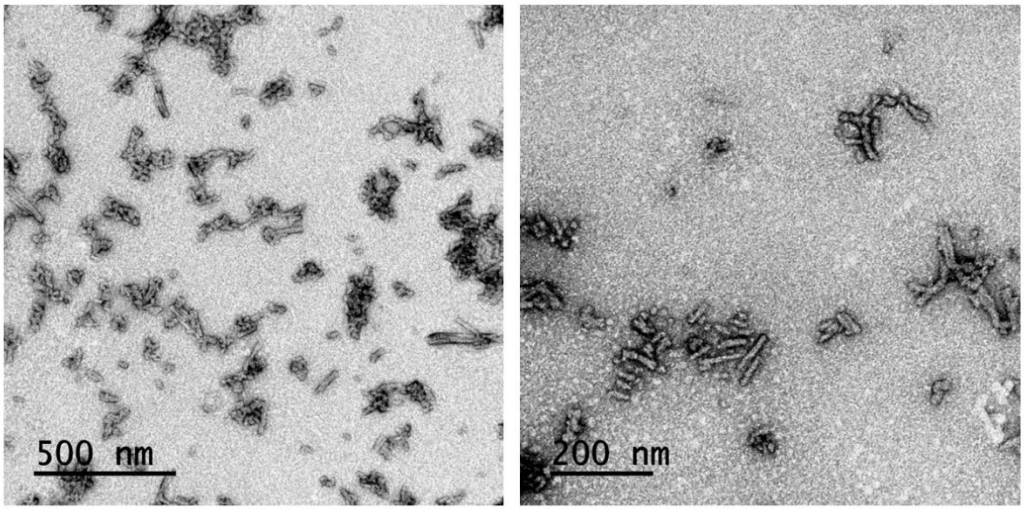

Figure 4. Representative TEM micrographs of sonicated Sup35-Z amyloid fibrils upon negative staining. The scale bar represents 500 nm and 200 nm, respectively.

Sup35-Z functionalized nanorods target human cells specifically.

We aimed to assess if Sup35-Z nanorods can target specific antigen in living cells, once they have been loaded with antibodies through their Z-domains. We decorated the nanorods with either an anti-EGFR antibody or an anti-CD3 antibody, labeled with Alexa 555 and Alexa 488, respectively, as described above. Anti-EGFR antibodies target the epidermal growth factor receptor (EGFR), which is highly expressed on the membrane of many epithelial cancer cells, such as HeLa cells. In contrast, anti-CD3 antibodies target the TCR/CD3 complex of T lymphocytes and consequently activate them.[35] First of all, we incubated the anti-EGFR antibody loaded nanorods (NRs-anti-EGFR) with HeLa cells. The majority of HeLa cells were red fluorescent when visualized by confocal microscopy, which indicated the NRs-anti-EGFR were able to recognize them. In contrast, when anti-CD3 antibody loaded Sup35-Z nanorods (NRs-anti-CD3) were incubated with HeLa cells, no cellular fluorescence was detected, consistent with the fact that this cell type does not express the CD3 complex (**Figure**

5). However, when the NRs-anti-CD3 were incubated with lymphocytes T, green fluorescent cells were. Thus, the data indicated that the recognition of human cells by antibody loaded nanorods was antibody-driven and specific.

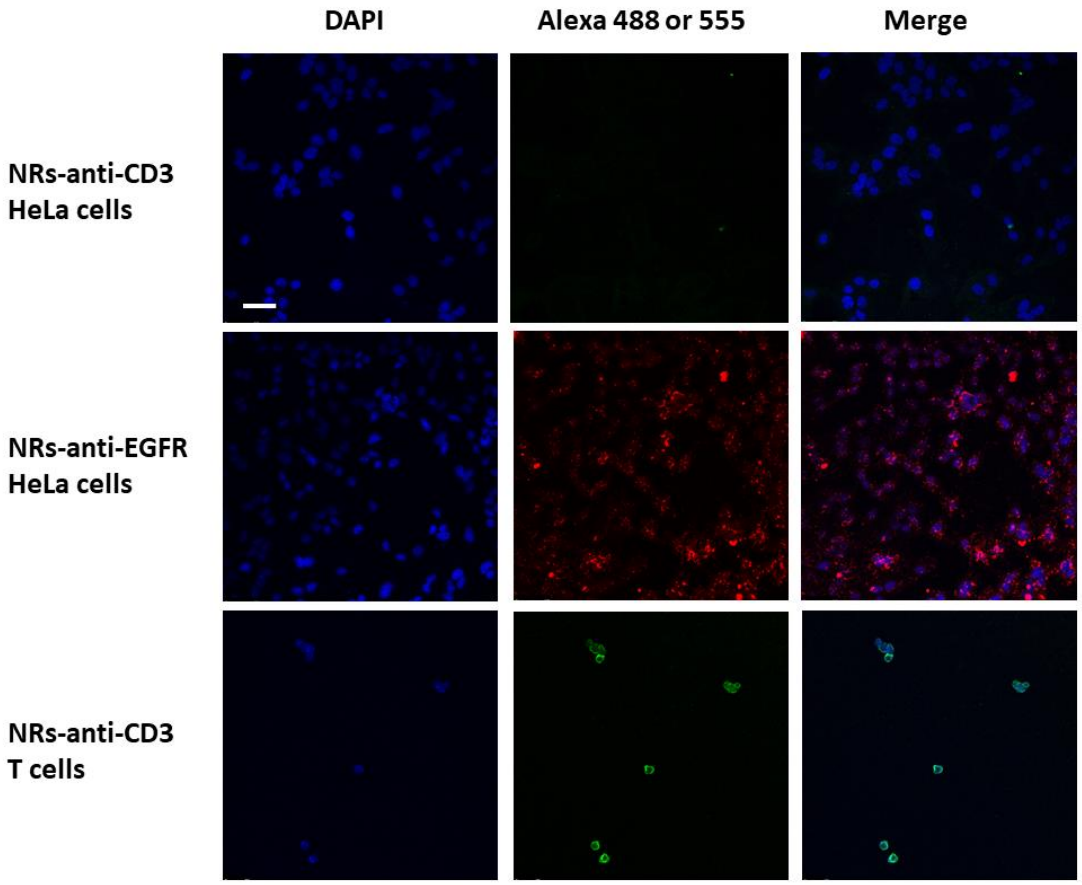

Figure 5. Binding specificity of functionalized Sup35-Z nanorods to human cells. Representative confocal microscopy images of HeLa cells incubated with nanorods conjugated with an anti-CD3 antibody (NRs-anti-CD3, Alexa 488) (upper panel) or an anti-EGFR antibody (NRs-anti-EGFR, Alexa 555) (middle panel), and lymphocyte T incubated with nanorods conjugated with an anti-CD3 antibody (NRs-anti-CD3, Alexa 488) (lower panel). The scale bar represents 50 μm.

To further assess if, besides targeting specifically T lymphocytes, NRs-anti-CD3 can activate them, we carried out a T cell proliferation assay implementing a modification of the typical antibody immobilization method [35] in which NRs-anti-CD3 acted as the antibody immobilizing agent. The T cell proliferation response was monitored using the PrestoBlue assay. The statistical analysis using a one-way ANOVA

test indicated that NRs-anti-CD3 significantly increase T cell proliferation, and at a level that is comparable or higher than the one resulting from incubation with the immobilized anti-CD3 antibody alone. In contrast, non-antibody loaded Sup35-Z nanorods had a negligible effect on T cell proliferation. (**Figure S9**). Therefore, the incorporated anti-CD3 antibody in the nanorods can efficiently target and significantly activate the T lymphocytes,

Dual antibody conjugated Sup35-Z nanorods direct T lymphocytes to HeLa cells.

We have shown that Sup35-Z fibrils can bind simultaneously to two different antibodies and that antibody-loaded nanorods can target specific cell types. This immediately suggested that these properties can be used to drive two different cell types in close proximity. To confirm this idea, we loaded Sup35-Z nanorods simultaneously with two antibodies, namely fluorescently labeled anti-EGFR and anti-CD3 (anti-EGFR-NRs-anti-CD3). We incubated the dual conjugated nanorods with HeLa cells for 20 min. Then the medium was removed, cells were rinsed with PBS, and T lymphocytes added. The mixture incubated for 20 min, after which the medium was again removed, cells cleaned, mounted, and imaged. The presence of circular T cells (25±5) and polygonal HeLa cells connected by yellow fluorescent nanostructures (merging the anti-EGFR and anti-CD3 fluorescence channels) was observed by confocal microscopy (**Figure 6, lower panel**). In contrast, Sup35-Z nanorods simultaneously loaded with anti-EGFR antibody and a fluorescent secondary anti-rabbit antibody (anti-EGFR-NRs-anti-rabbit) target HeLa cells but do not capture any T lymphocyte (**Figure 6, upper panel**). These data indicate that dual antibody loaded nanorods can bind at least two different antigens at the same time, and bring unrelated cell types spatially close. Thus, this nanomaterial can be applied for immunotherapy as a mimetic of bispecific antibodies (**Figure 7**), whose combination of binding activities can be tailored at will.

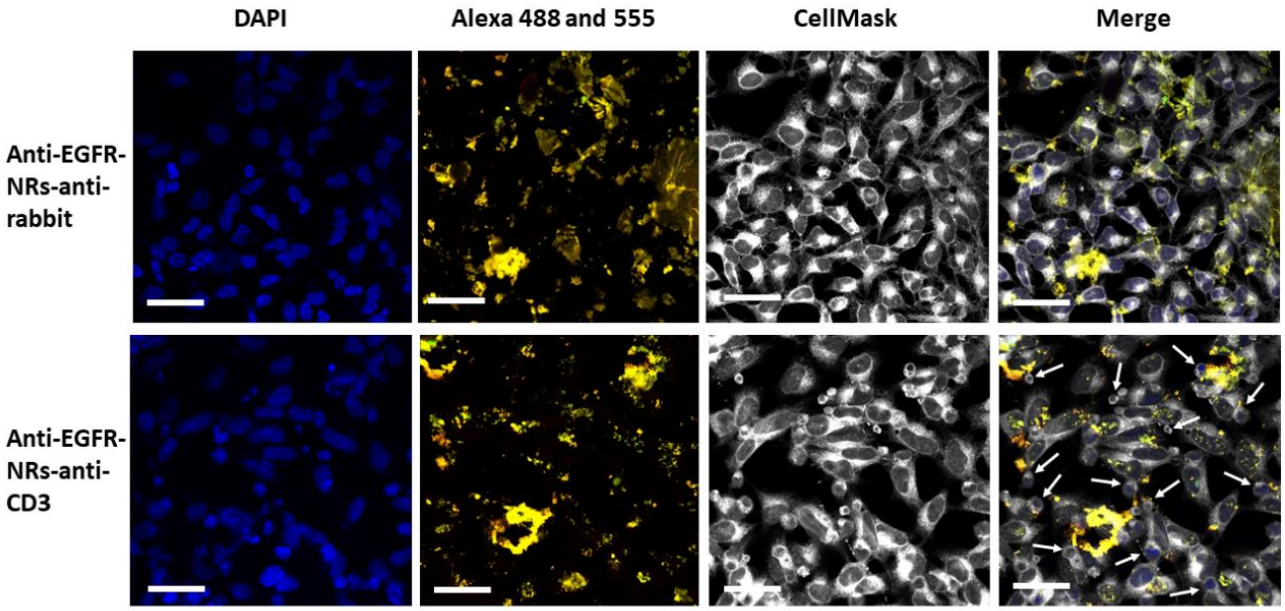

Figure 6. Double mAbs conjugated nanorods redirect the CD3 expressing T cells to EGFR expressing HeLa cells. Representative microscopy images of EGFR expressing HeLa cells and CD3 expressing T cells in the presence of anti-EGFR and anti-CD3 bound nanorods (anti-EGFR-NRs-anti-CD3, lower panel) and anti-EGFR and anti-rabbit bound nanorods (anti-EGFR-NRs-anti-rabbit, upper panel), respectively. The white arrows show the presence of lymphocyte T with circular and round shape. The anti-EGFR antibody and anti-CD3 antibody are labeled Alexa fluor 555 and Alexa fluor 488, respectively. The scale bar represents 50 μm.

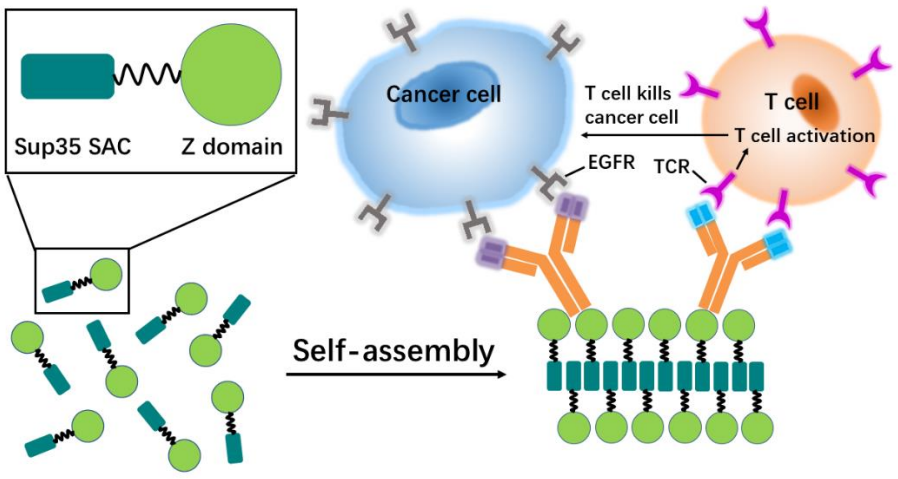

Figure 7. Schematic illustration on the dual-targeting functionality of mAbs-nanorods complex. The construct of Sup35-Z fusion consists of a Sup35 soft amyloid core (green square) and a Z-domain (green ball), that acts as an antibody capture domain, linked with a flexible linker (black line); Sup35-SAC induces the self-assembly of the fusion protein into antibody binding nanofibrils. Nanorods bound to two monoclonal antibodies (mAbs) direct the TCR/CD3 complex positive T cells to EGFR expressing tumor cells, and activated T cells will be able to kill the tumor cells.

Conclusions

We have built up an antibodies capturing amyloid fibrils by using a hybrid protein consisting of a Sup35 soft amyloid core (SAC) and globular protein Z-domain which holds a high affinity to antibodies. In agreement with our previous study, the Sup35-SAC allow the Z-domain to remain its native folded structure but also keep its natural functionality of binding antibody when embedded in nanofibrils. We further engineered the size of nanofibrils and obtain homogeneous nanorods. These homogenous nanorods are highly biocompatible for further biomedical applications. The monospecific antibodies conjugated nanorods (anti-EGFR or anti-CD3) complex can efficiently target the epitope of antigen on surface of different cells *in vitro*, and trigger the relative consequent reaction such as lymphocyte T inactivation. Moreover, the double antibodies conjugated nanorods (anti-EGFR × anti-CD3) complex can efficiently redirect the T cells to HeLa cells *in vitro*, which act as BsAbs. It clearly shows the potential of such Z-domain decorated nanorods in immunotherapy. Moreover, the Sup35-Z amyloid fibrils and the engineered nanorods described here appears to be a modular that can be readily used to conjugate different multiple specific antibodies for particular immunotherapy. The present work illustrate a straightforward strategy to obtain an multivalent functionalized nanomaterials and the proof-of-concept can be applied to exploit other nanomaterials for dual or multi-targeting.

Materials and Methods

Reagents and Materials. Reagents were purchased from Sigma-Aldrich (UK), unless otherwise stated. Antibodies were purchased from Thermo Fisher Scientific (UK). Carbon grid (400 square mesh copper) were purchased from Micro to Nano (Netherlands) and the uranyl acetate solution were provided by the microscopy service (Universitat Autònoma de Barcelona). Sup35-SAC 21-residues peptides were purchased from CASLO ApS (Scion Denmark Technical University).

Expression and Purification of Proteins. The cDNAs of Sup35-Z, consist of Sup35 soft amyloid core, 5 residues long linker and Z-domain of protein A, cloned in the plasmid pET28(b) with a His6 tag were acquired from GenScript (USA). The construct pET28(b)/Z-domain were acquired by using mutagenesis based on plasmid pET28(b)/Sup35-Z. *E.coli* BL21 (DE3) competent cells were transformed with the correspondent plasmids. Then, transformed cells were grown in 10 mL LB medium containing 50 μg/mL kanamycin, overnight at 37 °C, and transferred into 1 L fresh LB media containing 50 μg/mL kanamycin. After reaching an OD_{600} of 0.6, the culture was induced with 0.4 mM IPTG and grown at 20 °C for 16 h. Cells were collected by centrifugation at 5000 rpm for 15 min at 4 °C. The collected pellet was resuspended into 20 mL PBS pH 7.4 containing 20 mM imidazole, 1 mg/mL lysozyme and 1 mM PMSF. The solution was incubated on ice, followed by sonication for 20 min. The supernatant was collected by centrifugation at 15000 rpm for 30 min at 4 °C and, purified in an His-tag column, according to the manufacturer's protocol, followed by a gel filtration onto a HiLoad™ Superdex™ 75 prepgrade column (GE Healthcare,USA). The purified proteins were frozen with liquid nitrogen and stored at -80 °C. The purity of the sample was confirmed by SDS-PAGE. The concentration of the protein Z-domain and Sup35-Z was determined by UV absorption using a ε value of 1490 $L \cdot mol^{-1} \cdot cm^{-1}$ and 5960 $L \cdot mol^{-1} \cdot cm^{-1}$, respectively.

Conformational Characterization and Thermal Stability. Proteins were prepared at a final concentration of 10 μM in PBS pH 7.4 buffer, then samples were filtered through

a 0.22 μm Millipore filter and immediately analysed. Far-UV CD spectra were recorded from 260 nm to 200 nm at 1 nm bandwidth, response time of 1 second, and a scan speed of 100 nm/min in a Jasco-815 spectropolarimeter (Jasco Corporation, Japan), thermostated at 25 °C. Ten accumulations were averaged for each spectrum. For thermal stability, ellipticity was recorded at 222 nm each 0.5 °C with a heating rate 0.5 °C/min from 25 °C to 90 °C, using a Jasco-815 spectropolarimeter.

Purification of Antibody from Bovine Serum. Soluble proteins were prepared at a final concentration of 10 μM and a pull-down assay was performed. In particular, 50 μL protein solution were trapped into a His-tag column equilibrated with nickel ion and then incubated with bovine serum at room temperature for 30 min. The His-tag column was washed three times with PBS buffer. Bound IgG was eluted using EDTA (0.1 M) and the purity of IgG was analysed by SDS-PAGE. For the Sup35-Z fibrils, 200 μL incubated protein were precipitated and washed twice with PBS buffer. The fibrils were resuspended in bovine serum and incubated at room temperature for 30 min. The fibrils were then sedimented through centrifugation at 13200 rpm for 20 min and washed rigorously three times with PBS buffer. Bound IgG was eluted with 0.1 M glycine-HCl pH 3.0 buffer and the purity of IgG was analysed by SDS-PAGE.

Aggregation Assay. Sup35-Z protein and Sup35-SAC peptides were prepared at 200 μM in PBS pH 7.4, and filtered through a 0.22 μm filter. The samples were incubated at 37 °C with agitation at 600 rpm for 5 days. Z-domain were incubated at the same concentrations and conditions as controls.

Amyloid Dyes Binding Assay. Thioflavin T (Th-T) and Congo red (CR) were used to determine the formation of amyloid fibrils. For the Th-T binding assay, incubated proteins were diluted to a final concentration of 20 μM in PBS pH 7.4, in the presence of 25 μM Th-T. Emission fluorescence was recorded in the 460-600 nm range, using an excitation wavelength of 445 nm and emission bandwidth of 5 nm on a Jasco FP-8200 Spectrofluorometer (Jasco Corporation, Japan). For the CR binding assay, incubated proteins were prepared at final concentration of 20 μM and, CR was mixed

to a final concentration of 20 μM. Optical absorption spectra were recorded in the range from 375 to 700 nm in a Specord 200 Plus spectrophotometer (Analytik Jena, Germany). Spectra of protein alone and buffer were acquired to subtract protein scattering.

Transmission Electron Microscopy (TEM). For TEM samples preparation, 10 μL of the incubated proteins or incubated proteins sonicated for 5 min were deposited on a carbon-coated copper grid for 10 min and the excess liquid was removed with filter paper, followed by a negative stain with 10 μL of 2 %(w/v) uranyl acetate for 1min. Grids were exhaustively scanned using a JEM 1400 transmission electron microscope (JEOL ltd, Japan) operating at 80 kV, and images were acquired with a CCD GATAN ES1000W Erlangshen camera (Gatan Inc., USA). The width of fibrils was analysed by Image J (National Health Institute), averaging the measures of 10 individual fibrils for each fusion protein.

Fourier Transform Infrared Spectroscopy (FTIR). 30 μL of the prepared Sup35-Z fibrils at 200 μM were centrifuged at 12000g for 30 min and resuspended in 10 μL of water. Samples were placed on the ATR crystal and dried out under N_2 flow. The experiments were carried out in a Bruker Tensor 27 FTIR (Bruker Optics, USA) supplied with a Specac Golden Gate MKII ATR accessory. Each spectrum consists of 32 acquisitions measured at a resolution of 1 cm^{-1} using the three-term Blackman-Harris Window apodization function. Data were acquired and normalized, using the OPUS MIR Tensor 27 software (Bruker Optics, USA). IR spectra were fitted employing a nonlinear peak-fitting equation using Origin 8.5 (OriginLab Corporation). The area for each Gaussian curve was calculated in the amide I region (1700 to 1600 cm^{-1}) using a second derivative deconvolution method.

Antibody Binding Activity and Binding Capacity of Fibrils. Sup35-Z fibrils were washed twice and prepared at different concentration (0.1-0.4 μM) in PBS pH 7.4. 2 μg of secondary IgG labelled Alexa fluor 488 was incubated with fibrils at room temperature for 30 min. For two IgG binding assay, 1 μg of secondary IgG labelled with Alexa 488 and Alexa 555 were incubated with fibrils at room temperature for 30

min, respectively. The fibrils were then precipitated by centrifugation at 13200 rpm for 20 min and resuspended in PBS with three times washing steps. The fluorescence spectra of original IgG, supernatant and resuspended fibrils were recorded in the range of 510 nm to 600 nm using an excitation wavelength of 488 nm and emission bandwidth of 5 nm on a Jasco FP-8200 Spectrofluorometer (Jasco Corporation, Japan). The reduced fluorescence was calculated and fitted to linear equation using Origin 8.5 (OriginLab Corporation). 10 μL of the resuspended fibrils were dropped onto a clean glass slide (Deltalab, 26×76 mm) and covered by a cover slide (Deltalab, 22×22mm). Fluorescence imaging of nanofibers was carried out on an Eclipse 90i epifluorescence optical microscopy equipped with a Nikon DXM1200F (Nikon,Japan) camera and ACT-1 software. Images were acquired with an excitation filter of 465-495 nm and detecting fluorescence emission in a range of 515-555 nm. Sup35 peptides fibrils were prepared at 0.4 μM and treated with same conditions as a control.

Functionality of Bound Antibody Displayed in Sup35-Z Fibrils. 20 μL incubated Sup35-Z protein was sedimented and washed twice with PBS buffer pH 7.4. 1 μg of mouse anti-GFP IgG was incubated with fibrils at room temperature for 30 min. The fibrils were then washed three times and incubated with 10 μg of GFP and 2 μg of goat anti-mouse IgG Alexa 555, respectively. The fibrils were washed three times and resuspended in PBS buffer. 10 μL of the resuspended fibrils were dropped onto a clean glass slide (Deltalab, 26×76 mm) and covered by a cover slide (Deltalab, 22×22mm). The fluorescence imaging of captured GFP in the fibrils were analysed and observed on an Eclipse 90i epifluorescence optical microscopy as previous operation. The Alexa 555 fluorescence of secondary IgG was analysed on a Jasco FP-8200 Spectrofluorometer (Jasco Corporation, Japan) as described above. The GFP alone and secondary IgG alone were incubated with fibrils and analysed with same conditions as negative controls.

Cells and Cell Culture. Human HeLa cell line and Jurkat, clone E6-1 cell line (T lymphocyte) were obtained from American Type Culture Collection (ATCC). HeLa cells were maintained in Minimum Essential Medium Alpha media (MEM-α),

supplemented with 10% fetal bovine serum (FBS). Jurkat cell were maintained in Rosewell Park Memorial Institute media (RPMI 1640), supplemented with 10% fetal bovine serum. Both cells were incubated at 37 ℃ with 5% CO_2.

Cytotoxicity of Nanorods. HeLa cells were cultured on a 96-well plate at concentration of 3×10^3/well for 24 h. The nanorods were prepared in the range 1-25 μM and incubated with HeLa cells. Each sample was triplicate. The plate were incubated at 37℃ with 5% CO_2 for 48h. The PBS alone instead of fibrils were used as control and the medium without cells were used as blank control. 10 μL of PrestoBlue reagent (ThermoFisher Scientific) were added to each well and incubated for another 1 h. The fluorescence were analysed on a Victor III Multilabel Plate Reader (Perkin Elmer,USA), equipped with 530/10 nm CW-lamp filter and 590/20 nm emission filter. The viability of cells were calculated as follow equation:

$$\text{Viability (\%)} = (I_{test}-I_{blank}) / (I_{control}-I_{blank}) \times 100\%$$

Where the I_{test}, I_{blank} and $I_{control}$ are the fluorescence intensity of test, blank and control group, respectively. The significance test of difference between the test group and the control were analysed by one-way Analysis of Variance (ANOVA) using Origin 8.5 program (OriginLab Corporation).

Preparation of Antibody Conjugated Nanorods. The incubated protein Sup35-Z was precipitated by using centrifugation at 13000 rpm for 30 min. The precipitate was sonicated 1 min and resuspended in PBS buffer pH 7.4 containing 1 μg antibody and then incubated for 30 min. For the two antibody conjugated nanorods, each antibody was used 1 μg. The concentration of nanorods was determined by the reduction of absorbance at 280 nm in the supernatant fraction. Three labelling antibodies (Thermo Fisher Scientific, USA): anti-EGFR antibody labelled with Alexa fluor 555, anti-CD3 antibody labelled with Alexa fluor 488 and goat anti-rabbit antibody labelled with Alexa 555. Then the antibody conjugated nanorods were washed three times with PBS buffer to remove the unbound antibodies and resuspended in PBS buffer. The antibody loaded nanorods (NRs-anti-EGFR, NRs-anti-CD3, anti-EGFR-NRs-anti-CD3, anti-

EGFR-NRs-anti-rabbit) were used for consequent experiments immediately.

Antibodies Conjugated Sup35-Z Nanorods Target Human Cells. HeLa cells were cultured on a 8-well Millicell® EZ slide (Millipore, Germany) to a final confluence of 70-80%. Then the medium was replaced with fresh medium containing 10 μM of NRs-anti-EGFR. Then the slide was incubated at 37 ℃, 5% CO_2 for 20 min. The anti-CD3 antibody loaded nanorods (NRs-anti-CD3) was used as control. Then the medium was removed. The adherent cells on slide were rinsed three times with fresh medium. For the lymphocyte, the cells were harvested resuspended in fresh medium containing 10 μM of NRs-anti-CD3. Then the cells was incubated at 37 ℃, 5% CO_2 for 20 min. Then the cells were precipitated and washed three times with fresh medium. 150 μL of suspension was transferred to the wells of the slide and incubated at 37 ℃, 5% CO_2 for 20 min. The medium was removed slightly and the cells were slightly rinsed three times. Cells were fixed with 4% PFA at room temperature for 20 min followed by a washing step with PBS buffer. The 4 tabs were break and 10 μL mounting medium containing DAPI was dropped onto the each well of slide. A coverslip was put on the slide. The slide was observed on a Leica TCS SP5 confocal microscope (Leica Biosystems, Germany). Images were acquired by using 405 nm, 488 and 561nm excitation laser for DAPI, Alexa fluor 488 and 555nm, respectively.

Proliferation Response of T Lymphocyte in the Presence of NRs-anti-CD3. 50 μL of anti-CD3 antibody alone and NRs-anti-CD3 resuspension were dispensed to each well of 96-well plate. Each sample was triplicate. 50 μL of sterile PBS and fibrils resuspension were used as controls. The plate was sealed with Parafilm™ and incubated at 37 ℃ for 2 h. Then the solution was removed and each well was rinsed three times with 200 μL PBS to remove all unbound IgG. Lymphocyte T were prepared at a final concentration of 10^6 /mL. 200 μL of cell suspension was added to each well and incubated at 37 ℃, 5% CO_2 for 2 days. The medium without cells were used as blank control. The proliferation of cell was analysed using PrestoBlue assay as described above. Statistical calculation was performed by using Origin 8.5 program (OriginLab Corporation) as described above.

Two mAbs conjugated Sup35-Z Nanorods Redirect Lymphocyte to HeLa Cells. HeLa cells were cultured on a 8-well Millicell® EZ slide (Millipore, Germany) to a final confluence of 70-80%. Then the medium was replaced with fresh medium containing 10 μM of anti-EGFR-NRs-anti-CD3. Then the slide was incubated at 37 ℃, 5% CO_2 for 20 min. The anti-EGFR-NRs-anti-rabbit was used as control. The medium was removed and each well was rinsed three times with PBS buffer. 150 μL of lymphocyte cells suspension were added and incubated for another 20 min. The medium was removed and each well was rinse three times with PBS buffer. Cells were stained with CellMask Deep Red for 10 min. Then cells were fixed with 4% PFA at room temperature for 20 min followed by a washing step with PBS buffer. The 4 tabs were break and 10 μL mounting medium containing DAPI was dropped onto the each well of slide. A coverslip was put on the slide. The slide was observed on a Leica TCS SP5 confocal microscope (Leica Biosystems, Germany). Images were acquired by using 405 nm, 488 or 561nm and 633 nm excitation laser for DAPI, Alexa fluor 488 or 555nm and CellMask Deep Red, respectively.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This work was funded by the Spanish Ministry of Economy and Competitiveness BIO2016-78310-R to S.V and by ICREA, ICREA-Academia 2015 to S.V. Weiqiang Wang acknowledges financial support from the China Scholarship Council (CSC): NO. 201606500007.

References

1. Chen, D.; Li, B.; Cai, S.; Wang, P.; Peng, S.; Sheng, Y.; He, Y.; Gu, Y.; Chen, H., Dual targeting luminescent gold nanoclusters for tumor imaging and deep tissue therapy. *Biomaterials* **2016,** *100*, 1-16.

2. Fang, J.; Nakamura, H.; Maeda, H., The EPR effect: unique features of tumor blood vessels for drug delivery, factors involved, and limitations and augmentation of the effect. *Advanced drug delivery reviews* **2011,** *63* (3), 136-151.

3. Huynh, E.; Zheng, G., Cancer nanomedicine: addressing the dark side of the enhanced permeability and retention effect. *Nanomedicine* **2015,** *10* (13), 1993-1995.

4. Sapsford, K. E.; Algar, W. R.; Berti, L.; Gemmill, K. B.; Casey, B. J.; Oh, E.; Stewart, M. H.; Medintz, I. L., Functionalizing nanoparticles with biological molecules: developing chemistries that facilitate nanotechnology. *Chemical reviews* **2013,** *113* (3), 1904-2074.

5. Li, Z.; Huang, P.; Zhang, X.; Lin, J.; Yang, S.; Liu, B.; Gao, F.; Xi, P.; Ren, Q.; Cui, D., RGD-conjugated dendrimer-modified gold nanorods for in vivo tumor targeting and photothermal therapy. *Molecular pharmaceutics* **2010,** *7* (1), 94-104.

6. Song, H.; He, R.; Wang, K.; Ruan, J.; Bao, C.; Li, N.; Ji, J.; Cui, D., Anti-HIF-1α antibody-conjugated pluronic triblock copolymers encapsulated with Paclitaxel for tumor targeting therapy. *Biomaterials* **2010,** *31* (8), 2302-2312.

7. Vicent, M. J.; Duncan, R., Polymer conjugates: nanosized medicines for treating cancer. *Trends in biotechnology* **2006,** *24* (1), 39-47.

8. van der Meel, R.; Vehmeijer, L. J.; Kok, R. J.; Storm, G.; van Gaal, E. V., Ligand-targeted particulate nanomedicines undergoing clinical evaluation: current status. *Advanced drug delivery reviews* **2013,** *65* (10), 1284-1298.

9. Cheng, K.; Shen, D.; Hensley, M. T.; Middleton, R.; Sun, B.; Liu, W.; De Couto, G.; Marbán, E., Magnetic antibody-linked nanomatchmakers for therapeutic cell targeting. *Nature Communications* **2014,** *5* (1), 1-9.

10. Gao, H., Perspectives on dual targeting delivery systems for brain tumors. *Journal of Neuroimmune Pharmacology* **2017,** *12* (1), 6-16.

11. Krishnamurthy, A.; Jimeno, A., Bispecific antibodies for cancer therapy: a review. *Pharmacology & therapeutics* **2018,** *185*, 122-134.

12. Kontermann, R. E.; Brinkmann, U., Bispecific antibodies. *Drug discovery today* **2015,** *20* (7), 838-847.

13. Nisonoff, A.; Mandy, W., Quantitative estimation of the hybridization of rabbit antibodies. *Nature* **1962,** *194* (4826), 355-359.

14. Kufer, P.; Lutterbüse, R.; Baeuerle, P. A., A revival of bispecific antibodies. *Trends in biotechnology* **2004,** *22* (5), 238-244.

15. Liu, H.; Saxena, A.; Sidhu, S. S.; Wu, D., Fc engineering for developing therapeutic bispecific antibodies and novel scaffolds. *Frontiers in immunology* **2017,** *8*, 38.

16. Knödler, M.; Körfer, J.; Kunzmann, V.; Trojan, J.; Daum, S.; Schenk, M.; Kullmann, F.; Schroll, S.; Behringer, D.; Stahl, M., Randomised phase II trial to investigate catumaxomab (anti-EpCAM× anti-CD3) for treatment of peritoneal carcinomatosis in patients with gastric cancer. *British journal of cancer* **2018,** *119* (3), 296-302.

17. von Stackelberg, A.; Locatelli, F.; Zugmaier, G.; Handgretinger, R.; Trippett, T. M.; Rizzari, C.; Bader, P.; O'brien, M. M.; Brethon, B.; Bhojwani, D., Phase I/phase II study of blinatumomab in pediatric patients with relapsed/refractory acute lymphoblastic leukemia. *Journal of Clinical Oncology* **2016,** *34* (36), 4381-4389.

18. Lv, Y.; Xu, C.; Zhao, X.; Lin, C.; Yang, X.; Xin, X.; Zhang, L.; Qin, C.; Han, X.; Yang, L., Nanoplatform assembled from a CD44-targeted prodrug and smart liposomes for dual targeting of tumor microenvironment and cancer cells. *Acs Nano* **2018,** *12* (2), 1519-1536.

19. Kosmides, A. K.; Sidhom, J.-W.; Fraser, A.; Bessell, C. A.; Schneck, J. P., Dual targeting nanoparticle stimulates the immune system to inhibit tumor growth. *ACS nano* **2017,** *11* (6), 5417-5429.

20. Giudice, M. C. L.; Meder, F.; Polo, E.; Thomas, S. S.; Alnahdi, K.; Lara, S.; Dawson, K. A., Constructing bifunctional nanoparticles for dual targeting: improved grafting and surface recognition assessment of multiple ligand nanoparticles. *Nanoscale* **2016,** *8* (38), 16969-16975.

21. Wei, G.; Su, Z.; Reynolds, N. P.; Arosio, P.; Hamley, I. W.; Gazit, E.; Mezzenga, R., Self-assembling peptide and protein amyloids: from structure to tailored function in nanotechnology. *Chemical Society Reviews* **2017,** *46* (15), 4661-4708.

22. Qi, G. B.; Gao, Y. J.; Wang, L.; Wang, H., Self-Assembled Peptide-Based Nanomaterials for Biomedical Imaging and Therapy. *Advanced Materials* **2018,** *30* (22), 1703444.

23. Mason, T. O.; Shimanovich, U., Fibrous Protein Self-Assembly in Biomimetic Materials. *Advanced Materials* **2018,** *30* (41), 1706462.

24. Hennig, R.; Pollinger, K.; Veser, A.; Breunig, M.; Goepferich, A., Nanoparticle multivalency counterbalances the ligand affinity loss upon PEGylation. *Journal of Controlled Release* **2014,** *194*, 20-27.

25. Wang, W.; Navarro, S.; Azizyan, R. A.; Baño-Polo, M.; Esperante, S. A.; Kajava, A. V.; Ventura, S., Prion soft amyloid core driven self-assembly of globular proteins into bioactive nanofibrils. *Nanoscale* **2019,** *11* (26), 12680-12694.

26. Tashiro, M.; Tejero, R.; Zimmerman, D. E.; Celda, B.; Nilsson, B.; Montelione, G. T., High-resolution solution NMR structure of the Z-domain of staphylococcal protein A. *Journal of molecular biology* **1997,** *272* (4), 573-590.

27. Forsgren, A.; Sjöquist, J., "Protein A" from S. aureus: I. Pseudo-immune reaction with human γ-globulin. *The Journal of Immunology* **1966,** *97* (6), 822-827.

28. Schmuck, B.; Sandgren, M.; Härd, T., A fine-tuned composition of protein nanofibrils yields

an upgraded functionality of displayed antibody binding domains. *Biotechnology Journal* **2017,** *12* (6), 1600672.

29. Ohta, S.; Glancy, D.; Chan, W. C., DNA-controlled dynamic colloidal nanoparticle systems for mediating cellular interaction. *Science* **2016,** *351* (6275), 841-845.

30. Azizyan, R. A.; Garro, A.; Radkova, Z.; Anikeenko, A.; Bakulina, A.; Dumas, C.; Kajava, A. V., Establishment of constraints on amyloid formation imposed by steric exclusion of globular domains. *Journal of molecular biology* **2018,** *430* (20), 3835-3846.

31. Levine Iii, H.; Scholten, J. D., [29] Screening for pharmacologic inhibitors of amyloid fibril formation. In *Methods in enzymology*, Elsevier: 1999; Vol. 309, pp 467-476.

32. Klunk, W. E.; Pettegrew, J.; Abraham, D. J., Quantitative evaluation of congo red binding to amyloid-like proteins with a beta-pleated sheet conformation. *Journal of Histochemistry & Cytochemistry* **1989,** *37* (8), 1273-1281.

33. Sant'Anna, R.; Fernández, M. R.; Batlle, C.; Navarro, S.; De Groot, N. S.; Serpell, L.; Ventura, S., Characterization of amyloid cores in prion domains. *Scientific reports* **2016,** *6* (1), 1-10.

34. Díaz-Caballero, M.; Navarro, S.; Ventura, S., Soluble assemblies in the fibrillation pathway of prion-inspired artificial functional amyloids are highly cytotoxic. *Biomacromolecules* **2020**.

35. Jenkins, M. K.; Chen, C.; Jung, G.; Mueller, D. L.; Schwartz, R. H., Inhibition of antigen-specific proliferation of type 1 murine T cell clones after stimulation with immobilized anti-CD3 monoclonal antibody. *The Journal of Immunology* **1990,** *144* (1), 16-22.

Supporting information for:

Dual antibody-conjugated amyloid nanorods to promote selective interactions between different cell types

*Weiqiang Wang[1], and Salvador Ventura[1]**

[1]Institut de Biotecnologia i de Biomedicina and Departament de Bioquímica i Biologia Molecular; Universitat Autònoma de Barcelona; 08193 Bellaterra (Barcelona), Spain.

E-mail: salvador.ventura@uab.es

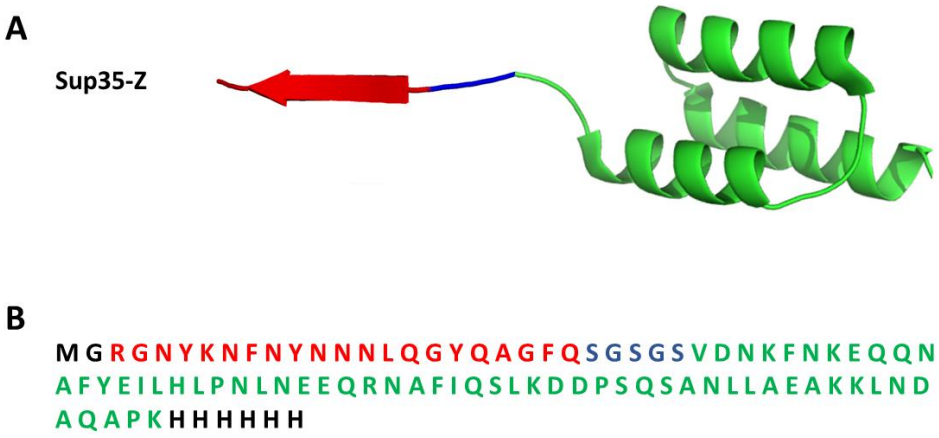

Figure S1. Schematic representation, sequence and expression of the Sup35-Z fusion protein. (A) Sup35-Z with Sup35 soft amyloid core (SAC) (residues 98-118) fused to Z domain (PDB: 1Q2N), an engineered analog of the B domain of *Staphylococcus aureus* protein A shown in cartoon representation. (B) Sequence of the Sup35-Z. The SAC, spacer linkers, globular structure and His6 tag are shown in red, blue, green and black, respectively.

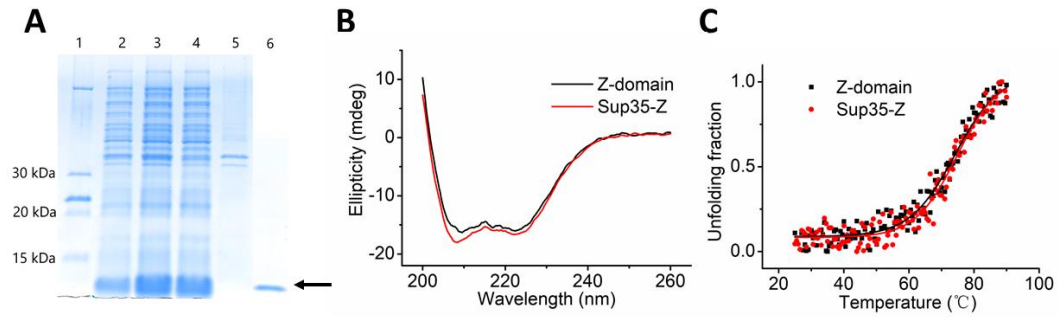

Figure S2. Solubility, conformation, and stability of Sup35-Z protein. (A) Analysis on SDS-PAGE of the expression of Sup35-Z fusion protein. *Lane 1*, corresponds to molecular weight marker, *lane 2*, non-induced culture, *lane 3*, total extract induced, *lane 4*, soluble fraction (supernatant) and, *lane 5*, insoluble fraction (pellet). *lane 6*, shows purified Sup35-Z by gel filtration (extracted from another gel). A black arrow indicates the band corresponding to Sup35-Z. (B) Far-UV CD spectra. (C) Thermal stability was analyzed by far-UV CD signal at 222 nm.

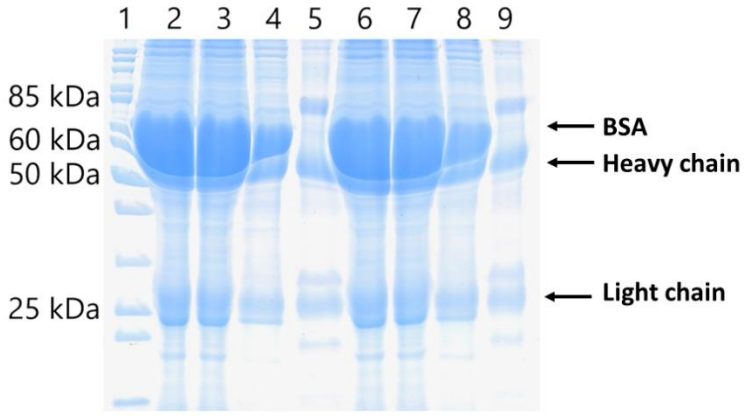

Figure S3. SDS-PAGE analysis on antibody binding affinity to soluble Z domain and Sup35-Z fusion. *Lane 1-5*, corresponds to Sup35-Z, *Lane 1*, corresponds to molecule weight marker, *lane 2*, bovine serum, *lane 3*, Flow-through of bovine serum after incubation with Sup35-Z domain loaded in His-tag column, *lane 4*, PBS buffer wash, *lane 5*, eluate with 0.1 M EDTA; *lane 6-9*, corresponds to the same fraction of Z domain as Sup35-Z fusion.

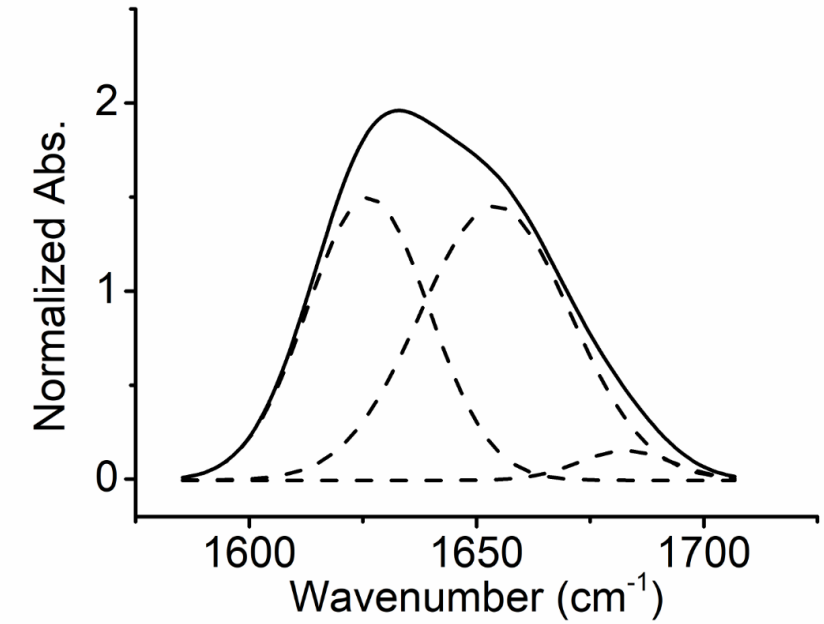

Figure S4. Conformational properties of Sup35-Z fusion protein fibrils. Sup35-Z protein solutions were incubated for 5 days. The absorbance spectra of Sup35-FF fibrils in the amide I region (solid line) and the components bands (dashed lines) are shown.

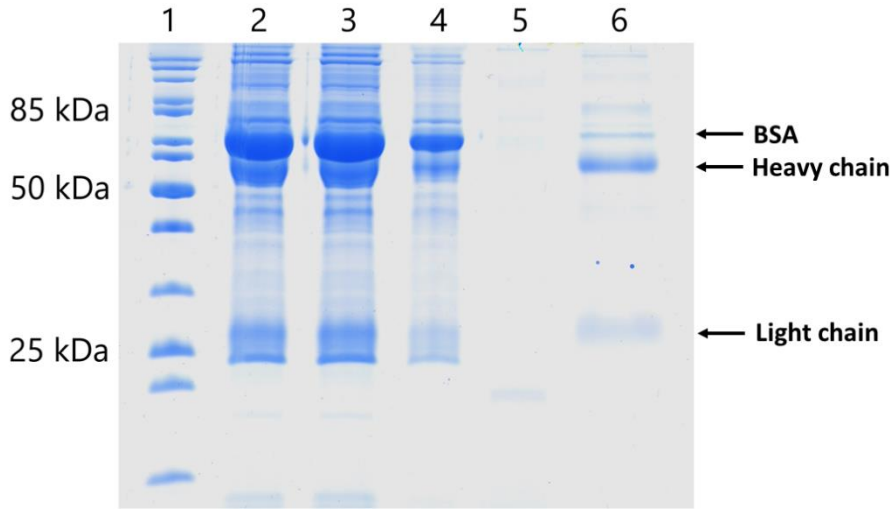

Figure S5. SDS-PAGE analysis of antibody binding affinity to Sup35-Z fusion protein fibrils. *Lane 1* corresponds to molecule weight marker, *lane 2*, bovine serum, *lane 3*, Supernatant of bovine serum after incubation with Sup35-Z fibrils, *lane 4*, PBS buffer wash, *lane 5*, insoluble fraction after elution with 0.1 M glycine buffer pH 3.0. *lane 6*, eluate with 0.1 M glycine buffer pH 3.0.

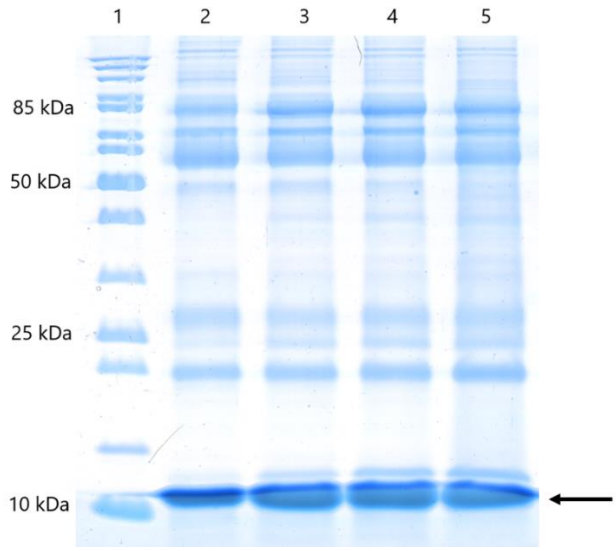

Figure S6. SDS-PAGE analysis of stability of Sup35-Z fusion protein fibrils in Bovine serum. *Lane 1* corresponds to molecule weight marker, *lane 2-5*, Sup35-Z fibrils incubated with bovine serum for 30 min, 0.5 day, 1 day and 3 day, respectively.

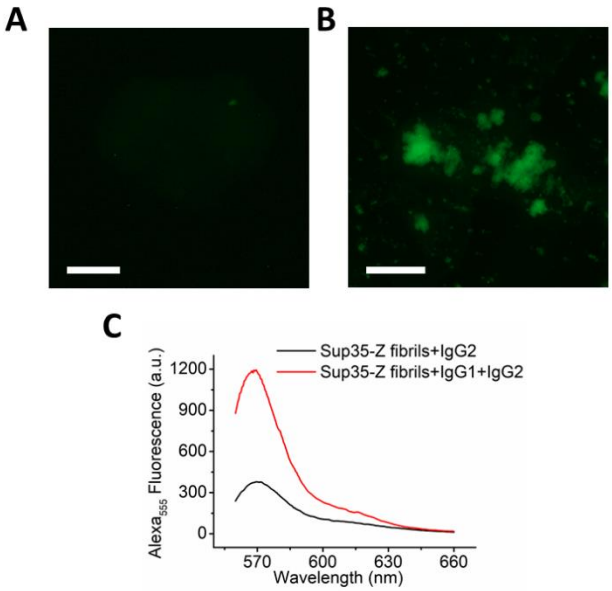

Figure S7. Functionality of the conjugated antibody on Sup35-Z fibrils. Fluorescence microscopy image of (A) Sup35-Z fibrils and (B) anti-GFP antibody conjugated Sup35-Z fibrils in the presence of GFP, scale bar represent 50 μm. (C) Fluorescence spectra of Sup35-Z fibrils in the presence of secondary antibody (black line) and primary and secondary antibody(red line).

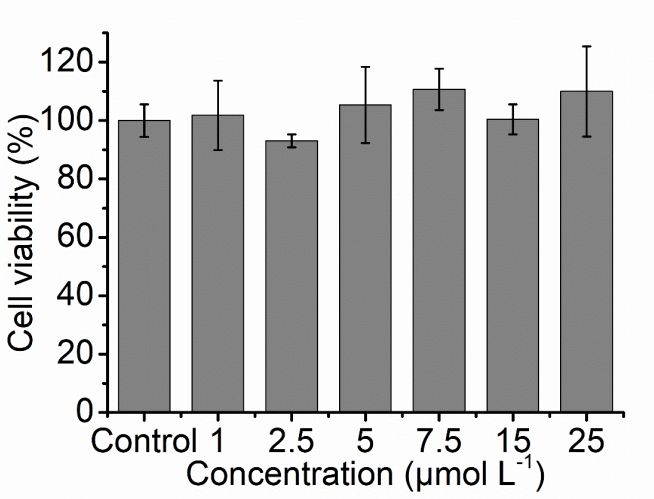

Figure S8. Cytotoxicity of the Sup35-Z nanorods. Results are expressed as means ± SD, n=3, and analyzed using a one-way ANOVA test. The statistical difference between the control group and the test group was established at $P < 0.05$.

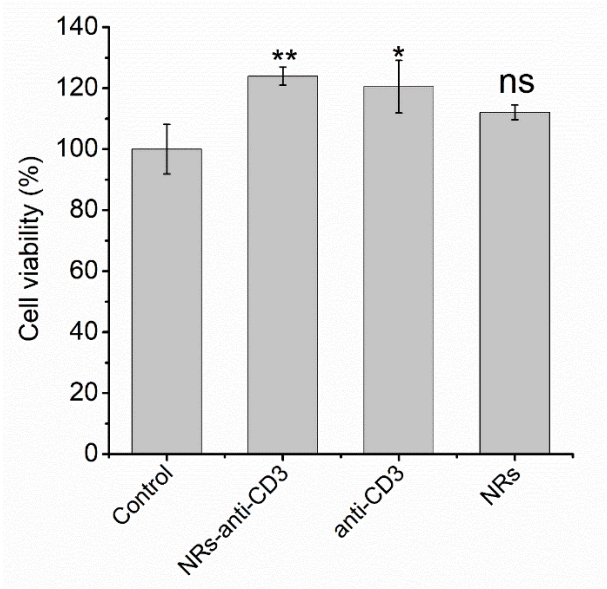

Figure S9. Proliferation response of the T lymphocyte. T lymphocyte proliferation was measured in the presence of coated anti-CD3 antibody conjugated nanorods (NRs-anti-CD3), anti-CD3 antibody, and nanorods (NRs) alone. Results are expressed as means ± SD, n=3, and analyzed using a one-way ANOVA test. The statistical differences between the control group and the test group were established at * $P < 0.05$ and ** $P < 0.01$.

Table S1. Assignment and area of the secondary structure components of Sup35-Z fibrils in the amide I region of the FTIR spectra

Assignments (%)	Sup35-Z fibrils
Inter β-sheet	44.0 (1626 cm^{-1})
α-helix/turns	52.6 (1654 cm^{-1})
β-sheet	3.4 (1682 cm^{-1})

General conclusions

Chapter I Prion soft amyloid core driven self-assembly of globular proteins into bioactive nanofibrils

1) Sup35-SAC does not affect the conformation, stability, folding properties of the adjacent globular proteins.

2) Sup35-SAC induce controlled self-assembly of fusion proteins into a functional amyloid fibrillar structure, in which the globular proteins remain in their native structure and hang from the core of amyloid fibrils.

3) The molecule modeling based on fusion Sup35-GFP suggests that a Sup35-SAC amyloid core surrounded by globular domains can be formed by stacking of either linear β-strands or β-arches without significant steric tension.

4) Sup35-SAC appears as a module that can be readily used to immobilize bioactive proteins of interest with different sizes and structures.

5) The modular genetic fusion approach described here can be applied to decorate fibrils with different functionalities, including active enzymes.

Chapter II Amyloidogenicity as a driving force for the formation of functional nanoparticles

1) By using molecular modeling, we propose that ARs linked with a globular protein can self-assemble into amyloid oligomers when the linker between the AR and the globular domain is shorter than the one allowing the formation of the infinite fibrils.
2) We can predict the assembly of such large systems containing molecules with amyloid-forming regions linked to globular structures by using CG-TMD and Rigid Body simulations.
3) This proof-of-concept study opens up further opportunities to fabricate nanostructures of defined size carrying multiple functional domains.

Chapter III Multifunctional amyloid oligomeric nanoparticles for specific cell targeting and drug delivery

1) We first demonstrated that by modulating inter-domain linker length, one could attain a tight control of the mesoscopic properties of the resulting amyloid-like nanostructures.

2) The approach allowed us to generate oligomeric amyloid-like spherical nanoparticles, which are homogenous in size, stable, and biocompatible.

3) We provide a proof-of-concept of the utility of these de novo designed nanostructures decorated with an antibody of interest, which direct the multivalent nanoparticles to the specific cell types expressing the selected antigen at its surface, allowing to discriminate between diseased and functional cells.

4) These hybrid nanoparticles appear as a very appealing strategy for targeted delivery of the drug in the proteinase enriched microenvironment of tumors, followed by the killing of tumor cells.

5) We propose a new and safe nanotechnologic modular scaffold with the potential of facilitating the specific delivery of agents to specific sites in the body, overpassing the major barrier for bioimaging and tissue-targeted therapies.

Chapter IV Dual antibody-conjugated amyloid nanorods to promote selective interactions between different cell types

1) We built up antibodies-capturing amyloid fibrils using a hybrid protein containing a SAC and the globular Z domain.

2) These amyloid fibrils can be further engineered and form homogenous and biocompatible amyloid nanorods.

3) Single antibody functionalized nanorods target specifically the cells expressing the intended receptor, while bivalent nanorods can efficiently bring two targeted cells together.

4) The amyloid fibrils or nanorods described in this work constitute nanometric scaffolds that can be readily used for cell multi-targeting with potential therapeutic applications.

References

1. Creighton, T. E., *Proteins: structures and molecular properties*. Macmillan: 1993.

2. Dobson, C. M., Protein folding and misfolding. *Nature* **2003,** *426* (6968), 884-890.

3. DeSaix, P.; Betts, J. G.; Johnson, E.; Johnson, J. E.; Korol, O.; Kruse, D. H.; Poe, B.; Wise, J. A.; Young, K. A., Anatomy & Physiology: OpenStax. **2018**.

4. Crick, F., Central dogma of molecular biology. *Nature* **1970,** *227* (5258), 561.

5. Cobb, M., 60 years ago, Francis Crick changed the logic of biology. *PLoS biology* **2017,** *15* (9), e2003243.

6. Nirenberg, M. W.; Matthaei, J. H., The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides. *Proceedings of the National Academy of Sciences* **1961,** *47* (10), 1588-1602.

7. Nirenberg, M.; Leder, P.; Bernfield, M.; Brimacombe, R.; Trupin, J.; Rottman, F.; O'neal, C., RNA codewords and protein synthesis, VII. On the general nature of the RNA code. *Proceedings of the National Academy of Sciences of the United States of America* **1965,** *53* (5), 1161.

8. Crick, F., Chapter 8: The genetic code. *What mad pursuit: a personal view of scientific discovery. New York: Basic Books* **1988**, 89-101.

9. Carrell, R. W.; Lomas, D. A., Conformational disease. *The Lancet* **1997,** *350* (9071), 134-138.

10. Leopold, P. E.; Montal, M.; Onuchic, J. N., Protein folding funnels: a kinetic approach to the sequence-structure relationship. *Proceedings of the National Academy of Sciences* **1992,** *89* (18), 8721-8725.

11. Gierasch, L. M.; King, J., *Protein folding: deciphering the second half of the genetic code*. American Association for the Advancement of Science: 1990.

12. Edsall, J. T., Hsien Wu and the first theory of protein denaturation (1931). In *Advances in protein chemistry*, Elsevier: 1995; Vol. 46, pp 1-5.

13. Wu, H., Studies on denaturation of proteins XIII. A theory of denaturation. In *Advances in protein chemistry*, Elsevier: 1995; Vol. 46, pp 6-26.

14. Anson, M.; Mirsky, A., On some general properties of proteins. *The Journal of general physiology* **1925,** *9* (2), 169.

15. Robertson, T. B., *The physical chemistry of the proteins*. Longmans, Green and Company: 1918.

16. Anson, M.; Mirsky, A., On haemochromogen and the relation of protein to the properties of the haemoglobin molecule. *The Journal of physiology* **1925,** *60* (1-2), 50.

17. Anson, M., Protein denaturation and the properties of protein groups. In *Advances in protein chemistry*, Elsevier: 1945; Vol. 2, pp 361-386.

18. Lumry, R.; Eyring, H., Conformation changes of proteins. *The Journal of physical chemistry*

1954, *58* (2), 110-120.

19. Anfinsen, C. B.; Haber, E.; Sela, M.; White Jr, F., The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proceedings of the National Academy of Sciences of the United States of America* **1961,** *47* (9), 1309.

20. Haber, E.; Anfinsen, C. B., Side-chain interactions governing the pairing of half-cystine residues in ribonuclease. *Journal of Biological Chemistry* **1962,** *237* (6), 1839-1844.

21. Epstein, C. J.; Goldberger, R. F.; Anfinsen, C. B. In *The genetic control of tertiary protein structure: studies with model systems*, Cold Spring Harbor symposia on quantitative biology, Cold Spring Harbor Laboratory Press: 1963; pp 439-449.

22. Dixon, G.; Wardlaw, A., Regeneration of insulin activity from the separated and inactive A and B chains. *Nature* **1960,** *188* (4752), 721-724.

23. Du, Y. C.; Zhang, Y. S.; Lu, Z.; Tsou, C., Resynthesis of insulin from its glycyl and phenylalanyl chains. *Scientia sinica* **1961,** *10*, 84.

24. Gutte, B.; Merrifield, R., The synthesis of ribonuclease A. *Journal of Biological Chemistry* **1971,** *246* (6), 1922-1941.

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

26. Matouschek, A.; Kellis, J. T.; Serrano, L.; Fersht, A. R., Mapping the transition state and pathway of protein folding by protein engineering. *Nature* **1989,** *340* (6229), 122-126.

27. Wells, J. A.; Ferrari, E.; Henner, D. J.; Estell, D. A.; Chen, E. Y., Cloning, sequencing, and secretion of Bacillus amyloliquefaciens subtillisin in Bacillus subtilis. *Nucleic acids research* **1983,** *11* (22), 7911-7925.

28. Silen, J.; McGrath, C.; Smith, K.; Agard, D., Molecular analysis of the gene encoding α-lytic protease: evidence for a preproenzyme. *Gene* **1988,** *69* (2), 237-244.

29. Baker, D.; Silen, J. L.; Agard, D. A., Protease pro region required for folding is a potent inhibitor of the mature enzyme. *Proteins: Structure, Function, and Bioinformatics* **1992,** *12* (4), 339-344.

30. Eder, J.; Rheinnecker, M.; Fersht, A. R., Folding of subtilisin BPN': characterization of a folding intermediate. *Biochemistry* **1993,** *32* (1), 18-26.

31. Franke, A. E.; Danley, D. E.; Kaczmarek, F. S.; Hawrylik, S. J.; Gerard, R. D.; Lee, S. E.; Geoghegan, K. F., Expression of human plasminogen activator inhibitor type-1 (PAI-1) in Escherichia coli as a soluble protein comprised of active and latent forms. Isolation and crystallization of latent PAI-1. *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology* **1990,** *1037* (1), 16-23.

32. Wimmer, E.; Harber, J.; Bibb, J.; Gromeier, M.; Lu, H.; Bernhardt, G., *Cellular receptors for animal viruses*. Cold Spring Harbor Laboratory Press Cold Spring Harbor: 1994; Vol. 28.

33. Baker, D.; Agard, D. A., Kinetics versus thermodynamics in protein folding. *Biochemistry* **1994,**

33 (24), 7505-7509.

34. Plaxco, K. W.; Dobson, C. M., Time-resolved biophysical methods in the study of protein folding. *Current opinion in structural biology* **1996,** *6* (5), 630-636.

35. Dobson, C. M.; Karplus, M., The fundamentals of protein folding: bringing together theory and experiment. *Current opinion in structural biology* **1999,** *9* (1), 92-101.

36. Dinner, A. R.; Šali, A.; Smith, L. J.; Dobson, C. M.; Karplus, M., Understanding protein folding via free-energy surfaces from theory and experiment. *Trends in biochemical sciences* **2000,** *25* (7), 331-339.

37. Levinthal, C., Are there pathways for protein folding? *Journal de chimie physique* **1968,** *65*, 44-45.

38. Baldwin, R. L., Matching speed and stability. *Nature* **1994,** *369* (6477), 183-184.

39. Dill, K. A., Dominant forces in protein folding. *Biochemistry* **1990,** *29* (31), 7133-7155.

40. Onuchic, J. N.; Wolynes, P. G., Theory of protein folding. *Current opinion in structural biology* **2004,** *14* (1), 70-75.

41. Radford, S. E.; Dobson, C. M., From computer simulations to human disease: emerging themes in protein folding. *Cell* **1999,** *97* (3), 291-298.

42. Kubelka, J.; Hofrichter, J.; Eaton, W. A., The protein folding 'speed limit'. *Current opinion in structural biology* **2004,** *14* (1), 76-88.

43. Ellis, R. J.; Minton, A. P., Protein aggregation in crowded environments. *Biological chemistry* **2006,** *387* (5), 485-497.

44. Hartl, F. U.; Hayer-Hartl, M., Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science* **2002,** *295* (5561), 1852-1858.

45. Dobson, C. M. In *Principles of protein folding, misfolding and aggregation*, Seminars in cell & developmental biology, Elsevier: 2004; pp 3-16.

46. Hartl, F. U.; Bracher, A.; Hayer-Hartl, M., Molecular chaperones in protein folding and proteostasis. *Nature* **2011,** *475* (7356), 324-332.

47. Capaldi, A. P.; Kleanthous, C.; Radford, S. E., Im7 folding mechanism: misfolding on a path to the native state. *Nature structural biology* **2002,** *9* (3), 209-216.

48. Clarke, G.; Collins, R. A.; Leavitt, B. R.; Andrews, D. F.; Hayden, M. R.; Lumsden, C. J.; McInnes, R. R., A one-hit model of cell death in inherited neuronal degenerations. *Nature* **2000,** *406* (6792), 195-199.

49. Eichner, T.; Kalverda, A. P.; Thompson, G. S.; Homans, S. W.; Radford, S. E., Conformational conversion during amyloid formation at atomic resolution. *Molecular cell* **2011,** *41* (2), 161-172.

50. Hammond, C.; Helenius, A., Quality control in the secretory pathway. *Current opinion in cell biology* **1995,** *7* (4), 523-529.

51. Kaufman, R. J.; Scheuner, D.; Schröder, M.; Shen, X.; Lee, K.; Liu, C. Y.; Arnold, S. M., The unfolded protein response in nutrient sensing and differentiation. *Nature Reviews Molecular Cell Biology* **2002,** *3* (6), 411-421.

52. Parsell, D. A.; Kowal, A. S.; Singer, M. A.; Lindquist, S., Protein disaggregation mediated by heat-shock protein Hspl04. *Nature* **1994,** *372* (6505), 475-478.

53. Gething, M.-J.; Sambrook, J., Protein folding in the cell. *Nature* **1992,** *355* (6355), 33-45.

54. Singleton, A.; Farrer, M.; Johnson, J.; Singleton, A.; Hague, S.; Kachergus, J.; Hulihan, M.; Peuralinna, T.; Dutra, A.; Nussbaum, R., α-Synuclein locus triplication causes Parkinson's disease. *Science* **2003,** *302* (5646), 841-841.

55. Singleton, A.; Myers, A.; Hardy, J., The law of mass action applied to neurodegenerative disease: a hypothesis concerning the etiology and pathogenesis of complex diseases. *Human molecular genetics* **2004,** *13* (suppl_1), R123-R126.

56. Conway, K. A.; Rochet, J.-C.; Bieganski, R. M.; Lansbury, P. T., Kinetic stabilization of the α-synuclein protofibril by a dopamine-α-synuclein adduct. *Science* **2001,** *294* (5545), 1346-1349.

57. Okochi, M.; Walter, J.; Koyama, A.; Nakajo, S.; Baba, M.; Iwatsubo, T.; Meijer, L.; Kahle, P. J.; Haass, C., Constitutive phosphorylation of the Parkinson's disease associated α-synuclein. *Journal of Biological Chemistry* **2000,** *275* (1), 390-397.

58. Steffan, J. S.; Agrawal, N.; Pallos, J.; Rockabrand, E.; Trotman, L. C.; Slepko, N.; Illes, K.; Lukacsovich, T.; Zhu, Y.-Z.; Cattaneo, E., SUMO modification of Huntingtin and Huntington's disease pathology. *Science* **2004,** *304* (5667), 100-104.

59. Ross, C. A.; Poirier, M. A., Protein aggregation and neurodegenerative disease. *Nature medicine* **2004,** *10* (7), S10-S17.

60. Kelly, J. W., The alternative conformations of amyloidogenic proteins and their multi-step assembly pathways. *Current opinion in structural biology* **1998,** *8* (1), 101-106.

61. Harper, J. D.; Lansbury Jr, P. T., Models of amyloid seeding in Alzheimer's disease and scrapie: mechanistic truths and physiological consequences of the time-dependent solubility of amyloid proteins. *Annual review of biochemistry* **1997,** *66* (1), 385-407.

62. Bolognesi, B.; Kumita, J. R.; Barros, T. P.; Esbjorner, E. K.; Luheshi, L. M.; Crowther, D. C.; Wilson, M. R.; Dobson, C. M.; Favrin, G.; Yerbury, J. J., ANS binding reveals common features of cytotoxic amyloid species. *ACS chemical biology* **2010,** *5* (8), 735-740.

63. Kayed, R.; Head, E.; Thompson, J. L.; McIntire, T. M.; Milton, S. C.; Cotman, C. W.; Glabe, C. G., Common structure of soluble amyloid oligomers implies common mechanism of pathogenesis. *Science* **2003,** *300* (5618), 486-489.

64. Dobson, C. M., Protein folding and disease: a view from the first Horizon Symposium. *Nature Reviews Drug Discovery* **2003,** *2* (2), 154-160.

65. Koo, E. H.; Lansbury, P. T.; Kelly, J. W., Amyloid diseases: abnormal protein aggregation in neurodegeneration. *Proceedings of the National Academy of Sciences* **1999,** *96* (18), 9989-9990.

66. Sunde, M.; Blake, C., The structure of amyloid fibrils by electron microscopy and X-ray diffraction. In *Advances in protein chemistry*, Elsevier: 1997; Vol. 50, pp 123-159.

67. Sunde, M.; Serpell, L. C.; Bartlam, M.; Fraser, P. E.; Pepys, M. B.; Blake, C. C., Common core structure of amyloid fibrils by synchrotron X-ray diffraction. *Journal of molecular biology* **1997,** *273* (3), 729-739.

68. O'Nuallain, B.; Wetzel, R., Conformational Abs recognizing a generic amyloid fibril epitope. *Proceedings of the National Academy of Sciences* **2002,** *99* (3), 1485-1490.

69. Biancalana, M.; Makabe, K.; Koide, A.; Koide, S., Molecular mechanism of thioflavin-T binding to the surface of β-rich peptide self-assemblies. *Journal of molecular biology* **2009,** *385* (4), 1052-1063.

70. Knowles, T. P.; Mezzenga, R., Amyloid fibrils as building blocks for natural and artificial functional materials. *Advanced Materials* **2016,** *28* (31), 6546-6561.

71. Chiti, F.; Webster, P.; Taddei, N.; Clark, A.; Stefani, M.; Ramponi, G.; Dobson, C. M., Designing conditions for in vitro formation of amyloid protofilaments and fibrils. *Proceedings of the National Academy of Sciences* **1999,** *96* (7), 3590-3594.

72. Aso, Y.; Shiraki, K.; Takagi, M., Systematic analysis of aggregates from 38 kinds of non disease-related proteins: identifying the intrinsic propensity of polypeptides to form amyloid fibrils. *Bioscience, biotechnology, and biochemistry* **2007,** *71* (5), 1313-1321.

73. DuBay, K. F.; Pawar, A. P.; Chiti, F.; Zurdo, J.; Dobson, C. M.; Vendruscolo, M., Prediction of the absolute aggregation rates of amyloidogenic polypeptide chains. *Journal of molecular biology* **2004,** *341* (5), 1317-1326.

74. Conchillo-Solé, O.; de Groot, N. S.; Avilés, F. X.; Vendrell, J.; Daura, X.; Ventura, S., AGGRESCAN: a server for the prediction and evaluation of" hot spots" of aggregation in polypeptides. *BMC bioinformatics* **2007,** *8* (1), 65.

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

76. Trovato, A.; Seno, F.; Tosatto, S. C., The PASTA server for protein aggregation prediction. *Protein Engineering, Design & Selection* **2007,** *20* (10), 521-523.

77. Tartaglia, G. G.; Vendruscolo, M., The Zyggregator method for predicting protein aggregation propensities. *Chemical Society Reviews* **2008,** *37* (7), 1395-1401.

78. Aguzzi, A.; Haass, C., Games played by rogue proteins in prion disorders and Alzheimer's disease. *Science* **2003,** *302* (5646), 814-818.

79. Hardy, J.; Selkoe, D. J., The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *science* **2002,** *297* (5580), 353-356.

80. Fowler, D. M.; Koulov, A. V.; Balch, W. E.; Kelly, J. W., Functional amyloid–from bacteria to humans. *Trends in biochemical sciences* **2007,** *32* (5), 217-224.

81. Chapman, M. R.; Robinson, L. S.; Pinkner, J. S.; Roth, R.; Heuser, J.; Hammar, M.; Normark,

S.; Hultgren, S. J., Role of Escherichia coli curli operons in directing amyloid fiber formation. *Science* **2002,** *295* (5556), 851-855.

82. Hervas, R.; Rau, M. J.; Park, Y.; Zhang, W.; Murzin, A. G.; Fitzpatrick, J. A.; Scheres, S. H.; Si, K., Cryo-EM structure of a neuronal functional amyloid implicated in memory persistence in Drosophila. *Science* **2020,** *367* (6483), 1230-1234.

83. Oh, J.; Kim, J.-G.; Jeon, E.; Yoo, C.-H.; Moon, J. S.; Rhee, S.; Hwang, I., Amyloidogenesis of type III-dependent harpins from plant pathogenic bacteria. *Journal of Biological Chemistry* **2007,** *282* (18), 13601-13609.

84. Fowler, D. M.; Koulov, A. V.; Alory-Jost, C.; Marks, M. S.; Balch, W. E.; Kelly, J. W., Functional amyloid formation within mammalian tissue. *PLoS Biol* **2005,** *4* (1), e6.

85. Maji, S. K.; Perrin, M. H.; Sawaya, M. R.; Jessberger, S.; Vadodaria, K.; Rissman, R. A.; Singru, P. S.; Nilsson, K. P. R.; Simon, R.; Schubert, D., Functional amyloids as natural storage of peptide hormones in pituitary secretory granules. *Science* **2009,** *325* (5938), 328-332.

86. Tycko, R.; Wickner, R. B., Molecular structures of amyloid and prion fibrils: consensus versus controversy. *Accounts of chemical research* **2013,** *46* (7), 1487-1496.

87. Makin, O. S.; Atkins, E.; Sikorski, P.; Johansson, J.; Serpell, L. C., Molecular basis for amyloid fibril formation and stability. *Proceedings of the National Academy of Sciences* **2005,** *102* (2), 315-320.

88. Sipe, J. D.; Cohen, A. S., History of the amyloid fibril. *Journal of structural biology* **2000,** *130* (2-3), 88-98.

89. Adler-Abramovich, L.; Aronov, D.; Beker, P.; Yevnin, M.; Stempler, S.; Buzhansky, L.; Rosenman, G.; Gazit, E., Self-assembled arrays of peptide nanotubes by vapour deposition. *Nature nanotechnology* **2009,** *4* (12), 849.

90. Romero, D.; Aguilar, C.; Losick, R.; Kolter, R., Amyloid fibers provide structural integrity to Bacillus subtilis biofilms. *Proceedings of the National Academy of Sciences* **2010,** *107* (5), 2230-2234.

91. Jacob, R. S.; Ghosh, D.; Singh, P. K.; Basu, S. K.; Jha, N. N.; Das, S.; Sukul, P. K.; Patil, S.; Sathaye, S.; Kumar, A., Self healing hydrogels composed of amyloid nano fibrils for cell culture and stem cell differentiation. *Biomaterials* **2015,** *54*, 97-105.

92. Li, C.; Adamcik, J.; Mezzenga, R., Biodegradable nanocomposites of amyloid fibrils and graphene with shape-memory and enzyme-sensing properties. *Nature nanotechnology* **2012,** *7* (7), 421.

93. Ryu, J.; Kim, S. W.; Kang, K.; Park, C. B., Mineralization of self-assembled peptide nanofibers for rechargeable lithium ion batteries. *Advanced Materials* **2010,** *22* (48), 5537-5541.

94. Wakabayashi, R.; Suehiro, A.; Goto, M.; Kamiya, N., Designer aromatic peptide amphiphiles for self-assembly and enzymatic display of proteins with morphology control. *Chemical communications* **2019,** *55* (5), 640-643.

95. Chernoff, Y. O.; Lindquist, S. L.; Ono, B.-i.; Inge-Vechtomov, S. G.; Liebman, S. W., Role of the chaperone protein Hsp104 in propagation of the yeast prion-like factor [psi+]. *Science* **1995,** *268* (5212), 880-884.

96. Taguchi, H.; Kawai-Noma, S., Amyloid oligomers: diffuse oligomer-based transmission of yeast prions. *The FEBS journal* **2010,** *277* (6), 1359-1368.

97. Kushnirov, V. V.; Vishnevskaya, A. B.; Alexandrov, I. M.; Ter-Avanesyan, M. D., Prion and nonprion amyloids: a comparison inspired by the yeast Sup35 protein. *Prion* **2007,** *1* (3), 179-184.

98. Horwich, A. L.; Weissman, J. S., Deadly conformations—protein misfolding in prion disease. *Cell* **1997,** *89* (4), 499-510.

99. Alper, T.; Cramp, W.; Haig, D. A.; Clarke, M. C., Does the agent of scrapie replicate without nucleic acid? *Nature* **1967,** *214* (5090), 764-766.

100. Griffith, J. S., Nature of the scrapie agent: Self-replication and scrapie. *Nature* **1967,** *215* (5105), 1043-1044.

101. Wickner, R. B., Yeast and fungal prions. *Cold Spring Harbor perspectives in biology* **2016,** *8* (9), a023531.

102. Suzuki, G.; Shimazu, N.; Tanaka, M., A yeast prion, Mod5, promotes acquired drug resistance and cell survival under environmental stress. *Science* **2012,** *336* (6079), 355-359.

103. Chien, P.; Weissman, J. S.; DePace, A. H., Emerging principles of conformation-based prion inheritance. *Annual review of biochemistry* **2004,** *73* (1), 617-656.

104. Uptain, S. M.; Lindquist, S., Prions as protein-based genetic elements. *Annual Reviews in Microbiology* **2002,** *56* (1), 703-741.

105. Ross, E. D.; Minton, A.; Wickner, R. B., Prion domains: sequences, structures and interactions. *Nature cell biology* **2005,** *7* (11), 1039-1044.

106. Hafner-Bratkovič, I.; Bester, R.; Pristovšek, P.; Gaedtke, L.; Veranič, P.; Gašperšič, J.; Manček-Keber, M.; Avbelj, M.; Polymenidou, M.; Julius, C., Globular domain of the prion protein needs to be unlocked by domain swapping to support prion protein conversion. *Journal of Biological Chemistry* **2011,** *286* (14), 12149-12156.

107. Aigle, M.; Lacroute, F., Genetical aspects of [URE3], a non-mitochondrial, cytoplasmically inherited mutation in yeast. *Molecular and General Genetics MGG* **1975,** *136* (4), 327-335.

108. Wickner, R. B., [URE3] as an altered URE2 protein: evidence for a prion analog in Saccharomyces cerevisiae. *Science* **1994,** *264* (5158), 566-569.

109. Cox, B., ψ, a cytoplasmic suppressor of super-suppressor in yeast. *Heredity* **1965,** *20* (4), 505-521.

110. Chernoff, Y. O.; Derkach, I. L.; Inge-Vechtomov, S. G., Multicopy SUP35 gene induces de-novo appearance of psi-like factors in the yeast Saccharomyces cerevisiae. *Current genetics* **1993,** *24* (3), 268-270.

111. Lund, P.; Cox, B., Reversion analysis of [psi−] mutations in Saccharomyces cerevisiae. *Genetics Research* **1981,** *37* (2), 173-182.

112. Wickner, R. B.; Shewmaker, F. P.; Bateman, D. A.; Edskes, H. K.; Gorkovskiy, A.; Dayani, Y.; Bezsonov, E. E., Yeast prions: structure, biology, and prion-handling systems. *Microbiol. Mol. Biol. Rev.* **2015,** *79* (1), 1-17.

113. Derkatch, I. L.; Bradley, M. E.; Hong, J. Y.; Liebman, S. W., Prions affect the appearance of other prions: the story of [PIN+]. *Cell* **2001,** *106* (2), 171-182.

114. Du, Z.; Park, K.-W.; Yu, H.; Fan, Q.; Li, L., Newly identified prion linked to the chromatin-remodeling factor Swi1 in Saccharomyces cerevisiae. *Nature genetics* **2008,** *40* (4), 460-465.

115. Alberti, S.; Halfmann, R.; King, O.; Kapila, A.; Lindquist, S., A systematic survey identifies prions and illuminates sequence features of prionogenic proteins. *Cell* **2009,** *137* (1), 146-158.

116. Glover, J. R.; Kowal, A. S.; Schirmer, E. C.; Patino, M. M.; Liu, J.-J.; Lindquist, S., Self-seeded fibers formed by Sup35, the protein determinant of [PSI+], a heritable prion-like factor of S. cerevisiae. *Cell* **1997,** *89* (5), 811-819.

117. Taylor, K. L.; Cheng, N.; Williams, R. W.; Steven, A. C.; Wickner, R. B., Prion domain initiation of amyloid formation in vitro from native Ure2p. *Science* **1999,** *283* (5406), 1339-1343.

118. Masison, D. C.; Wickner, R. B., Prion-inducing domain of yeast Ure2p and protease resistance of Ure2p in prion-containing cells. *Science* **1995,** *270* (5233), 93-95.

119. Paushkin, S. V.; Kushnirov, V. V.; Smirnov, V. N.; Ter-Avanesyan, M. D., In vitro propagation of the prion-like state of yeast Sup35 protein. *Science* **1997,** *277* (5324), 381-383.

120. King, C.-Y.; Diaz-Avalos, R., Protein-only transmission of three yeast prion strains. *Nature* **2004,** *428* (6980), 319-323.

121. Toombs, J. A.; McCarty, B. R.; Ross, E. D., Compositional determinants of prion formation in yeast. *Molecular and cellular biology* **2010,** *30* (1), 319-332.

122. Shewmaker, F.; Wickner, R. B.; Tycko, R., Amyloid of the prion domain of Sup35p has an in-register parallel β-sheet structure. *Proceedings of the National Academy of Sciences* **2006,** *103* (52), 19754-19759.

123. Baxa, U.; Taylor, K. L.; Wall, J. S.; Simon, M. N.; Cheng, N.; Wickner, R. B.; Steven, A. C., Architecture of Ure2p Prion Filaments THE N-TERMINAL DOMAINS FORM A CENTRAL CORE FIBER. *Journal of Biological Chemistry* **2003,** *278* (44), 43717-43727.

124. Baxa, U.; Speransky, V.; Steven, A. C.; Wickner, R. B., Mechanism of inactivation on prion conversion of the Saccharomyces cerevisiae Ure2 protein. *Proceedings of the National Academy of Sciences* **2002,** *99* (8), 5253-5260.

125. Wickner, R. B.; Shewmaker, F.; Edskes, H.; Kryndushkin, D.; Nemecek, J.; McGlinchey, R.; Bateman, D.; Winchester, C.-L., Prion amyloid structure explains templating: how proteins can be genes. *FEMS yeast research* **2010,** *10* (8), 980-991.

126. Glass, N. L.; Dementhon, K., Non-self recognition and programmed cell death in filamentous

fungi. *Current opinion in microbiology* **2006,** *9* (6), 553-558.

127. Wasmer, C.; Lange, A.; Van Melckebeke, H.; Siemer, A. B.; Riek, R.; Meier, B. H., Amyloid fibrils of the HET-s (218–289) prion form a β solenoid with a triangular hydrophobic core. *Science* **2008,** *319* (5869), 1523-1526.

128. Vázquez-Fernández, E.; Vos, M. R.; Afanasyev, P.; Cebey, L.; Sevillano, A. M.; Vidal, E.; Rosa, I.; Renault, L.; Ramos, A.; Peters, P. J.; Fernández, J. J.; van Heel, M.; Young, H. S.; Requena, J. R.; Wille, H., The Structural Architecture of an Infectious Mammalian Prion Using Electron Cryomicroscopy. *PLoS pathogens* **2016,** *12* (9), e1005835.

129. Spagnolli, G.; Rigoli, M.; Orioli, S.; Sevillano, A. M.; Faccioli, P.; Wille, H.; Biasini, E.; Requena, J. R., Full atomistic model of prion structure and conversion. *PLoS pathogens* **2019,** *15* (7), e1007864.

130. Schlieker, C.; Tews, I.; Bukau, B.; Mogk, A., Solubilization of aggregated proteins by ClpB/DnaK relies on the continuous extraction of unfolded polypeptides. *FEBS letters* **2004,** *578* (3), 351-356.

131. Winkler, J.; Tyedmers, J.; Bukau, B.; Mogk, A., Hsp70 targets Hsp100 chaperones to substrates for protein disaggregation and prion fragmentation. *Journal of Cell Biology* **2012,** *198* (3), 387-404.

132. Ter-Avanesyan, M. D.; Dagkesamanskaya, A. R.; Kushnirov, V. V.; Smirnov, V. N., The SUP35 omnipotent suppressor gene is involved in the maintenance of the non-Mendelian determinant [psi+] in the yeast Saccharomyces cerevisiae. *Genetics* **1994,** *137* (3), 671-676.

133. Bradley, M. E.; Liebman, S. W., The Sup35 domains required for maintenance of weak, strong or undifferentiated yeast [PSI+] prions. *Molecular microbiology* **2004,** *51* (6), 1649-1659.

134. Liu, J.-J.; Sondheimer, N.; Lindquist, S. L., Changes in the middle region of Sup35 profoundly alter the nature of epigenetic inheritance for the yeast prion [PSI+]. *Proceedings of the National Academy of Sciences* **2002,** *99* (suppl 4), 16446-16453.

135. Masison, D. C.; Maddelein, M.-L.; Wickner, R. B., The prion model for [URE3] of yeast: spontaneous generation and requirements for propagation. *Proceedings of the National Academy of Sciences* **1997,** *94* (23), 12503-12508.

136. Magasanik, B.; Kaiser, C. A., Nitrogen regulation in Saccharomyces cerevisiae. *Gene* **2002,** *290* (1-2), 1-18.

137. Patino, M. M.; Liu, J.-J.; Glover, J. R.; Lindquist, S., Support for the prion hypothesis for inheritance of a phenotypic trait in yeast. *Science* **1996,** *273* (5275), 622-626.

138. Bai, M.; Zhou, J.-M.; Perrett, S., The yeast prion protein Ure2 shows glutathione peroxidase activity in both native and fibrillar forms. *Journal of Biological Chemistry* **2004,** *279* (48), 50025-50030.

139. Baxa, U.; Keller, P. W.; Cheng, N.; Wall, J. S.; Steven, A. C., In Sup35p filaments (the [PSI+] prion), the globular C-terminal domains are widely offset from the amyloid fibril backbone. *Molecular microbiology* **2011,** *79* (2), 523-532.

140. Kryndushkin, D. S.; Wickner, R. B.; Tycko, R., The core of Ure2p prion fibrils is formed by the N-terminal segment in a parallel cross-β structure: evidence from solid-state NMR. *Journal of molecular biology* **2011,** *409* (2), 263-277.

141.Zhou, X. M.; Entwistle, A.; Zhang, H.; Jackson, A. P.; Mason, T. O.; Shimanovich, U.; Knowles, T. P.; Smith, A. T.; Sawyer, E. B.; Perrett, S., Self-assembly of amyloid fibrils that display active enzymes. *ChemCatChem* **2014,** *6* (7), 1961-1968.

142. Men, D.; Guo, Y.-C.; Zhang, Z.-P.; Wei, H.-p.; Zhou, Y.-F.; Cui, Z.-Q.; Liang, X.-S.; Li, K.; Leng, Y.; You, X.-Y., Seeding-induced self-assembling protein nanowires dramatically increase the sensitivity of immunoassays. *Nano letters* **2009,** *9* (6), 2246-2250.

143. Schmuck, B.; Sandgren, M.; Härd, T., A fine-tuned composition of protein nanofibrils yields an upgraded functionality of displayed antibody binding domains. *Biotechnology journal* **2017,** *12* (6), 1600672.

144. Men, D.; Zhang, Z.-P.; Guo, Y.-C.; Zhu, D.-H.; Bi, L.-J.; Deng, J.-Y.; Cui, Z.-Q.; Wei, H.-P.; Zhang, X.-E., An auto-biotinylated bifunctional protein nanowire for ultra-sensitive molecular biosensing. *Biosensors and Bioelectronics* **2010,** *26* (4), 1137-1141.

145. Leng, Y.; Wei, H. P.; Zhang, Z. P.; Zhou, Y. F.; Deng, J. Y.; Cui, Z. Q.; Men, D.; You, X. Y.; Yu, Z. N.; Luo, M., Integration of a Fluorescent Molecular Biosensor into Self-Assembled Protein Nanowires: A Large Sensitivity Enhancement. *Angewandte Chemie International Edition* **2010,** *49* (40), 7243-7246.

146. Zhou, X.-M.; Shimanovich, U.; Herling, T. W.; Wu, S.; Dobson, C. M.; Knowles, T. P.; Perrett, S., Enzymatically active microgels from self-assembling protein nanofibrils for microflow chemistry. *ACS nano* **2015,** *9* (6), 5772-5781.

147. Altamura, L.; Horvath, C.; Rengaraj, S.; Rongier, A.; Elouarzaki, K.; Gondran, C.; Maçon, A. L.; Vendrely, C.; Bouchiat, V.; Fontecave, M., A synthetic redox biofilm made from metalloprotein–prion domain chimera nanowires. *Nature chemistry* **2017,** *9* (2), 157-163.

148. Schmuck, B.; Gudmundsson, M.; Blomqvist, J.; Hansson, H.; Härd, T.; Sandgren, M., Production of ready-to-use functionalized Sup35 nanofibrils secreted by Komagataella pastoris. *ACS nano* **2018,** *12* (9), 9363-9371.

149. DiCosimo, R.; McAuliffe, J.; Poulose, A. J.; Bohlmann, G., Industrial use of immobilized enzymes. *Chemical Society Reviews* **2013,** *42* (15), 6437-6474.

150. Toombs, J. A.; Petri, M.; Paul, K. R.; Kan, G. Y.; Ben-Hur, A.; Ross, E. D., De novo design of synthetic prion domains. *Proceedings of the National Academy of Sciences* **2012,** *109* (17), 6519-6524.

151. Ross, E. D.; Edskes, H. K.; Terry, M. J.; Wickner, R. B., Primary sequence independence for prion formation. *Proceedings of the National Academy of Sciences* **2005,** *102* (36), 12825-12830.

152. Sabate, R.; Rousseau, F.; Schymkowitz, J.; Batlle, C.; Ventura, S., Amyloids or prions? That is the question. *Prion* **2015,** *9* (3), 200-206.

153. Sant'Anna, R.; Fernández, M. R.; Batlle, C.; Navarro, S.; De Groot, N. S.; Serpell, L.; Ventura,

S., Characterization of amyloid cores in prion domains. *Scientific reports* **2016,** *6* (1), 1-10.

154. Batlle, C.; de Groot, N. S.; Iglesias, V.; Navarro, S.; Ventura, S., Characterization of soft amyloid cores in human prion-like proteins. *Scientific reports* **2017,** *7* (1), 1-16.

155. Kawai-Noma, S.; Pack, C.-G.; Kojidani, T.; Asakawa, H.; Hiraoka, Y.; Kinjo, M.; Haraguchi, T.; Taguchi, H.; Hirata, A., In vivo evidence for the fibrillar structures of Sup35 prions in yeast cells. *Journal of Cell Biology* **2010,** *190* (2), 223-231.

156. Duernberger, Y.; Liu, S.; Riemschoss, K.; Paulsen, L.; Bester, R.; Kuhn, P.-H.; Schölling, M.; Lichtenthaler, S. F.; Vorberg, I., Prion replication in the mammalian cytosol: Functional regions within a prion domain driving induction, propagation, and inheritance. *Molecular and cellular biology* **2018,** *38* (15), e00111-18.

157. Zambrano, R.; Conchillo-Sole, O.; Iglesias, V.; Illa, R.; Rousseau, F.; Schymkowitz, J.; Sabate, R.; Daura, X.; Ventura, S., PrionW: a server to identify proteins containing glutamine/asparagine rich prion-like domains and their amyloid cores. *Nucleic acids research* **2015,** *43* (W1), W331-W337.

158. Díaz-Caballero, M.; Navarro, S.; Fuentes, I.; Teixidor, F.; Ventura, S., Minimalist prion-inspired polar self-assembling peptides. *ACS nano* **2018,** *12* (6), 5394-5407.