UAB

Universitat Autònoma de Barcelona

Universitat Autònoma de Barcelona

Departament de Bioquímica i Biologia Molecular

Institut de Biotecnologia i Biomedicina

DESIGN OF IMPROVED THERAPEUTIC PROTEINS AND NOVEL PROTEIN-BASED NANOMATERIALS

Doctoral thesis submitted by Marcos Gil García for the degree of PhD in Biochemistry, Molecular Biology and Biomedicine from the Universitat Autònoma de Barcelona.

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

Marcos Gil García   Prof. Salvador Ventura Zamora

Bellaterra, 2022

SUMMARY

Proteins are involved in a myriad of biological processes such as catalysis, transport, regulation, defense, and providing structure to the cell. Most proteins need to fold into a defined 3D structure to perform such functions. However, this energetically directed process can be kinetically disturbed, resulting in the formation of stable misfolded supramolecular assemblies, including different types of protein aggregates. Protein aggregation stems from the establishment of aberrant intermolecular interactions and has been associated with the onset of several degenerative diseases. In addition, this unwanted reaction limits the development of proteins of biotechnological interest, impacting the production, storage, and commercialization of protein-based drugs. In this regard, strategies aimed to diminish the impact of protein aggregation on therapeutic proteins are pivotal to ensure their correct development as safe and active drugs. In the present doctoral thesis, we demonstrate that a structure-based aggregation predictor, which considers protein stability, successfully assists the redesign of two structurally unrelated proteins, improving their solubility without compromising their active conformation. This approach might replace the expensive trial-and-error assays employed by the pharmaceutical industry, providing an economical alternative for accelerating the development of protein-based drugs.

Proteins are widely understood as the building blocks of life and can act as self-assembling entities involved in the creation of different supramolecular structures. Hence, there is an increasing interest in using polypeptides to build up functional and biocompatible nanomaterials; among them, protein inclusion bodies (IBs) have emerged as an attractive architecture. Traditionally considered useless protein deposits formed by misfolded conformations, recent data converge to indicate that IBs act as reservoirs of stable and active protein. These submicrometric particles are produced efficiently and cost-effectively, and are usually sustained by an amyloid-like scaffold where the protein of interest is trapped. Nevertheless, this amyloid conformation necessarily impacts the activity of the protein of interest and can be potentially cytotoxic. To overcome these drawbacks, we aimed at generating novel and improved functional IBs based on the α-helical architecture characteristic of coiled-coils. Using a naturally encoded coiled-coil domain as the scaffolding entity, α-helix-rich and biocompatible functional IBs were obtained in a ready-to-use form. Interestingly, coiled-coil-based IBs present a higher specific activity than their amyloid-like counterparts, since their assembly is guided by native interactions that

do not interfere with the folding of the functional moieties. We demonstrate that these coiled-coil-inspired protein nanoparticles can display fluorescent and antibody-capturing activities simultaneously, being biocompatible, stable, and targeting specific antigens when decorated with a single or a combination of antibodies; thus, emerging as a promising technology for biomedical applications.

Overall, the work described in the present thesis attempts to provide useful strategies aimed at (I) redesigning therapeutic proteins with enhanced biophysical properties employing *in silico* tools and (II) developing versatile and tunable multifunctional protein-based nanomaterials sustained by α-helical interactions.

RESUMEN

Las proteínas están involucradas en innumerables procesos biológicos como catálisis, transporte, regulación, defensa y estructura celular. Para poder llevar a cabo estas funciones, la gran mayoría de ellas necesitan plegarse en una estructura 3D definida. Sin embargo, este proceso energéticamente favorable puede ser perturbado cinéticamente, dando lugar a la formación de ensamblajes supramoleculares formados por proteínas desplegadas, los agregados proteicos. La agregación de proteínas consiste en el establecimiento de interacciones intermoleculares aberrantes, estando estrechamente relacionada con la aparición de varias enfermedades degenerativas. Además, esta reacción indeseada limita el desarrollo de proteínas con interés biotecnológico, afectando los procesos de producción, almacenamiento y su posterior comercialización. De este modo, las estrategias dirigidas a disminuir el impacto de los procesos de agregación en proteínas terapéuticas son imprescindibles para asegurar el correcto desarrollo de éstas como fármacos seguros y activos. En la presente tesis doctoral, demostramos que un predictor de agregación basado en estructura, y que además tiene en cuenta la estabilidad proteica, asiste de manera exitosa el rediseño de dos proteínas sin relación estructural, mejorando su solubilidad sin comprometer su conformación activa. Esta estrategia podría reemplazar los costosos ensayos de prueba-y-error llevados a cabo por la industria farmacéutica, proporcionando una alternativa económica destinada a acelerar el desarrollo de proteínas terapéuticas.

Las proteínas son comúnmente conocidas por ser los cimientos de la vida, pudiendo actuar como entidades capaces de autoensamblarse, dando lugar a la formación de varias estructuras supramoleculares. Por ello, hay un gran interés en el uso de polipéptidos para crear nanomateriales funcionales y biocompatibles; entre los que destacan los cuerpos de inclusión proteicos (IBs). Aunque estos IBs han sido tradicionalmente considerados como depósitos de proteínas inservibles, los datos más recientes indican que los IBs actúan como reservorios de proteína activa y estable. Estas partículas submicrométricas son producidas de manera eficiente y barata, y normalmente están sostenidas por un esqueleto amiloide en el cual reside la proteína de interés. Sin embargo, esta conformación amiloide tiene un impacto en la actividad de la proteína de interés y puede ser potencialmente citotóxica. Para superar estas limitaciones, nuestra estrategia consistió en la creación de nuevos IBs funcionales usando la estructura de hélice-α, característica de los *coiled-coils*. De este

modo, usando un dominio *coiled-coil* natural como andamio, se obtuvieron IBs funcionales, biocompatibles, ricos en hélice-α y listos para ser usados. De manera interesante, los IBs formados por *coiled-coils* tienen una actividad específica mayor que sus homólogos amiloides, ya que su ensamblaje viene guiado por interacciones nativas que no interfieren en el correcto plegamiento de las partes funcionales. En esta tesis doctoral, demostramos que las nanopartículas proteicas inspiradas en la estructura *coiled-coil* pueden albergar fluorescencia y tener la capacidad de unir anticuerpos de manera simultánea, siendo estables y biocompatibles, y pudiendo dirigirse hacia antígenos específicos cuando son decoradas con un anticuerpo o una combinación de éstos, siendo por lo tanto una tecnología prometedora en biomedicina.

En resumen, el trabajo descrito en la presente tesis trata de proporcionar estrategias útiles destinadas a (I) rediseñar proteínas terapéuticas con propiedades biofísicas mejoradas mediante el uso de herramientas *in silico* y (II) desarrollar nanomateriales proteicos versátiles y multifuncionales, basados en interacciones entre hélices-α.

RESUM

Les proteïnes estan involucrades en innombrables processos biològics com catàlisis, transport, regulació, defensa i estructura cel·lular. Per a poder dur a terme aquestes funcions, la gran majoria d'elles necessiten plegar-se en una estructura 3D definida. No obstant això, aquest procés energèticament favorable pot ser pertorbat cinèticament, donant lloc a la formació d'ensamblatges supramoleculars formats per proteïnes desplegades, els agregats proteics. L'agregació de proteïnes consisteix en l'establiment d'interaccions intermoleculars aberrants, estant estretament relacionada amb l'aparició de diverses malalties degeneratives. A més, aquesta reacció no desitjada limita el desenvolupament de proteïnes amb interès biotecnològic, afectant els processos de producció, emmagatzematge i la seva posterior comercialització. D'aquesta manera, les estratègies dirigides a disminuir l'impacte dels processos d'agregació en proteïnes terapèutiques són imprescindibles per a assegurar el correcte desenvolupament d'aquestes com a fàrmacs segurs i actius. En la present tesi doctoral, vam demostrar que un predictor d'agregació basat en estructura, i que a més té en compte l'estabilitat proteica, assisteix amb èxit el redisseny de dues proteïnes sense relació estructural, millorant la seva solubilitat sense comprometre la seva conformació activa. Aquesta estratègia podria reemplaçar els costosos assajos de prova-i-error duts a terme per la indústria farmacèutica, proporcionant una alternativa econòmica destinada a accelerar el desenvolupament de proteïnes terapèutiques.

Les proteïnes són comunament conegudes per ser els fonaments de la vida, podent actuar com a entitats capaces d'autoensamblar-se, donant lloc a la formació de diverses estructures supramoleculars. Per això, hi ha un gran interès en l'ús de polipèptids per a crear nanomaterials funcionals i biocompatibles, entre els quals destaquen els cossos d'inclusió proteics (IBs). Encara que aquests IBs han estat tradicionalment considerats com a dipòsits de proteïnes inservibles, les dades més recents indiquen que els IBs actuen com a dipòsits de proteïna activa i estable. Aquestes partícules submicromètriques són produïdes de manera eficient i barata, i normalment estan sostingudes per un esquelet amiloide en el qual resideix la proteïna d'interès. No obstant això, aquesta conformació amiloide té un impacte en l'activitat de la proteïna d'interès i pot ser potencialment citotòxica. Per a superar aquestes limitacions, la nostra estratègia va consistir en la creació de nous IBs funcionals fent servir l'estructura d'hèlix-α, característica dels *coiled-coils*.

D'aquesta manera, usant un domini *coiled-coil* natural com a bastida, es van obtenir IBs funcionals, biocompatibles, rics en hèlix-α i llestos per a ser usats. De manera interessant, els IBs formats per *coiled-coils* tenen una major activitat específica comparat amb els seus homòlegs amiloides, ja que el seu ensamblatge ve guiat per interaccions natives que no interfereixen en el correcte plegament de les parts funcionals. En aquesta tesi doctoral, vam demostrar que les nanopartícules proteiques inspirades en l'estructura *coiled-coil* poden albergar fluorescència i tenen la capacitat d'unir anticossos de manera simultània, sent estables i biocompatibles, i podent dirigir-se cap a antígens específics quan són decorades amb un anticòs o una combinació d'aquests, sent per tant una tecnologia prometedora en biomedicina.

En resum, el treball descrit en la present tesi tracta de proporcionar estratègies útils destinades a (I) redissenyar proteïnes terapèutiques amb propietats biofísiques millorades mitjançant l'ús d'eines *in silico* i (II) desenvolupar nanomaterials proteics versàtils i multifuncionals, basats en interaccions entre hèlixs-α.

CONTENTS

LIST OF PUBLICATIONS

This thesis is composed of the following published works:

1. **Gil-Garcia M**, Bañó-Polo M, Varejão N, Jamroz M, Kuriata A, Díaz-Caballero M, Lascorz J, Morel B, Navarro S, Reverter D, Kmiecik S, Ventura S. Combining Structural Aggregation Propensity and Stability Predictions to Redesign Protein Solubility. Mol Pharm. 2018 Sep 4;15(9):3846-3859. DOI: 10.1021/acs.molpharmaceut.8b00341.

2. **Gil-Garcia M**, Navarro S, Ventura S. Coiled-coil inspired functional inclusion bodies. Microb Cell Fact. 2020 Jun 1;19(1):117. DOI: 10.1186/s12934-020-01375-4.

3. **Gil-Garcia M**, Ventura S. Multifunctional antibody-conjugated coiled-coil protein nanoparticles for selective cell targeting. Acta Biomater. 2021 Sep 1;131:472-482. DOI: 10.1016/j.actbio.2021.06.040.

4. **Gil-Garcia M**, Ventura S. Coiled-Coil Based Inclusion Bodies and Their Potential Applications. Front Bioeng Biotechnol. 2021 Aug 17;9:734068. DOI: 10.3389/fbioe.2021.734068.

Other co-authored articles not included in this thesis:

5. Fernández MR, Batlle C, **Gil-García M**, Ventura S. Amyloid cores in prion domains: Key regulators for prion conformational conversion. Prion. 2017 Jan 2;11(1):31-39. DOI: 10.1080/19336896.2017.1282020.

6. Fraga H, Pujols J, **Gil-Garcia M**, Roque A, Bernardo-Seisdedos G, Santambrogio C, Bech-Serra JJ, Canals F, Bernadó P, Grandori R, Millet O, Ventura S. Disulfide driven folding for a conditionally disordered protein. Sci Rep. 2017 Dec 5;7(1):16994. DOI: 10.1038/s41598-017-17259-4.

7. Wang W, **Gil-Garcia M**, Ventura S. Dual Antibody-Conjugated Amyloid Nanorods to Promote Selective Cell-Cell Interactions. ACS Appl Mater Interfaces. 2021 Apr 7;13(13):14875-14884. DOI: 10.1021/acsami.0c21996.

8. Batlle C, Calvo I, Iglesias V, J Lynch C, **Gil-Garcia M**, Serrano M, Ventura S. MED15 prion-like domain forms a coiled-coil responsible for its amyloid conversion and propagation. Commun Biol. 2021 Mar 26;4(1):414. DOI: 10.1038/s42003-021-01930-8.

9. **Gil-Garcia M**, Iglesias V, Pallarès I, Ventura S. Prion-like proteins: from computational approaches to proteome-wide analysis. FEBS Open Bio. 2021 Sep;11(9):2400-2417. DOI: 10.1002/2211-5463.13213.

ABBREVIATIONS

3D structure	Three-dimensional structure
A3D	AGGRESCAN 3D
A3D 2.0	AGGRESCAN 3D 2.0
ADC	Antibody-drug conjugate
AI	Artificial intelligence
APR	Aggregation-prone region
Aß	Amyloid-β peptide
Å	Angstrom
BsAb	Bispecific antibody
CD	Circular dichroism
CDR	Complementarity-determining region
CR	Congo Red
CryoEM	Cryo-electron microscopy
DI	Developability index
DLS	Dynamic light scattering
DNA	Deoxyribonucleic acid
DSC	Differential scanning calorimetry
E. coli	*Escherichia coli*
EGFR	Epidermal growth factor receptor
FC	Flow cytometry
FDA	U.S. Food and Drug Administration
FTIR	Fourier transform infrared spectroscopy
GFP	Green fluorescent protein
GRAVY	Grand average of hydropathicity

GuHCl	Guanidinium hydrochloride
IB	Inclusion Body
IDR/IDP	Intrinsically disordered region/protein
IgG	Immunoglobulin G
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LLPS	Liquid-liquid phase separation
mAb	Monoclonal antibody
MD	Molecular dynamics
PBS	Phosphate buffered saline
PDB	Protein data bank
pI	Isoelectric point
SAC	Soft amyloid core
SAP	Spatial aggregation propensity
scFv	Single-chain variable fragment
sdAb	Single-domain antibody
SEM	Scanning electron microscopy
ssNMR	Solid-state nuclear magnetic resonance
STAP	Structural aggregation propensity
TEM	Transmission electron microscopy
TFE	2,2,2-Trifluoroethanol
Th-T	Thioflavin-T
V_H fragment	Variable heavy fragment
V_{HH} antibody	Camelid single-domain antibody
V_L fragment	Variable light fragment
wt	*Wild-type*

INTRODUCTION

1. Proteins and their folding

Proteins are biomacromolecules formed by a combination of 20 amino acids covalently joined by peptide bonds, with their side chains encoding for defined physicochemical properties. Considering the cellular context and quoting to Dr. Arthur M Lesk; *"In the drama of life on a molecular scale, proteins are where the action is"*, proteins are the workhorses of the cell and are involved in a myriad of molecular functions such as transport, catalysis, signalling, gene regulation, and structure. In order to execute these functions, the majority of proteins need to acquire a specific three-dimensional structure (3D structure), known as the native state, promoted by the establishment of non-covalent interactions between neighbouring residues in the polypeptide chain. This energetically and kinetically favourable process is known as protein folding (Dobson, 2003).

Proteins are structured according to different levels of organization: The first stage is the **primary structure**, corresponding to the amino acid sequence and sustained by peptide bonds between contiguous residues. Then, the **secondary structure**, is dictated by the primary sequence, and mainly stems from the establishment of hydrogen bonds within the peptide backbone and gives rise to the formation of regular secondary patterns like α-helices and ß-sheets. Subsequently, proteins form a highly compacted structure, the **tertiary structure**, promoted by interactions between the amino acid' side chains of the polypeptide (hydrogen bonds, Van der Waals packing, electrostatic interactions and disulfide bonds), and, in some cases, different polypeptide chains can establish non-covalent intermolecular interactions, leading to the acquisition of the **quaternary structure** (Herczenik and Gebbink, 2008) (Figure 1).
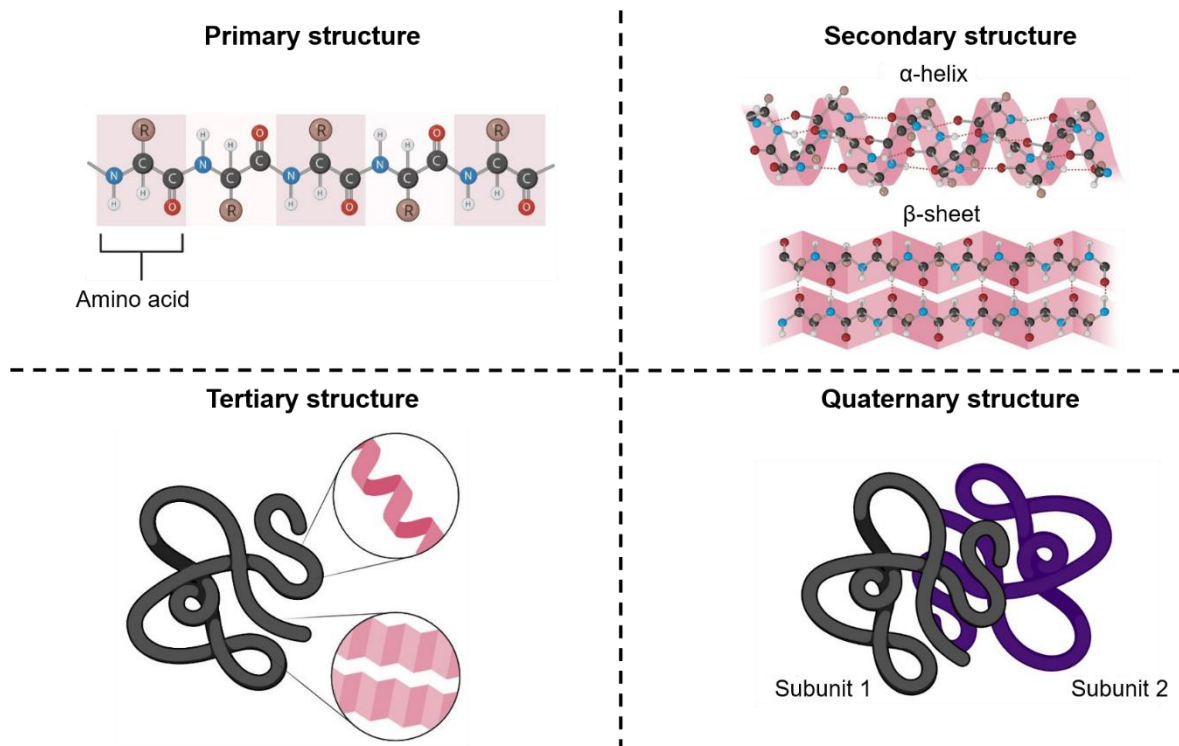
Figure 1. Representation of the different structural organization levels of proteins. The primary sequence consists of a specific sequence of amino acids linked by peptide bonds. The secondary structure is achieved by a local folding of the protein prompted by hydrogen bonding within the peptide backbone. The tertiary structure is a 3D folding pattern sustained by amino acid' side chain interactions, and the quaternary structure arises from non-covalent interactions between polypeptide chains. Adapted from "Protein Structure", by BioRender.com (2021). Retrieved from https://app.biorender.com/biorender-templates.

The folding of a polypeptide from unstructured conformations to a functional 3D structure is a key process in biochemistry and cell biology. Errors during this process drive proteins to populate partially or totally unfolded states, promoting non-native intermolecular interactions related to the formation of protein aggregates. This aberrant oligomerization is associated with the onset of over 40 human disorders such as Alzheimer's disease, Parkinson's disease and type II diabetes (Chiti and Dobson, 2017). A correct understanding of folding mechanisms is an essential requirement for the successful prediction of protein structures and the design and redesign of improved proteins (Baker, 2000; Woolfson, 2021). In this sense, outstanding advances in the prediction of the protein structure have been recently achieved. An artificial intelligence (AI) algorithm developed by Google DeepMind, known as AlphaFold, provided accurate predictions of protein structures departing from the protein sequence (Senior et al., 2020). Strikingly, a novel version of the original program, AlphaFold2, improved its accuracy to a level comparable with

experimental data in the last biennial Critical Assessment of Techniques for Protein Structure Prediction (CASP) challenge, a tournament aimed at assessing the state-of-the-art of protein structure modelling (Jumper et al., 2021b, 2021a). This scientific breakthrough might contribute to solve the "protein folding problem", and pave the way for a burgeoning era of rational protein design. Indeed, Google DeepMind in collaboration with the European Molecular Biology Laboratory (EMBL) has released the most complete dataset of predicted protein structures, including those of all human proteins and 20 additional organisms relevant for biological research (Tunyasuvunakool et al., 2021). These recent breakthroughs are freely available to the scientific community, and demonstrate the pivotal role that AI might play in the discovery of new therapeutic approaches for the treatment of protein-associated diseases.

One of the first insights into the protein folding problem come from the well-known studies of Christian B. Anfinsen in the early '60s. Anfinsen and co-workers focused on the refolding process of the enzyme ribonuclease and demonstrated the importance of the specific amino acid sequence to achieve the native state (Anfinsen et al., 1961). Furthermore, Anfinsen observed that the acquisition of the 3D structure was a spontaneous process, leading to the "Anfinsen's dogma" that asserts that the 3D, folded and functional state of proteins in their biological contexts corresponds to the most stable state and is only determined by the primary sequence (Anfinsen, 1973). However, several exceptions of the rule have been reported. For instance, the native state does not necessarily imply the lowest energetic level, since more stable conformations can be reached under physiological conditions (e.g., amyloid fibrils). Moreover, the existence of chameleon sequences able to adopt different secondary structures, together with the discovery of functional intrinsically disordered regions/proteins (IDRs or IDPs) lacking a defined 3D structure, defy this postulate (Bahramali et al., 2017; Oldfield and Dunker, 2014; Pinheiro et al., 2021).

Afterwards, the work of Cyrus Levinthal illustrated how the required time for a protein to acquire its native conformation by random interactions would be enormous. However, proteins spontaneously adopt their functional conformation in the range of milli- or even micro-seconds, leading Levinthal to propose specificity in the protein folding process, without the need to explore the whole conformational space (Levinthal, 1968).

In good agreement with Levinthal's ideas, several studies have revealed the crucial role of native-like interactions to guide the protein folding process towards a low energetic state (Daggett and Fersht, 2003). Different folding models have been proposed to explain the surprisingly high speed of this process in comparison with a random process. These models are: (I) The "framework-model" based on the establishment of secondary structure elements as the first formed units (Karplus and Weaver, 1976), (II) the "hydrophobic collapse" where hydrophobic interactions promote a rapid compaction of the structure, limiting the available conformational options (Tanford, 1962), and (III) the "nucleation-condensation" model that combines both models and is the most accepted. According to this model, a hydrophobic collapse is needed for the formation and stabilization of partially folded secondary and tertiary structures, prompting the formation of a set of compacted conformers known as the "molten globule", and finally, allowing a conformational refinement to acquire the native state (Fersht, 1995; Itzhaki et al., 1995).

From a physical point of view, protein folding is governed by thermodynamic and kinetic laws, being a thermodynamically favoured process promoted by the establishment of intramolecular interactions. These contacts, mainly hydrogen bonds, and hydrophobic or electrostatic interactions, trigger the acquisition of the native structure through a decrease in the free energy of the system.

This thermodynamic protein stability is represented by the difference in Gibbs free energy (ΔG) between the folded (native state) and unfolded conformations, and is determined by two factors according to the following equation:

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

Where ΔH and ΔS denote the enthalpic and entropic contributions, respectively, and T denotes the absolute temperature (Kelvin) (Shirdel and Khalifeh, 2019).

The favourable enthalpic contribution implies the formation of stabilizing native interactions such as electrostatic and hydrophobic interactions, while the entropic contribution of protein folding is generally unfavourable due to the gain of order associated with the acquisition of a folded conformation. Nevertheless, the hydrophobic effect counterbalances this entropic penalty promoting a disorder in the surrounding water molecules and leading to a moderate increment of entropy upon folding.
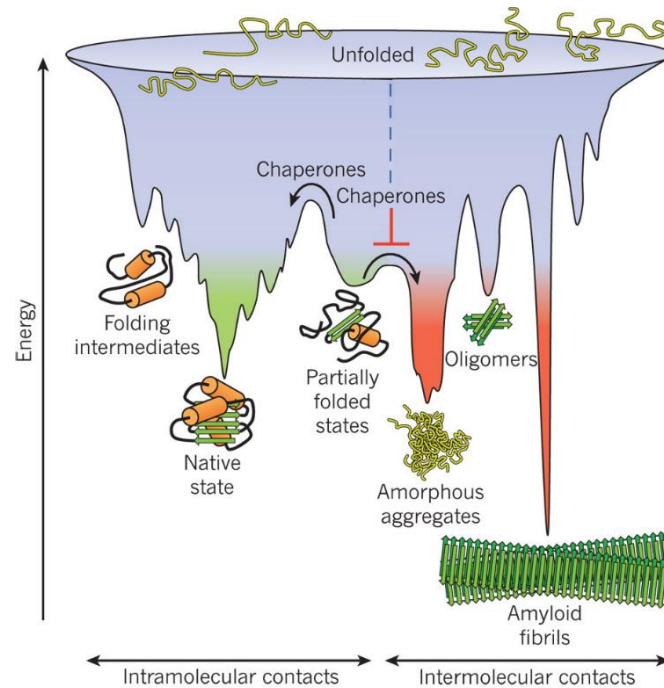
During the protein folding process, there is a huge diversity of possible conformations accessible for each polypeptide, which constitutes its conformational space. This space is usually illustrated by a bidimensional diagram, named as "the energy landscape", where the unfolded conformations populate the highest energetic states while only a few species, native or native-like conformations, occupy the most stable regions associated with minima energy levels (Dobson et al., 1998) (Figure 2). The folding funnel hypothesis describes the protein folding process as a continuous progress from the most unstable conformations (unfolded state) towards the most stable species, the native state (Karplus, 2011). This process implies a small number of possible conformations for short proteins exhibiting a "funnel-like" landscape, leading to a rapid two-state reaction, however, in the landscape of large proteins, these possible conformers dramatically increase, with diverse metastable folding intermediates, introducing significant complexity to the folding process.

2. Protein aggregation

Errors during the protein folding process trigger the formation and accumulation of misfolded metastable conformations. These misfolded states are sustained by hydrophobic interactions and hydrogen bonding similarly to the protein folding process, but in this case, these contacts involve different molecules (Chiti and Dobson, 2017) (Figure 2). These non-native interactions are mainly triggered by the solvent exposure of hydrophobic regions, which are buried when the protein is correctly folded. To counterbalance this anomaly, nature has evolved a set of quality control mechanisms aimed at avoiding the accumulation of misfolded species in the cytosol, such as the chaperone-assisted folding and the ubiquitin-proteasome degradation pathway (Hartl et al., 2011), among others. Failures in such quality control machinery (exacerbated by aging) trigger the accumulation of potentially cytotoxic protein aggregates in the cytoplasm (Hipp et al., 2019).

In order to perform their biological functions, proteins have to be dynamic and fluctuating entities. However, this dynamism might be associated with the appearance of partially and locally unfolded conformations (overpassing the kinetic barrier of protein folding) that can interact with other misfolded species, starting an assembly reaction and leading to the formation of soluble oligomers, or insoluble and extremely stable fibrils. These self-assembled conformations are related to a myriad of human pathologies, ranging from neurodegenerative disorders (e.g., Alzheimer's

and Parkinson's diseases) in which aggregates are deposited in the brain, to systemic amyloidoses involving multiple tissues (Invernizzi et al., 2012).

Figure 2. Hypothetical energy landscape of protein folding and protein aggregation. The protein folding process is a thermodynamically favoured reaction guided by the establishment of intramolecular interactions, leading to the acquisition of a free energy minimum conformation, the native state. However, errors during this process evoke the establishment of intermolecular and aberrant interactions and the subsequent accumulation of extremely stable amyloid fibrils. Reproduced with permission from Hartl et al. 2011.

2.1.   Determinants of protein aggregation

Although proteins related to conformational diseases do not share specific sequential or structural features (there is a huge diversity of structures able to form amyloids, such as all-α, all-ß, α/ß or disordered proteins), all of them adopt an organized structure when the aggregation process occurs (Almeida and Brito, 2020). This structural promiscuity together with the discovery of functional amyloids formed at physiological conditions (Otzen and Riek, 2019) and the aggregation of native-like conformations such as in therapeutic proteins, strongly suggest that the protein aggregation process is not only restricted to pathological conditions, and must be understood as a generic property of proteins. Therefore, protein aggregation is a pivotal and evolutionary-conserved reaction that cannot be completely purged out by evolution (Sanchez de Groot et al., 2012).

Although departing from different primary sequences, the amyloid architecture generally consists of a highly ordered assembly, known as the cross-ß conformation, sustained by the stacking of parallel or antiparallel ß-sheets perpendicularly oriented to the fibril axis (Ke et al., 2020).

This characteristic architecture has been extensively studied using high-resolution structural techniques such as X-ray diffraction, solid-state nuclear magnetic resonance (ssNMR) and cryo-electron microscopy (cryoEM) (Gallardo et al., 2020; Loquet et al., 2018; Riek and Eisenberg, 2016), elucidating that the side chains of the residues in the ß-strands are accommodated into a compact, interdigitated and dry interface, called "the steric zipper". Generally, these amyloid fibrils possess a diameter in the range of 5-20 nm and a length of several μm. Moreover, despite the common preservation of this cross-ß topology, amyloid fibrils derived from the same protein might display different polymorphic structures associated with distinct patterns of neuropathology (Tycko, 2015). In this way, a recent paper published in *Nature* reports different disease-related amyloid structures of Tau protein, providing with a clinically useful classification of tauopathies depending on their amyloid fold (Shi et al., 2021). These polymorphisms, fundamentally identified by cryoEM, defy our idea of amyloids as a monolithic structural arrangement, and pave the way for an exciting analysis of the structural attributes of these supramolecular assemblies and their relevance in different disease-associated phenotypes.

The cross-ß signature of amyloids can specifically bind several amyloid dyes such as Thioflavin-T (Th-T) and Congo Red (CR), whose spectral properties change upon binding to amyloid-like conformations, and the biophysical properties of amyloids can be unveiled monitoring their conformational traits using Fourier transform infrared spectroscopy (FTIR) and circular dichroism (CD). These spectroscopic techniques can be complemented with the visualization of amyloid fibrils using transmission electron microscopy (TEM) and atomic force microscopy (AFM), providing with a plethora of feasible and cost-effective strategies for the study of amyloids (Sabaté and Ventura, 2013).

2.1.1. Intrinsic determinants of protein aggregation

Protein aggregation process is strongly dependent on the sequential determinants encoded in the primary sequence. Mutagenesis experiments evidenced the importance of the physicochemical properties of amino acid's side chains in the aggregation propensity, remarking that this process is mainly dictated by the

hydrophobicity and the ß-sheet propensity of given residues (Chiti et al., 2002, 2003). In this way, the substitution of polar residues by non-polar and hydrophobic amino acids (aliphatic (Ile, Leu, Val) and aromatic (Phe, Tyr, Trp)) results in an increment in the overall aggregation propensity, while charged (Arg, Lys, Asp and Glu) and ß-breaker (Pro and Gly) residues work against this tendency. The net charge of the polypeptide plays a pivotal role, with electrostatic repulsion forces between molecules being anti-aggregational (Lawrence et al., 2007).

2.1.2. Environmental factors impacting protein aggregation

The aggregation tendency of a protein is not only dictated by the aforementioned intrinsic determinants, but also influenced by the surrounding milieu. In this sense, extrinsic factors such as the ionic strength, the pH, the temperature and the protein concentration are crucial in the modulation of the aggregation tendency of a protein (DuBay et al., 2004; Morel et al., 2010; Wang et al., 2010).

The pH of the solution influences the protonation states of the carboxyl and amide groups within amino acid side chains, modifying the net charge and the polarity of the polypeptide. When the pH reaches the isoelectric point (pI) of the polypeptide, the net charge of the protein is 0, precluding electrostatic repulsions and favoring the establishment of non-specific interactions between different monomers (Tedeschi et al., 2017). Similarly, the ionic strength can impact the electrostatic and repulsive forces between molecules, favouring or disfavouring self-association processes.

Since the conformational stability of a protein is affected by the temperature, this parameter can destabilize the native state, promoting the appearance of transient partially unfolded conformations that expose previously buried hydrophobic regions to the solvent, triggering the formation of intermolecular contacts. These interactions between molecules can also potentially arise from a local increment in protein concentration (Ciryam et al., 2015; Tartaglia et al., 2007b; Vecchi et al., 2020).

2.2. Aggregation-Prone Regions (APRs) and *gatekeeper* residues

As previously commented, protein aggregation is strongly influenced by the physicochemical properties of amino acids in the polypeptide chain. However, it has been demonstrated that not all the protein sequence contributes equally to the aggregation process. There are short stretches (10-15 residues in length) characterized by an enrichment in hydrophobic residues able to nucleate the

aggregation reaction (Teng and Eisenberg, 2009; Ventura et al., 2004). These aggregation-prone stretches, commonly known as Aggregation-Prone Regions (APRs), are usually located at the hydrophobic core in the native conformation (Ventura et al., 2002), stabilizing and facilitating the acquisition of the tertiary structure, and in the interfaces between protein subunits, sustaining the quaternary structure indispensable for the correct function (Castillo and Ventura, 2009). In both cases, nature has evolved protective strategies to prevent these APRs from establishing aberrant interactions, however, because proteins are not static molecules in solution, they might undergo aggregation reactions by the ephemeral exposure of APRs to the solvent (Garcia-Pardo et al., 2014).

In order to diminish the aggregation tendency of APRs, nature has accommodated charged and ß-breaker residues flanking these regions (Beerten et al., 2012; De Baets et al., 2014). The function of these *gatekeeper* residues is to complicate aberrant interactions between APRs, providing electrostatic repulsions or disfavouring ß-sheet conformations. Although the presence of these *gatekeepers* negatively impacts the folding kinetics and is associated with an overall destabilizing effect (each *gatekeeper* residue can decrease the thermodynamic stability of the folded state by 0.8 kcal/mol) (De Baets et al., 2014), its effect is more prominent disfavouring the kinetics of protein aggregation (Houben et al., 2020), suggesting their emergence as the result of a selective pressure against aggregation. Intriguingly, there are important differences between negatively and positively charged *gatekeeper* residues. Positive *gatekeepers*, such as lysine or arginine, are more dehydrated and harbour more hydrophobic side chains than their negatively-charged counterparts (aspartic and glutamic acid), displaying a lower aggregation-breaking efficiency. However, these positive residues can be better incorporated to the globular structure of the protein than the negative ones, whose side chains (shorter and less hydrophobic) are not so well accommodated in the native structure (Houben et al., 2021). This fact is extremely important for the redesign of protein solubility, because a destabilizing effect, caused by the incorporation of wrong *gatekeepers* in the structure, could act in the opposite direction, evoking the partial exposure of hidden APRs and accelerating the aggregation process.

2.3. Mechanism of protein aggregation and amyloid fibril formation

In the last years, different aggregation formation models have been proposed to illustrate the kinetics of protein aggregation. These models are the "Templated

Assembly" developed by Griffith (Griffith, 1967) or the "Monomer directed Conversion" postulated by Prusiner (Prusiner, 1982), both of them based on the analyses of the prion protein' aggregation kinetics. However, the model that best explains the kinetics of protein aggregation for the majority of proteins, is the "Nucleation-dependent Polymerization" model (Jarrett and Lansbury, 1993) that divides this process in 3 different phases (Figure 3):

The first stage is the **nucleation process**, a thermodynamically unfavourable reaction where protein monomers start to interact and associate, leading to the formation of oligomers or prefibrillar species. These protein oligomers are considered as the nucleation elements from which the propagation reaction begins, and represent one of the most pathogenic conformations in amyloid diseases (Kreiser et al., 2020). This first step is deeply affected by protein concentration since elevated concentrations imply a higher probability of monomer-monomer interactions; and can be accelerated by the addition of pre-formed aggregated species by a *seeding* process. Interestingly, experiments aimed at analysing *self-seeding* (seeding using the same APR) and *cross-seeding* (seeding using different APR) reactions, suggested a specificity in the assembly reaction, with an acceleration of the aggregation process when the same APR was used (Sabaté et al., 2010).

Later on, the second stage, known as **elongation**, is a rapid and thermodynamically favoured process consisting in the polymerization and growth of the amyloid fibril through monomer addition.

Finally, the incorporation of molecules to the fibril is associated with a decrease in the available monomers in solution, decelerating the elongation process and leading to the third stage, the **steady or saturation phase**. At this point, the assembly/disassembly reactions are in a chemical equilibrium because of the recycling of monomers in the mature fibril. In addition, different amyloid fibrils can laterally interact between them, forming a mature amyloid fiber composed of several protofilaments.

This process can be easily monitored following the increments in Th-T fluorescence and light scattering. In the first stage, the soluble and monomeric species are not able to bind the Th-T dye and do not scatter the incident light, whereas the increase of amyloid structures in solution directly correlates with an increment of both signals (Figure 3).
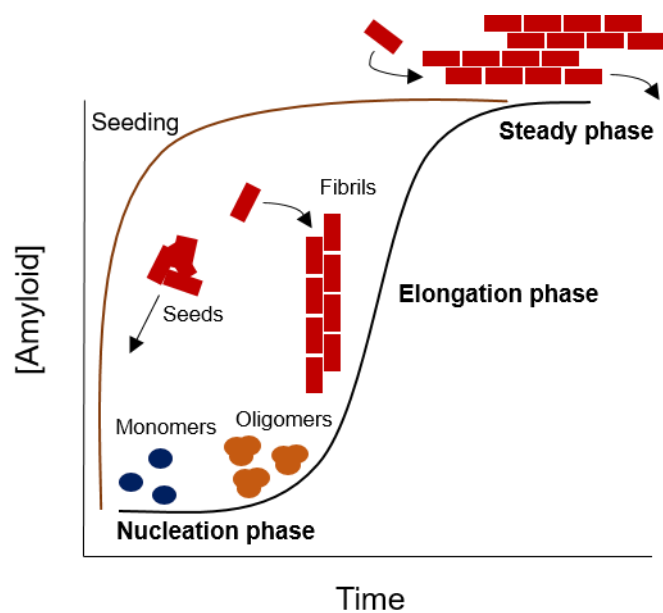
Figure 3. Protein aggregation formation according to the nucleation-dependent polymerization model. The assembly of amyloid fibrils comprises three stages: The concentration-dependent nucleation phase related to the formation of oligomeric species, the elongation phase characterised by the formation of amyloid fibrils after monomer addition, and the steady or saturation phase consisting in a dynamic equilibrium between the assembled and disassembled states. The seeding process, promoted by the incorporation of mature aggregated species into the first steps of the reaction, shortens the nucleation phase resulting in a logarithmic curve (red line).

2.4.   Protein aggregation in therapeutic proteins

Protein therapeutics have emerged as promising life-saving drugs for human diseases due to their excellent specificity and low toxicity. The "therapeutic protein" term covers a wide range of biological molecules such as hormones, growth factors, interferons, recombinant vaccines, replacement enzymes, gene therapy vectors, monoclonal antibodies (mAbs) and next-generation antibody formats such as bispecific antibodies (BsAbs), antibody-drug conjugates (ADCs), single-chain variable fragments (scFvs) and single-domain antibodies (sdAbs) (Khongorzul et al., 2020; Narayanan et al., 2021a; Wang et al., 2019a).

Antibody-based therapies are perhaps the most successful approach to treat pathologically diverse human diseases such as cancer and autoimmune disorders. Specifically, mAbs are the most lucrative single product class, constituting almost the 50% of therapeutic proteins approved by the U.S. Food and Drug Administration (FDA), and doubling approvals in the last twenty years (Figure 4). Indeed, mAbs represent eight of the top ten products by sales in 2017, and the total sales of these molecular therapies reached $123 billion in 2017 (Walsh, 2018).
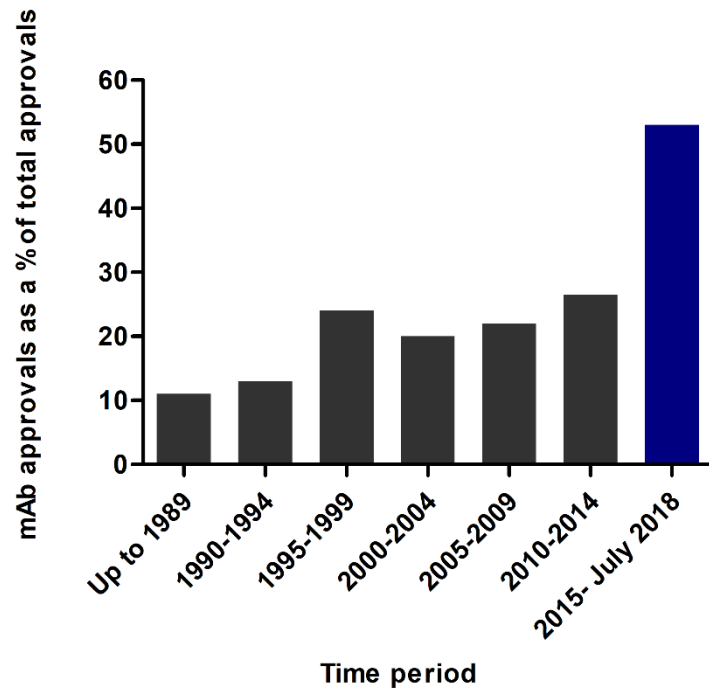
Figure 4. mAb approvals for the first time in percentage of total approvals. The percentage of approvals in the period 2015-July 2018 (in blue color) is two times higher than the one for the previous twenty years. The tendency clearly indicates an increment in the approvals of these therapeutic proteins in the last years. Adapted from Walsh, 2018.

Despite the increasing success of antibody-based therapies, their commercialization is still constrained by the difficulties associated with the production and manufacturing processes. For instance, in most cases, these therapeutic proteins present a very large size (around 150 kDa for mAbs) and complexity (post-translational modifications), making unaffordable their chemical synthesis. A more feasible alternative to produce such proteins is their recombinant expression using microbial and eukaryotic cell factories such as *Escherichia coli*, *Saccharomyces cerevisiae* and Chinese hamster ovary (CHO) cells, being the last ones, the preferred options for the insertion of post-translational modifications.

Moreover, there are challenging barriers regarding the production and manufacturing of these drugs, and in the end, impacting their commercialization (Roberts, 2014). Among the different degradation reactions that therapeutic proteins might suffer during their production, aggregation is the most common, being strongly associated with the extremely high protein concentrations required for most therapies (above 100 mg/mL) (Wolf Pérez et al., 2022). Protein aggregation is not only associated with the production stages, but also with the storage and

administration, because these drugs must be stable and active for long periods of time for which they have not been evolutionary adapted. This harmful process is also related to a diminished protein activity and undesired immunological responses, having regulatory implications (Wolf Pérez et al., 2022). Consequently, the FDA dictates specific and rigorous recommendations for the manufacturing and delivery of such protein drugs, compelling pharmaceutical and biotechnological companies to spend substantial efforts to prevent the eventual aggregation of their products, using trial-and-error assays to screen different buffers, salts, or additives that are time-consuming and marginally successful.

An emerging concept in the field of biotherapeutics is the developability of a product. The developability concept consists of a series of properties that ensure the correct stability and safety of a biological drug during production and formulation processes. In other words, developability is the combination of factors aimed at guarantying the successful development of a therapeutic protein as a safe, active and stable drug (Wolf Pérez et al., 2022). In this way, significant efforts have been dedicated to improve the developability of antibodies and economize the development of these highly demanded molecules, selecting the most idoneous candidates according to a panel of sophisticated methods such as AI or microfluidic technology, among others (Kopp and Arosio, 2018; Narayanan et al., 2021b).

Although mAbs and their relatives are pivotal for cancer research and therapies, these molecules display significant limitations associated with their large size (Figure 5), hampering their tumor accessibility (Thurber et al., 2008), and with their short lifetime. Actually, there is a trend in the pharmaceutical industry towards the use and development of small antibody fragments with enhanced tumor penetration properties and easy-to-produce using microbial cell factories. These small proteins include a vast diversity of antibody formats as the scFvs and the sdAbs (the human single-domain antibodies (the variable heavy chain (V_H) or the variable light chain (V_L)) (Figure 5). However, these small and functional fragments tend to show poor biophysical properties, displaying a high aggregation propensity caused by the lack of the interdomain stabilization elements, present in their larger counterparts (Holliger and Hudson, 2005). In order to improve their properties, different strategies have been devoted to screening mutants with enhanced aggregation resistance. One of these approaches is phage display, a high-throughput method successfully applied by Christ and co-workers to select aggregation-resistant

variants of V_H fragments derived from human germlines (Dudgeon et al., 2012). However, 6 mutations that decrease the aggregation propensity of these antibodies were located at the Complementarity-Determining Regions (CDRs) which are indispensable for the correct function of the antibody. This fact is expected since CDRs are sticky antigen-recognition sites, but these mutations can harm the affinity and selectivity for the antigen. A feasible alternative are nanobodies, small antibodies derived from camelids that show better solubility profiles than human sdAbs, however, they are not naturally expressed in humans and can evoke undesired immunogenic responses after administration. Approaches aimed at "humanizing" such molecules based on the replacement of surface residues by human sequences are strongly associated with a decrease in solubility (Yang and Shah, 2020). For that reason, there is a notable interest in obtaining highly soluble version of human sdAbs.

Rational redesign has emerged as a way to avoid the insertion of mutations that can jeopardize the antibody's functionality. This approach requires knowledge about the structure of the target protein and the role of specific residues in its correct folding and function, and consists in the introduction of specific point mutations aimed to improve one or more of these properties. The *in silico* rational design of new antibodies with improved characteristics and tailored functionalities is emerging as a promising strategy to diminish the efforts and time devoted to antibody discovery (Sormanni et al., 2018).

Since protein aggregation is one of the main drawbacks hampering the development of novel therapeutic proteins, and is associated with different human pathologies, substantial efforts have been addressed to develop computational tools aimed at forecasting the aggregation tendency of a given polypeptide. The use of such predictors can allow to improve the biophysical properties of protein-based drugs and to elucidate the molecular determinants underlying human amyloidosis (Santos et al., 2020).
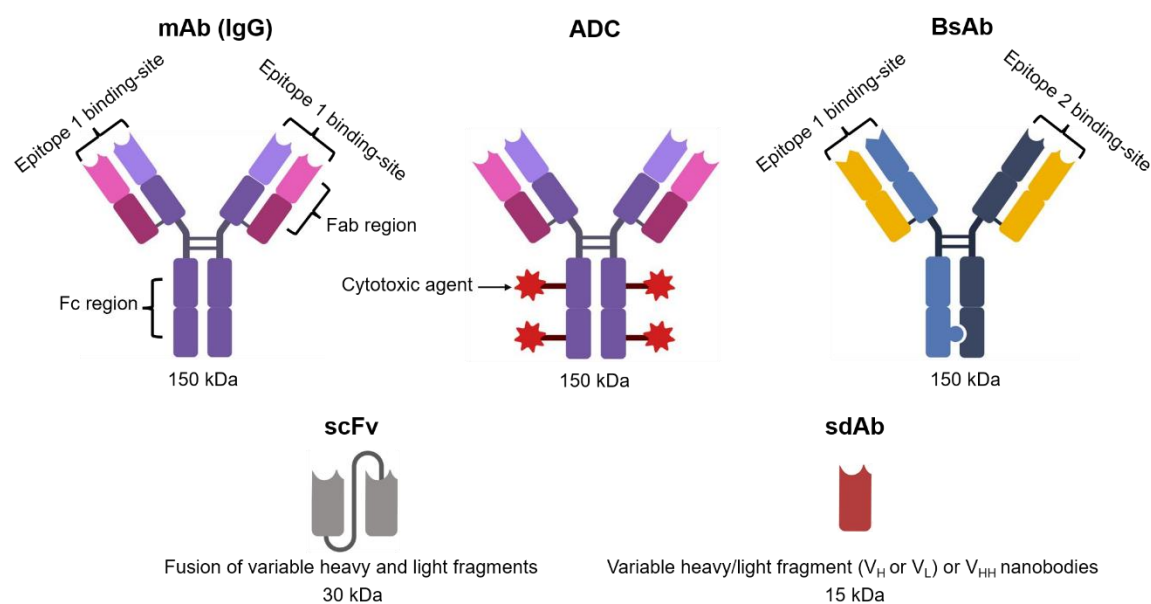
Figure 5. Schematic representation of different antibody formats developed for biomedical purposes. The main hallmarks of each format together with their structure and approximate size (in kDa) are shown. It is important to note that other different formats as bispecific single-chain antibodies are currently available (Wang et al., 2019a), but in this figure we sought to briefly provide an overview of the most representative examples. Created with Biorender.com.

2.5. Prediction of protein aggregation

The exhaustive study of protein aggregation during the last 20 years, and the identification of intrinsic factors encoded in the primary sequence influencing the aggregation tendency of a given polypeptide, have fuelled the development of computational approaches aimed at deciphering the aggregation propensity of proteins. Among these factors, APRs containing a high hydrophobicity, a low net charge, and an enrichment in residues adopting ß-sheet conformations have been the main target of such predictors (Navarro and Ventura, 2019; Santos et al., 2020).

The development of these numerous *in silico* tools (to date, more than 30 algorithms have been released) has been essential for a better understanding of the role of this process in nature (e.g., proteome-wide analyses has unveiled the relation between protein function and aggregation) and for assisting the redesign of protein solubility in therapeutic proteins (Ganesan et al., 2016; Sormanni et al., 2015; van der Kant et al., 2017).

In this section, we will delve into the *state-of-the-art* aggregation predictor toolbox, distinguishing programs according to the different rationale that they employ for their predictions and providing useful information for their use.

2.5.1. Sequence-based aggregation predictors

The first generation of such computational tools are known as *sequence-based* aggregation predictors and attempt to identify linear APRs across the primary sequence. These predictors are potentially useful for mapping the aggregation tendency of: (I) IDPs lacking of a defined 3D structure, (II) nascent polypeptides which are not already folded and still expose their linear APRs, essential for the hydrophobic collapse, and (III) destabilized globular proteins than can transiently expose their buried APRs to the solvent.

To date, more than 20 *sequence-based* predictors have been developed, weighing different parameters. The first subclass of predictors are the *phenomenological*, whose predictions stem from experimental data. Two notable examples of these *in silico* tools are **AGGRESCAN** (Conchillo-Solé et al., 2007) and **Zyggregator** (Tartaglia and Vendruscolo, 2008).

AGGRESCAN was the first predictor based on a scale derived from *in vivo* experimental data. The AGGRESCAN scale was defined measuring the green fluorescence of single mutation variants of the amyloid peptide (Aß42) fused to GFP in the *E. coli* cytoplasm, establishing a direct correlation between solubility and intrinsic fluorescence. This smart strategy allowed the ponderation of the contribution of each single mutation for the solubility of the protein, resulting in an experimentally derived aggregation scale. AGGRESCAN predicts the protein aggregation propensity by calculating the average aggregation propensity for each residue of the primary sequence, and APRs are defined as a stretch of five or more consecutive residues harbouring an aggregation propensity score higher than the threshold (average value of the scale).

Zyggregator implements in its predictions the scale developed by Pawar et al., accounting for hydrophobicity, secondary structure tendency and net charge (Pawar et al., 2005). This algorithm also considers the effect of *gatekeeper* residues along the sequence, taking into account the local structural stability of the target protein implemented in the CamP method (Tartaglia et al., 2007a), allowing the detection of putative APRs in folded proteins.

The second subclass of these predictors forecast the aggregation propensity according to the *theoretical* properties associated with protein aggregation. One example of these algorithms is **TANGO** (Fernandez-Escamilla et al., 2004), an

algorithm aimed at identifying APRs considering the physicochemical properties and energetic contributions associated with the β-sheet conformation. To do that, TANGO evaluates the propensity of each amino acid to form different structural elements such as α-helix and α-helical aggregation, β-sheet and β-sheet aggregation, including the impact of different environmental factors (e.g., pH, temperature and ionic strength) on its predictions. Other widely used *theoretical* predictors are **Prediction of Amyloid Structure Aggregation 2.0 (PASTA 2.0)** (Walsh et al., 2014), **FoldAmyloid** (Garbuzynskiy et al., 2010) and **Waltz** (Maurer-Stroh et al., 2010). Briefly, PASTA 2.0 predictions rely on the assumption that amyloid β-sheets correspond to pairings of parallel or antiparallel β-sheets. The estimation of β-sheet-based intermolecular contacts relies on experimentally-resolved β-sheet structures coupled with a statistical energy function able to identify fibril forming segments. FoldAmyloid bases its predictions on the "average packing density" concept that assumes high packing densities and H-bonding propensity for amyloid structures; and Waltz exploits a position-specific scoring matrix derived from the structural analysis of 48 fibril-forming hexapeptides for predicting the aggregation propensity.

Furthermore, several *machine learning* methods such as **FISH Amyloid** (Gasior and Kotulska, 2014), **APPNN** (Família et al., 2015) and **NetCSSP** (Kim et al., 2009) employ neural networks to identify sequential patterns associated with protein aggregation, obtaining similar or even better results than the previously commented algorithms.

Finally, the last subset of *sequence-based* predictors takes profit of the aforementioned tools to create *consensus* predictions. These predictors minimize the biases of each singular algorithm, improving the robustness of the output. Two notable examples are **MetAmyl** (Emily et al., 2013) and **AMYLPRED2** (Tsolis et al., 2013).

2.5.2. 3D structure-based aggregation predictors

Although sequence-based algorithms are appropriate for the prediction of protein aggregation during protein synthesis (before the acquisition of the 3D structure) or in the case of IDPs (devoid of a 3D structure), they are clearly blind to APRs placed in a folded conformation. APRs might play pivotal roles to promote the correct protein folding, stabilizing the hydrophobic core, being thus, solvent-inaccessible. Therefore, these buried APRs make a neglectable contribution to the aggregation propensity of a folded protein, even if sequence-based algorithms would consider

them as important hot spots of aggregation, resulting in a significant overprediction of aggregation relevant stretches.

In order to overcome the limitations associated with *sequence-based* prediction tools, substantial efforts have been addressed to develop improved algorithms aimed at forecasting the aggregation tendency of a folded protein. These *structure-based* aggregation predictors can detect structural clusters of hydrophobic residues that are not contiguous in sequence, but that, when the protein folds, rearrange very close in space. These aggregation-prone regions formed by the clustering of sequentially distant residues are named Structural Aggregation Propensity Regions (STAPs), and *sequence-based* predictors are blind to them. S*tructure-based* software are valuable tools for the identification of solvent-exposed STAPs in therapeutic proteins, holding potential for biomedical and biotechnological applications.

In the present thesis, we will describe four *structure-based* aggregation predictors whose performance has been already experimentally validated (Table 1).

The first predictor intended to foresee the aggregation tendency of globular proteins was **Spatial Aggregation Propensity (SAP)** (Chennamsetty et al., 2009). Initially, SAP was conceived for the creation of aggregation-resistant antibodies. SAP exploits Molecular Dynamics (MD) to simulate the structural fluctuations in the folded state and assigns an aggregation score (SAP score) to each amino acid, calculating the mean SAP value of a residue considering the side chains of residues located at 5 Å. The spatial contribution for each amino acid is calculated considering the solvent accessible area of the atoms within the sphere and the corresponding hydrophobicity of each residue. To do that, SAP applies the *Black and Mould* scale (Black and Mould, 1991), dismissing the contribution of solvent-buried residues, and mapping the SAP value onto the protein surface according to a colour code. A posterior version of SAP, the **Developability Index (DI)** algorithm (Lauer et al., 2012), exploits the SAP pipeline to accurately forecast the aggregation tendency of mAbs considering both their tertiary structure and net charge.

The SAP method has been employed for the study of APRs within the different subclasses of antibodies, allowing to redesign their aggregation propensity by rational mutagenesis. This strategy is considered the first attempt to modulate the aggregation propensity of antibodies using *in silico* tools (Chennamsetty et al., 2009). Furthermore, SAP has been successfully applied for: (I) the prediction and identification of protein binding sites in two model proteins, the epidermal growth

factor receptor (EGFR) and an immunoglobulin G1 (IgG1), (II) the understanding of the aggregation propensity of light-exposed antibodies, and (III) the modelling of the aggregation tendency and toxicity of Aβ42 variants (Bommana et al., 2018; Chennamsetty et al., 2011; Tiwari and Kepp, 2015). Although this algorithm is appropriate to redesign the protein solubility, its availability is restricted to a commercial license.

Other experimentally-validated tool is **CamSol** (Sormanni et al., 2015), intended to forecast the aggregation propensity of folded proteins considering the contribution of additional physicochemical properties, relative to SAP. CamSol predictions are based on the Zyggregator sequence-based algorithm, aimed at predicting protein solubility. Although solubility and aggregation are similar concepts, they differ from a physical point of view, reflecting different properties of the energy landscape. While solubility is a thermodynamic parameter associated with the free energy difference between the native and the aggregated state, the aggregation propensity is kinetically connected to the free energy barrier between these two states. Briefly, the CamSol pipeline consists of two steps: First, each residue of the primary sequence is scored according to its hydrophobicity, charge and secondary structure propensity, and a correction factor is applied to the resulting profile, smoothing the result in a 7-residue segment context that considers the influence of sequentially nearby residues. Then, this solubility profile is structurally corrected, analysing the contribution of the initial 7-residue window and the residues located at 8 Å radius sphere from the central amino acid, computing the impact on protein solubility after an evaluation of the relative solvent exposure. Thus, the structurally-corrected solubility score for each amino acid is derived from the ponderation of its intrinsic score in the linear profile and the structural context contribution, together with the solvent accessibility. The structure is finally coloured according to the resulting profile in order to assist the user in the identification of STAPs.

Furthermore, the CamSol method includes a semi-automated mode able to detect and mutate the strongest STAPs, providing the user with the predicted most soluble variants. However, despite CamSol avoids mutating important regions in terms of structure and function, it does not predict the thermodynamic effect of such mutations in the overall structure upon mutation, being blind to undesired destabilizing changes. This computational tool has been successfully applied for (I) the computationally-assisted redesign of aggregation-resistant anti-Aβ42 antibodies,

(II) the identification of genetic variants of leptin with reduced solubility, (III) the analysis of the relationship between aggregation propensity, stability and conformational flexibility of ß2-microglobulin and its W60G variant, and (IV) the selection of mAb variants with enhanced solubility, among others (Camilloni et al., 2016; Haglund et al., 2018; Shan et al., 2018; Sormanni et al., 2015). This predictor is implemented as a free webserver for academic purposes.

Another *structure-based* tool derived from a *sequence-based* predictor is **SolubiS** (De Baets et al., 2015; Van Durme et al., 2016). The SolubiS method forecasts the aggregation propensity of folded proteins identifying linear APRs using TANGO, evaluating the impact on protein stability of each APR using FoldX, and allowing the insertion of solubilizing and non-destabilizing mutations (van der Kant et al., 2019). FoldX is an empirical forcefield able to analyse the contribution of different energetic terms on protein stability. Briefly, these energetic terms are the solvation energies for polar and non-polar amino acids, the electrostatic and Van der Waals interactions, the hydrogen bonds, the entropic cost of fixing a side chain in a particular conformation, and the steric clashes between atoms in the structure (Schymkowitz et al., 2005). The SolubiS output consists of a two-dimensional "Stretch-plot" where predicted APRs are plotted as a single point accounting for the intrinsic aggregation propensity and the stability contributions. SolubiS assists the rational redesign of protein aggregation propensity above a desired stability threshold, excluding potential destabilizing mutations that would negatively impact the protein solubility. The algorithm also offers an automatic search of solubilizing and stability neutral mutations based on the *gatekeeper* strategy.

The SolubiS method has been effective in the redesign of biotechnologically relevant proteins such as the human α-galactosidase, widely used for the treatment of Fabry disease, and the protective antigen of *Bacillus anthracis* (Ganesan et al., 2016). Moreover, this software has successfully assisted the rational redesign of aggregation-resistant mAbs without altering their antigen binding capacity and stability (van der Kant et al., 2017). However, since SolubiS is trained to identify linear APRs, it is absolutely blind for the identification of STAPs formed by residues that are scattered in the sequence. This computational tool is freely available as a webserver and as a plugin for the YASARA software.

Last but not least, our group has developed **AGGRESCAN 3D (A3D)** (Zambrano et al., 2015), an experimentally-validated *structure-based* aggregation predictor

derived from AGGRESCAN. Similar to CamSol, A3D foresees the structural aggregation propensity of a given protein by correcting the intrinsic aggregation prediction for each amino acid by the solvent exposure and the structural environment. This structural contribution derives from the sum of the corrected aggregation propensity for each residue within a 10 Å radius sphere from the Cα carbon of the particular amino acid, although A3D allows to reduce this radius sphere to 5 Å.

Using a PDB structure as an input (derived from experimental data or computationally modelled), the A3D pipeline starts with an energy minimization of the structure using the FoldX forcefield in order to remove the unfavourable conformations (e.g., incorrect torsion angles, steric hindrance between side chains and non-optimized rotamer orientations). Then, the repaired PDB is analysed by A3D and the aggregation score for each residue is graphically represented, implying positive values an aggregation-prone tendency while negative values entail a protective effect against aggregation.

Similar to SAP, A3D offers a *dynamic mode* to calculate the impact of molecular fluctuations in the protein aggregation propensity using MD simulations. These MD simulations are performed using CABS-flex, a high-resolution coarse-grained molecular modelling algorithm able to reproduce the dynamic fluctuations of the near-native ensembles (Jamroz et al., 2013). Furthermore, an improved version of this predictor that arises in part from the experimental results obtained in the present thesis, **A3D 2.0** (Kuriata et al., 2019b), incorporates protein stability calculations upon mutation (using the FoldX forcefield) and a "automated mutations" mode aimed at identifying the most solubilizing and non-destabilizing variants, allowing the redesign of protein solubility without impacting protein stability. This new released version affords to run the *dynamic mode* for proteins larger than 400 residues and multichain molecules, and has been recently used by Ventura and colleagues in combination with AlphaFold to forecast the aggregation propensity of the complete human proteome (unpublished data). These improvements in A3D can be extremely useful for researchers not familiarized with protein design, and the software is freely accessible for academic and non-profit users. Interestingly enough, it is also implemented as a standalone version in Python, offering a suitable framework for merging it with other available computational tools (Kuriata et al., 2019a).

As previously mentioned, the experimental validation of the ability of A3D to assist the redesign of protein solubility without altering protein stability is part of the present doctoral thesis (Chapter 2) (Gil-Garcia et al., 2018).

Table 1. Hallmarks of experimentally-validated 3D structure-based aggregation predictors.

	SAP	CamSol	SolubiS	A3D 2.0
Score derived from experimental data	No	Yes	No	Yes
Dynamic mode	Yes	No	No	Yes
Auto-mutation mode	No	Yes	Yes	Yes
Assessment of impact on protein stability	No	No	Yes	Yes
Corrected by atom distance	Yes	Yes	No	Yes
Corrected by solvent exposure	Yes	Yes	No	Yes
Experimentally validated	Yes	Yes	Yes	Yes
Availability	Under license	Free webserver	Free webserver or YASARA plugin	Free webserver and standalone version

3. Protein-based self-assembled functional nanomaterials

Nature has evolved functional nanostructures exploiting the physicochemical properties of peptides and proteins. The use of proteins as building blocks to create structured biomaterials is conserved across the different kingdoms of life. In this way, viruses are assembled via protein-protein interactions, creating nanocages able to encapsulate the genetic material (Perlmutter and Hagan, 2015), the bacterial biofilm of *Escherichia, Salmonella* and *Pseudomonas* contains significant amounts of amyloid fibrils in order to increase its rigidity and strength, improving cell adhesion and host colonization (Van Gerven et al., 2018); and, in the eukaryotic cell, actin fibrils are essential to perform different functions such as cell polarity or motility (Pollard and Borisy, 2003). Moreover, the Balbiani body in oocytes, a compartment involved in the recruitment of different organelles (e.g., mitochondria and endoplasmic reticulum) during the early phases of oocyte maturation, is sustained by amyloid-like interactions (Boke et al., 2016; Boke and Mitchison, 2017). Other functional amyloids are prions, a type of self-perpetuating proteins able to switch between a soluble state and a transmissible amyloid conformation, being

involved in heritable phenotypic adaptations (Batlle et al., 2017). Despite prion proteins have been extensively identified in yeast, the development of a set of prion predictors has fuelled the discovery of functional prion-like proteins in different and unrelated organisms (Gil-Garcia et al., 2021) such as the ones involved in the flowering process in plants (Chakrabortee et al., 2016; Jung et al., 2020).

Considering more complex organisms, the silk secreted by spiders and silkworms is composed of self-assembled amyloid fibrils formed through a natural spinning process (Römer and Scheibel, 2008). In *Drosophila*, an aggregated conformation of Orb2A protein controls the long-lasting memory (Gill et al., 2017), and inside the human body; (I) a network of collagen fibrils shapes the mechanical properties of biological tissues (Bielajew et al., 2020) and (II) an essential biological process as blood coagulation is promoted by the polymerization of fibrin polypeptides (Weisel and Litvinov, 2013). Strikingly, the secretory granules of the human endocrine system act as functional amyloids storing peptide hormones until their release (Maji et al., 2009).

The use of protein-based self-assembled materials in nature has opened the door for the creation of versatile and biocompatible nanostructures with tunable properties and tailored functionalities. In comparison with synthetic nanostructures (e.g., polymeric nanoparticles and liposomes), these proteinaceous building blocks are easy to produce, biocompatible, biodegradable and can be tuned by simple genetic redesign (Jain et al., 2018).

3.1. Amyloid-based self-assembled nanomaterials

Amyloid fibrils have attracted increasing interest due to their promising potential as biomaterials for biomedical and biotechnological purposes. As previously commented, the amyloid architecture consists of intermolecular ß-sheets oriented perpendicularly to the fibril axis. These amyloid fibrils possess an optimal strength, stiffness and toughness as a consequence of their hierarchical organisation, harbouring idoneous mechanical properties for their use as biomaterials. Furthermore, amyloids are tunable structures in which shape and size can be modified playing with the environmental conditions (ionic strength, temperature, pH) (Wei et al., 2017). Therefore, amyloid-based nanomaterials are a promising platform with potential applications in biomedicine, biotechnology and environmental sciences.

One of the most important applications of amyloids is their use as scaffolds for cell growth and proliferation. For instance, different amyloid fibrils formed by lysozyme or by amyloidogenic peptides obtained from the α1 chain of mouse laminin have been employed as biocompatible networks for promoting cell adhesion and spreading (Kasai et al., 2004; Reynolds et al., 2014). In addition, amyloid-based assemblies have been exploited to create biosensors. In an elegant application, Men et al., created amyloid nanowires composed of two fusion proteins consisting of Sup35 prion domain as the assembling unit and two unrelated functional moieties, the protein G for antibody binding and the enzyme methyl parathion hydrolase. Doping the solution with different ratios of both fusion proteins, the authors created multifunctional nanowires with a high ration of enzyme to protein G, displaying a 100-fold higher sensitivity for the detection of F1 antigen from *Yersinia pestis* than classical HRP-conjugated antibody-based ELISA (Men et al., 2009). In subsequent studies, this group achieved the decoration of amyloid nanowires by exploiting the selective biotin-avidin interaction, allowing the attachment of any desired protein to the amyloid scaffold and skipping the production of different fusion proteins for each intended application (Men et al., 2010).

Amyloid fibrils have also found remarkable applications in environmental science and technology, acting as cheap and environmentally friendly purifying nanostructures. In this way, the Eisenberg lab has developed amyloid fibrils able to catch CO_2 from air (Li et al., 2014).

Our group has devoted significant efforts to the creation of functional amyloid fibrils using two strategies that stem from the characteristic amino acid composition of prion domains, enriched in polar and uncharged residues and harbouring a cryptic amyloid propensity.

The first strategy relies on the use of short-and-minimalist binary peptides able to self-assemble into ß-sheet-enriched nanofibrils with electrocatalytic activities and without any associated-toxicity. The polar nature of these peptides results in a slow aggregation reaction, allowing a kinetic control of it (Díaz-Caballero et al., 2018). Subsequently, Díaz-Caballero et al., successfully created pH-responsive self-assembling peptides, composed of tyrosine and histidine, that can trigger the formation of reversible hydrogels upon pH variations, holding hydrolytic and electrocatalytic activities. Interestingly, these biocompatible HY peptides exemplify

that the self-assembling properties and functionality of an amyloid fibril can be encoded by the same residues (Díaz-Caballero et al., 2021).

The second strategy exploits the "soft amyloid core (SAC)" technology based on the use of cryptic amyloid sequences embedded within prion domains to drive the self-assembly reaction into amyloid-based fibrils (Fernández et al., 2017). These SACs are more polar and longer than classical amyloid peptides, allowing the expression of a fusion protein consisting of these peptides and the desired protein as a soluble entity, while retaining a sufficient aggregation tendency to trigger the self-assembly process *in vitro*. This approach has been exploited for the creation of a wide variety of functional amyloids harbouring different activities such as fluorescence, catalysis and antibody capturing (Wang et al., 2019b, 2021). Moreover, modifying the length of the flexible linker between the SAC peptide and the appended globular domains, Wang et al., have successfully created amyloid oligomers capable of delivering a cytotoxic compound (methotrexate) to cancer cells specifically (Wang et al., 2020).

3.2. Coiled-coil-based self-assembled structures

Although amyloid-based nanomaterials are perhaps the most studied protein-based nanostructures, they have limitations, mostly derived from the associated-toxicity of the amyloid architecture (Díaz-Caballero et al., 2020). Certain natural functional amyloid-based assemblies are formed and work at physiological conditions, but this requires the existence of control mechanisms, developed by nature, to ensure that the self-assembly process is completed and the formation of toxic oligomeric species is avoided (Jackson and Hewitt, 2017). Artificial amyloid-based structures are not under this natural control, and the release of toxic oligomeric species during or after their polymerization cannot be discarded.

In order to overcome these amyloid-associated drawbacks, several research groups have exploited the coiled-coil architecture sustained by α-helices to build up safer protein-based materials (Wu and Collier, 2017).

From a biochemical point of view, coiled-coils present a characteristic structure, exploiting the *knobs-into-holes* strategy, relying in an ordered packing of the residues of the different helices that promotes the interdigitation of the side chains (Crick, 1953). Canonical coiled-coils share a particular amino acid composition, associated with a heptad repeat denoted as *abcdefg*, where *a* and *d* positions tend to be occupied by hydrophobic residues, and the *eg* positions by polar or charged amino

acids; although, other possible periodicities such as hendecad and pentadecad repeats have been described (Lupas and Bassler, 2017).

Coiled-coils are considered an optimal architecture to create protein-based biomaterials since their shape and morphology can be fine-tuned by simple modifications of paired interactions between residues, displaying a negligible associated-toxicity. The easy modification of coiled-coil interfaces by protein design has allowed the creation of precisely defined structures such as nanofibers, nanoparticles, networks and hydrogels, paving the way for the build-up of biomaterials for biomedical and biotechnological applications such as drug delivery or vaccine development (Utterström et al., 2021).

Similar to amyloid-based nanoassemblies, coiled-coil nanoparticles have been successfully exploited as stimulus-responsive drug delivery systems. Due to the pH and/or temperature sensitivity of these structures and their hydrophobic core, coiled-coils have been used as nanocarriers to improve the circulation time in blood of a determined cargo molecule, allowing a selective release upon environmental variations. In this way, Eriksson and co-workers have demonstrated the ability of a coiled-coil domain to transport the chemotherapeutic drug, cisplatin, to tumoral cells, demonstrating elevated cytotoxic effects in comparison with free cisplatin *in vitro* (Eriksson et al., 2009).

However, the most outstanding application of the coiled-coil architecture, in comparison with the amyloid one, is related to its ability to self-assemble into symmetric and well-defined nanocages. The versatility of this assembly has been pivotal to create coiled-coil based vaccines able to incorporate epitopes derived from diverse viruses as the severe acute respiratory syndrome virus (SARS) and the human immunodeficiency virus (HIV), triggering an antigen-specific humoral response without needing co-adjuvants (Pimentel et al., 2009; Wahome et al., 2012). This nanocage is assembled via pentameric and trimeric coiled-coils, displaying a spherical shape, an icosahedral symmetry and a diameter around 30 nm. Interestingly, this strategy provides a modular platform to present the desired antigen in an assembled conformation, increasing the valency of a virus component-free format.

3.3. Supramolecular architectures sustained by self-assembling proteins

The modular self-assembly of protein building blocks allows the creation of a plethora of different supramolecular architectures displaying different properties and opening an avenue for a new era of biodegradable, biocompatible and tailored biomaterials. Here, we will mention the main hallmarks of such structures (Figure 6):

Fibers: The formation of nanofibrils has been widely exploited by nature to create functional nanostructures. Determined proteins can self-assemble into spatially ordered and linear nanostructures, typically stabilized by extended supramolecular α-helical structures (coiled-coil nanofibrils) or β-sheet networks (amyloid fibrils), that share excellent mechanical properties, independently of the scaffolding polypeptide. Among the extensive applications of such structures, they can be exploited as antibody capturing entities or reusable catalysts, depending on the constituent active moiety (Wang and Ventura, 2020). The modular activity of these protein fibers stems from the ability of genetically append specific proteins to the assembling entity, whereas an in-depth understanding of the interactions sustaining these assemblies (α-helix or β-sheet) allow to fine-tune their structural properties by simple modification of the environmental conditions (pH, ionic strength, temperature…) (Shen et al., 2021).

2D-Films or arrays: Nature uses the self-assembling properties of specific polypeptides to create 2D-protein arrays. The most known example is the S-layer, a 2D-protein film frequently found in the surface of bacteria and archaea, and implicated in cell protection, adherence and topology (Zhu et al., 2017). Significant efforts have been devoted to mimicking this natural architecture, creating synthetic 2D-films with tailored properties and functionalities (Liu et al., 2021). Several constructions departing from native proteins have been successfully achieved, however, their geometry and symmetry was strongly limited by the characteristics of the oligomeric state of the native polypeptide (Bai et al., 2016). To overcome this limitation, Di Maio and Baker labs have created *de novo* α-helical bundles with the desired geometry for obtaining customizable, one-component 2D-arrays (Chen et al., 2019). Nevertheless, 2D-arrays formed by two or more components are more attractive. In this regard, the Baker lab has recently achieved an outstanding milestone, creating biologically active binary 2D-arrays that provide with potential advantages in the modulation of the assembly dynamics and in the biomaterial's functionality (Ben-Sasson et al., 2021). These binary arrays can self-assemble at

nanomolar concentrations of each component and harbour defined activities, being able to assemble at the cell surface without suffering endocytic processes, and thus, holding a promising future in the modulation of cell responses and reshaping of living systems. Another biotechnologically-oriented application of 2D-protein films exploits their ability to specifically absorb heavy metal ions from water when decorated with active carbon, resulting in an eco-friendly approach to fight against water contamination (Bolisetty and Mezzenga, 2016).

Protein cages: Inspired by nature, the design of protein cages formed by the self-assembly of different subunits to create drug delivery systems has become a promising therapeutic strategy (Bhaskar and Lim, 2017). The concept of creating protein cages with specific structures and geometries has been extensively exploited by nature for the assembly of viruses able to protect their genome and deliver it into the host cell, or for the storage and transport of iron molecules by ferritin, a self-assembled protein cage ubiquitously found in different biological systems (Chakraborti and Chakrabarti, 2019; Perlmutter and Hagan, 2015). These nanometric protein cages are built up by the establishment of intermolecular non-covalent interactions between different protein building blocks, constraining the assembly to a well-defined cage with specific geometries. Advances in protein design are offering an increasing catalogue of protein cages created from scratch, displaying different geometries, symmetries and properties. These macromolecular structures are often sustained by α-helical polypeptides arranged via helical bundles or coiled-coil architectures (Lapenta et al., 2018). In this way, and inspired by the DNA origami technology, Roman Jerala and colleagues have developed a semi-automated computational design strategy to assist the rational design of protein cages in any polyhedral shape of interest. In their pioneering work, they achieved the assembly of cages *in vitro* and *in vivo* with tetrahedron, four-sided pyramid, and triangular prism shapes without any associated toxicity (Ljubetič et al., 2017). The biomedical applications of such nanostructures are limitless, acting as drug delivery agents able to transport cargo molecules inside the cage and specifically deliver them at the selected destination, or as vaccines able to present different epitopes to the immune system in a safe way, acting as virus-like particles. A recent study developed by David Baker and colleagues has reported the *de novo* creation of protein cages able to assemble antibodies into well-defined architectures with controlled valency and geometry, enhancing the viral neutralization of SARS-CoV-2 virus and assisting the *ex vivo* T cell activation and proliferation required for Chimeric Antigenic Receptor

(CAR)-T cell therapies (Divine et al., 2021). Strikingly, such antibody-based nanocages can selectively target any desired epitope and contain cargo molecules in their interior, expanding the available toolbox for the *a la carte* creation of highly selective and modular drug delivery systems.
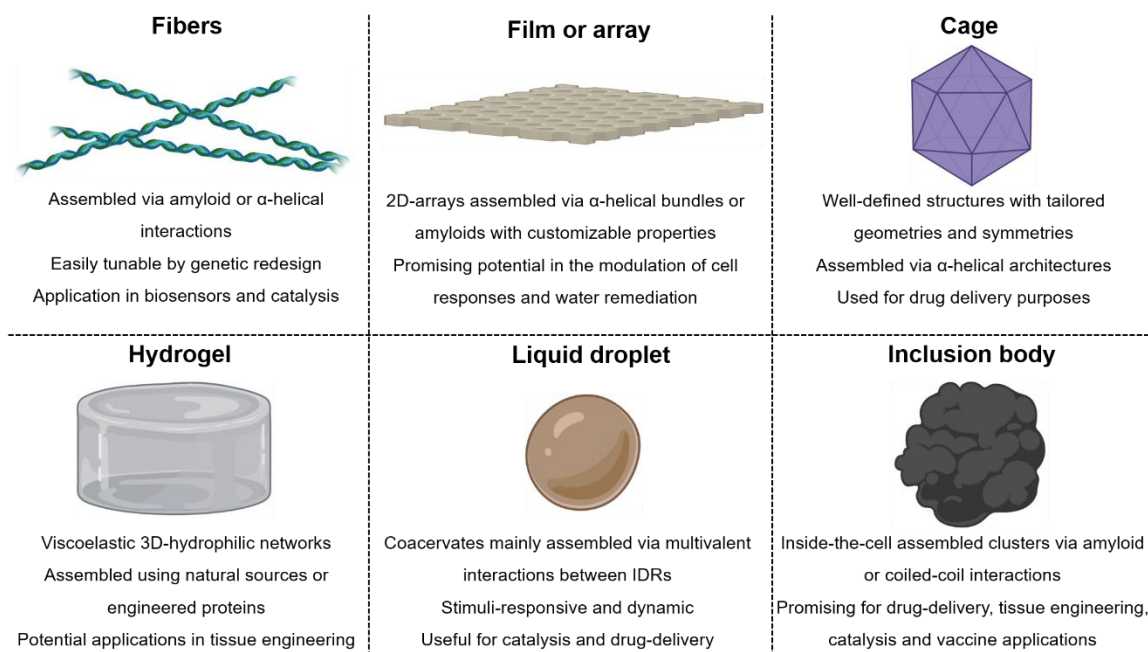
Hydrogels: Protein hydrogels are 3D-hydrophilic networks formed by physical gelation or chemical crosslinking able to retain large amounts of water. These supramolecular structures are inherently viscoelastic, biodegradable, biologically-friendly and share mechanical properties with most biological tissues, holding a great potential for tissue engineering and biomedical applications (Huerta-López and Alegre-Cebollada, 2021). They can be built using natural proteins isolated from vegetal or animal sources (e.g., collagen, elastin, silk proteins and soy) or created from scratch using synthetic proteins. As it occurs with the aforementioned assemblies, the mechanical properties of hydrogels can be tuned by mutating the constituent protein building blocks, offering a variety of stimuli-responsive structures with unique physicochemical properties. Due to their versatility, protein hydrogels composed of collagen or keratin have found application in cosmetology, for the treatment of skin problems such as wrinkles and scars (Mandal et al., 2020; Mitura et al., 2020). In tissue engineering, these structures can provide not only with a mechanical scaffold for the adhesion and proliferation of cells, but also with the chemical stimuli for promoting specific cell functions and behaviour (Schloss et al., 2016), and, they can act as drug delivery agents able to specifically release cargo molecules upon external cues (Narayanaswamy and Torchilin, 2019).

Protein-liquid droplets: The formation of protein coacervates or droplets via liquid-liquid phase separation (LLPS) has been identified as a pivotal process underlying the regulation of a wide variety of cellular functions (Boeynaems et al., 2018; Shin and Brangwynne, 2017). Intense research has been devoted to identifying the molecular determinants behind this process, involving the mixing of proteins and other molecules such as nucleic acids into an oil-like phase. This protein coacervation is often prompted by the establishment of multivalent intermolecular interactions (e.g., electrostatic interactions, π-π interactions, cation-π interactions and hydrogen bonding) between residues mainly located at IDRs or IDPs, which can be modulated by modifying the pH, the temperature, the protein concentration and the ionic strength of the solution (Dignon et al., 2020). Thenceforth, the assembly of such supramolecular architectures can be modulated by environmental control, and their

composition easily modified by fusing different IDRs or IDPs to the protein/s of interest. This simple genetic design allows embedding a target protein inside a liquid-like structure, sustained by the interactions between the appended IDRs. Importantly, a detailed analysis of the molecular grammar governing this process has provided the stickers-and-spacers model of phase separation, deciphering the sequence features underpinning this reaction and opening the door for modulating the assembly process at the residue level (Choi et al., 2020; Martin et al., 2020). This has permitted to create multifunctional microcompartments sustained by natural or artificial IDPs where enzymatic reactions can occur at higher rates than in soluble conditions (Küffner et al., 2020). Moreover, the decoration of these enzymatic compartments with inorganic and magnetic nanoparticles enables their manipulation by applying external magnetic fields, representing an important advance toward the creation of stimuli-sensitive and biological-inorganic systems capable of performing multiple functions (Capasso Palmiero et al., 2020). These liquid-like compartments have been also exploited for the sequestration and functional isolation of proteins inside the cell, allowing to control endogenous cellular processes by annulling/programming protein functions or controlling the release of cargo molecules to the cytosol (Garabedian et al., 2021; Hastings and Boeynaems, 2021; Krauss, 2021; Reinkemeier et al., 2019).

Inclusion bodies: The use of microbial cell factories to produce recombinant proteins is a routine procedure for obtaining high yields of therapeutic proteins in the biotechnological industry (Sanchez-Garcia et al., 2016). However, this process is often associated with the accumulation of the desired protein as insoluble deposits, protein inclusion bodies (IBs), in the bacterial cytosol. These refractile and submicrometric intracellular particles contain significant amounts of active protein, providing assembled and functional particles cost-effectively and in a ready-to-use format (De Marco et al., 2019). Functional IBs composed of a wide variety of proteins allowed obtaining reusable catalysts able to perform enzymatic cascade reactions, drug delivery agents for the treatment of cancer, and protein-based vaccines or immunostimulants capable of protecting against harmful bacterial infections (Céspedes et al., 2020; Jäger et al., 2020; Kloss et al., 2018; Krauss et al., 2017; Pesarrodona et al., 2019; Roca-Pinilla et al., 2020; Schetters et al., 2020; Torrealba et al., 2016).

In the present doctoral thesis, we delve into the field of functional IBs, providing a review focused on the recent advances and applications of this technology (Gil-Garcia and Ventura, 2021a) (Chapter 1) and two research articles that describe the use of the coiled-coil architecture to create multifunctional IBs (Gil-Garcia et al., 2020; Gil-Garcia and Ventura, 2021b) (Chapters 3 and 4).

Fibers

Assembled via amyloid or α-helical interactions
Easily tunable by genetic redesign
Application in biosensors and catalysis

Film or array

2D-arrays assembled via α-helical bundles or amyloids with customizable properties
Promising potential in the modulation of cell responses and water remediation

Cage

Well-defined structures with tailored geometries and symmetries
Assembled via α-helical architectures
Used for drug delivery purposes

Hydrogel

Viscoelastic 3D-hydrophilic networks
Assembled using natural sources or engineered proteins
Potential applications in tissue engineering

Liquid droplet

Coacervates mainly assembled via multivalent interactions between IDRs
Stimuli-responsive and dynamic
Useful for catalysis and drug-delivery

Inclusion body

Inside-the-cell assembled clusters via amyloid or coiled-coil interactions
Promising for drug-delivery, tissue engineering, catalysis and vaccine applications

Figure 6. Schematic representation of the supramolecular architectures exploited to create functional biomaterials. A schematic illustration of different protein-based supramolecular assemblies is shown together with a short description of their main properties and potential applications. Created with Biorender.com.

RESEARCH OBJECTIVES

The main objectives of the present doctoral thesis are the creation of improved therapeutic proteins and novel protein-based nanomaterials with potential application in biomedicine. Specifically, the strategy consisted in (I) the rational redesign of protein solubility without altering the active conformation using *in silico* tools, and (II) the creation of novel and versatile multifunctional protein-based nanomaterials stabilized by α-helical contacts.

Chapter 1: Coiled-Coil Based Inclusion Bodies and Their Potential Applications

The production of recombinant proteins using bacterial cells is often associated with the formation of insoluble protein aggregates in their cytosol. Traditionally, these refractile deposits, commonly known as protein inclusion bodies (IBs), have been considered as undesired and useless products without any significant value. However, several studies have corroborated that these proteinaceous elements can be a reservoir of functional protein, suggesting a potential application as functional nanomaterials. In principle, any protein, independently of its intrinsic solubility, can be assembled as IBs when fused to aggregation-prone sequences. These insoluble polypeptides, referred as IB-tags, are easily joined to the target protein, triggering the assembly of amyloid-based nanoparticles. Over the last years, this simple but effective approach has allowed the creation of functional IBs for different biomedical and biotechnological applications. However, this amyloid conformation is necessarily associated with a decreased functionality and a potential intrinsic cytotoxicity, compromising the biomedical application of such nanoparticles. To overcome these drawbacks, the use of non-amyloid scaffolds, based on the self-assembling properties of α-helical coiled-coils, is emerging as an alternative strategy to create highly functional IBs.

In the present mini-review, we aimed at summarizing the state-of-the-art of functional IBs technology, delving into the potential advantages of using coiled-coils as IB-tags, and addressing their promising applications in biotechnology and biomedicine.

Chapter 2: Combining Structural Aggregation Propensity and Stability Predictions to Redesign Protein Solubility

The production and downstream processing of therapeutic proteins often implies concentrations that are orders of magnitude beyond their cellular levels, resulting in their aggregation. This undesired and harmful reaction can occur at every step of protein-based drug development, jeopardizing the functionality and stability of the therapeutic protein, and entailing an inherent toxicity. Thus, biotechnological and pharmaceutical companies devote large efforts to avoid such deleterious reaction, usually by performing trial-and-error assays that are expensive and time-consuming. In this regard, the use of computational tools aimed at foreseeing the structural aggregation propensity of a given polypeptide can provide an alternative to optimize the selection of candidates in a cost-effective way. Here, we aimed to demonstrate that an in-house developed structural aggregation predictor, that considers protein stability, might allow to redesign the solubility of structurally unrelated proteins, without impacting their native conformation.

Specifically, the main objectives of this chapter were:

- To validate the structure-based aggregation predictor AGGRESCAN 3D (A3D) as a tool for assisting the redesign of protein solubility.
- To redesign the solubility of a folding reporter version of the green fluorescent protein (fr-GFP).
- To determine the 3D structure of this improved version of the fr-GFP and confirm that solubilizing mutations do not alter the native conformation.
- To study the impact of protein thermodynamic stability in the overall protein solubility.
- To include stability calculations in the prediction pipeline to ensure a correct interplay between solubility and thermodynamic stability.
- To redesign the solubility of a human antibody without compromising its conformation and stability.

Chapter 3: Coiled-Coil Inspired Functional Inclusion Bodies

The creation of functional IBs is often triggered by the fusion of aggregation-prone peptides or domains to the protein of interest, leading to the formation of amyloid-like nanostructures, where the functional protein is trapped. Over the last years, a wide range of aggregation-prone tags, artificial peptides and natural proteins, have been employed to induce IBs' formation. The majority of these tags promote the formation of IBs sustained by intermolecular β-sheet interactions. This fact can negatively impact their activity and safety, since a significant fraction of the target protein is misfolded and incorporated to the amyloid scaffold, and the release of oligomeric species can be potentially cytotoxic. In order to offer an alternative to these classical amyloid-based IBs, here, we attempted to create functional IBs exploiting the self-assembling properties of α-helical coiled-coils.

The specific objectives of this chapter were:

- To computationally identify an optimal protein candidate to build up α-helix-based IBs.
- To study the self-assembling properties of this protein.
- To create fluorescent IBs sustained by α-helical contacts and characterize their conformational traits.
- To compare the activity and biocompatibility of coiled-coil based IBs with those of amyloid-like IBs.

Chapter 4: Multifunctional Antibody-Conjugated Coiled-Coil Protein Nanoparticles for Selective Cell Targeting

Monoclonal antibodies are the most lucrative protein-based drugs due to their application in bioimaging and therapeutics. However, their commercialization is often hampered by their low stability and half-life. In order to overcome these limitations, different strategies consisting in the covalent conjugation of antibodies to nanoparticles have been described. This antibody-decoration of nanosized particles provides two advantages: (I) The antibody is stabilized in a nanoscopic assembly, increasing its *in vivo* half-life, and (II) provides multivalency, resulting in an enhanced binding avidity and agonism through receptor clustering. These nanoparticles can be simultaneously conjugated with different antibodies, mimicking the function of bispecific antibodies. However, covalent conjugation steps can negatively impact the antibody's functionality, since the antigen-binding sites can be buried upon binding. In this chapter, we aimed to generate biocompatible and fluorescent protein-based nanoparticles capable of binding monospecific antibodies non-covalently and simultaneously, for potential applications in biomedicine.

The specific goals of this project were:

- To create coiled-coil-based IBs displaying fluorescent and antibody-capturing activities.
- To assess their conformational traits and bifunctionality.
- To evaluate their biocompatibility in human cells.
- To test the ability of antibody-conjugated IBs to target specific antigens in cells without associated toxicity.
- To build up dual antibody-decorated nanoparticles.
- To assess if double antibody-conjugated IBs can bring two unrelated cell types in close proximity.

CHAPTER 1

Coiled-Coil Based Inclusion Bodies and Their Potential Applications

Marcos Gil-Garcia* and Salvador Ventura*

*co-corresponding authors

Coiled-Coil Based Inclusion Bodies and Their Potential Applications

Marcos Gil-Garcia* and Salvador Ventura*

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

The production of recombinant proteins using microbial cell factories is frequently associated with the formation of inclusion bodies (IBs). These proteinaceous entities can be sometimes a reservoir of stable and active protein, might display good biocompatibility, and are produced efficiently and cost-effectively. Thus, these submicrometric particles are increasingly exploited as functional biomaterials for biotechnological and biomedical purposes. The fusion of aggregation-prone sequences to the target protein is a successful strategy to sequester soluble recombinant polypeptides into IBs. Traditionally, the use of these IB-tags results in the formation of amyloid-like scaffolds where the protein of interest is trapped. This amyloid conformation might compromise the protein's activity and be potentially cytotoxic. One promising alternative to overcome these limitations exploits the coiled-coil fold, composed of two or more α-helices and widely used by nature to create supramolecular assemblies. In this review, we summarize the state-of-the-art of functional IBs technology, focusing on the coiled-coil-assembly strategy, describing its advantages and applications, delving into future developments and necessary improvements in the field.

Keywords: functional inclusion bodies, biomaterials, tags, coiled-coil, biotechnology, biomedicine, protein assembly

INTRODUCTION

The use of microorganisms as cell factories to produce biological products (as therapeutic proteins), often leads to the formation of insoluble protein deposits in their cytoplasm (Villaverde and Carrió, 2003). These protein deposits, commonly known as inclusion bodies (IBs), have been considered a significant bottleneck for obtaining high yields of well-folded and soluble protein since they were understood as reservoirs of inactive protein (Baneyx and Mujacic, 2004).

IBs are refractile and submicron protein nanoparticles (between 50–1,000 nm) that usually accumulate at the poles of the cell (Luo et al., 2006; Margreiter et al., 2008; García-Fruitós et al., 2009, García-Fruitós et al., 2012; Castellanos-Mendoza et al., 2014). These proteinaceous inclusions are mechanically stable in a wide range of temperatures and pHs, present a pseudo-spherical shape, and are mainly composed of the target protein (De Marco et al., 2019); although the composition and purity of the IBs significantly depends on the identity of the recombinant protein. Moreover, a significant fraction of the protein trapped in these IBs can be well-folded and functional, suggesting that these nanoparticles might be a source of active protein (Ventura and Villaverde, 2006). Due to their inherent stability, they can be easily purified by simple cell disruption methods, obtaining high yields of pure protein cost-effectively. Furthermore, purified IBs can be stored in lyophilized form or at −80°C for extended periods (García-Fruitós et al., 2009).

IBs usually present an amyloid-like architecture responsible for their excellent mechanical robustness, as corroborated by several biophysical techniques (Carrió et al., 2005; Morell et al., 2008; Wang et al., 2008). However, despite this amyloidogenic character and the intrinsic toxicity of amyloid oligomeric forms (Díaz-Caballero et al., 2020), these IBs are assumed to be non-toxic materials, and thanks to their submicrometric size and functional character, they have found biomedical and biotechnological applications (Villaverde et al., 2015).

The formation of insoluble IBs can be induced by fusing aggregation-prone peptide or protein tags to the globular protein of interest, irrespectively of its solubility. Typically, the tag provides the driving force for forming the intermolecular β-sheet contacts that sustain the IBs amyloid-like nanostructure. Different research groups have used coiled-coil domains as IB-inducing tags (IB-tags) with significant success in the last years, suggesting that these α-helix-based tags are a feasible alternative to the β-sheet-based ones (Jäger et al., 2020).

In this mini-review, we illustrate recent advances in the field of functional IBs, discussing their biotechnological and biomedical applications, with particular emphasis on the uses of the new class of coiled-coil IBs. We also describe the assembly of cell component-free artificial IBs.

PROTEIN INCLUSION BODIES AS AN UNEXPECTED SOURCE OF ACTIVE PROTEIN

The development of recombinant DNA technologies has allowed producing a wide diversity of proteins using heterologous expression systems. This includes therapeutic proteins such as replacement enzymes, hormones, or antibodies (Sanchez-Garcia et al., 2016; Walsh, 2018).

Proteins function at concentrations at which they remain soluble, whereas exceeding their solubility limits results in their aggregation and inactivation, according to "the life on the edge" hypothesis (Tartaglia et al., 2007; Vecchi et al., 2020). Therefore, the proteins' cellular levels are tightly regulated, both in time and space, according to their intrinsic solubilities (de Groot and Ventura, 2010; Castillo et al., 2011; Sanchez de Groot et al., 2015). The production of such polypeptides recombinantly often surpasses several orders of magnitude these solubility limits, which together with the lack of post-translational modifications or the deregulation of the refolding machinery, trigger the occurrence of protein misfolding and aggregation events inside the cell. This unavoidable connection between intracellular protein concentration and aggregation constitutes one of the main limitations for producing recombinant proteins at preparative levels in the biotechnological and pharmaceutical industry, evoking a considerable loss of resources and time (Roberts, 2014). Therefore, a significant effort has been devoted to implementing strategies aimed to increase the yield of soluble and functional protein, such as the redesign of the intrinsic protein solubility (Gil-Garcia et al., 2018) or the fusion of solubility tags (Costa et al., 2014), among others. These

approaches try to push the solubility of the target protein above its natural limit. Soluble and folded protein has been obtained from IBs by denaturing-renaturing procedures, but the process is cost-ineffective and needs to be tailored for the protein of interest (Singhvi et al., 2020).

Traditionally understood as undesired byproducts of protein production processes, IBs have been ignored and avoided for decades. Two independent studies published in the 80–90s reported biological activity in the IBs formed by two different enzymes (β-galactosidase and endoglucanase D) (Worrall and Goss, 1989; Tokatlidis et al., 1991). However, they were considered anecdotic, and only when, in 2005, two additional studies recapitulated these data, our perception of IBs molecular properties changed dramatically. Jevsevar and coworkers obtained IBs of the human granulocyte-colony stimulating factor from where the functional protein could be easily isolated without the need for denaturing-renaturing steps (Jevsevar et al., 2005). In parallel, García-Fruitós and coworkers demonstrated that the IBs formed by two fluorescent proteins (blue and green fluorescent proteins) and two enzymes (human dihydrofolate reductase (DHFR) and E. coli β-galactosidase) were functional (García-Fruitós et al., 2005).

The awareness that IBs were functional particles, together with their easy and cost-effective production and the remarkable mechanical properties of such nanoparticles, paved the way for creating rationally designed IBs with a wide variety of functionalities.

THE USE OF IB-TAGS AS A SMART STRATEGY FOR OBTAINING COST-EFFECTIVE AND READY-TO-USE FUNCTIONAL IBs

The formation of IBs involves the establishment of homotypic intermolecular interactions, a process that is strongly dependent on the microenvironment. In this way, it has been demonstrated that different cell culture variables such as the pH of the solution (Castellanos-Mendoza et al., 2014; Calcines-Cruz et al., 2018), the production temperature (de Groot and Ventura, 2006; Peternel et al., 2008; Seras-Franzoso et al., 2014), the time of induction (Margreiter et al., 2008) or the concentration of inducer agent (Luo et al., 2006), influence not only the kinetics of IBs formation but also the size, stability, and activity of these nanoparticles (De Marco et al., 2019).

Strange as it may seem, it is as challenging to produce an aggregation-prone protein in a soluble conformation as it is to force a soluble protein to form IBs efficiently, regardless of the culture conditions. In this latter case, the fusion of an aggregation-prone sequence to the target protein, commonly known as IB-tag, may solve the problem and incorporate the protein of interest into IBs.

In the last years, different IB-tags, consisting of small artificial peptides or large protein domains, have been exploited to create functional IBs (Krauss et al., 2017; Jäger et al., 2020).
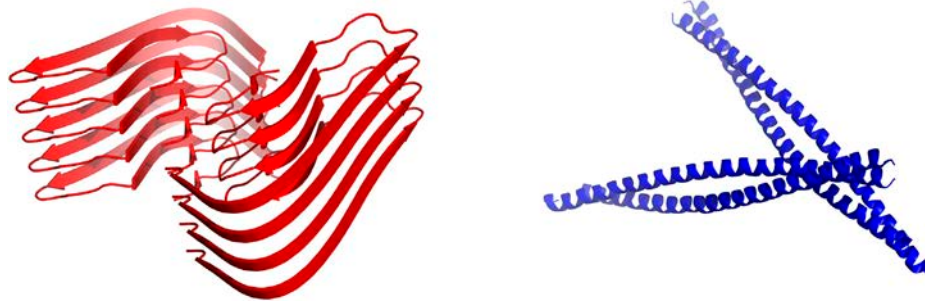
FIGURE 1 | Structural building blocks of amyloid based and coiled-coil based IBs. Schematic representation of cross-β-sheet (red, according to PDB: 5OQV) and coiled-coil (blue, according to PDB: 2JEE) structures. They intend to illustrate the kind of scaffolds sustaining amyloid and coiled-coil based IBs.

A large group of such tags corresponds to small artificial peptides such as the β-sheet forming ELK16 or the hydrophobic GFIL8 peptides, that have been exploited for forming catalytically active IBs composed of oxidases, hydrolases, and oxidoreductases (Wu et al., 2011; Wang et al., 2015).

Apart from these small IB-tags, other aggregation-tags involve natural protein domains such as the viral capsid protein (VP1) of the foot-and-mouth disease virus and the variant F19D of the amyloid β-peptide (Aβ42). The VP1 domain has been exploited for obtaining functional IBs in the yeast *Pichia pastoris*, constituting an alternative for producing IBs of proteins needing post-translational modifications (Rueda et al., 2016). In the case of Aβ42, it has been used by our group to model amyloid aggregation and identifying chemical modulators of this deleterious reaction (De Groot et al., 2006; Dasari et al., 2011; Villar-Piqué et al., 2012).

Alternative aggregation-prone domains for obtaining catalytic IBs are the cellulose-binding domain (CBD) of different organisms such as *Clostridium cellulovorans* and *Cellulomonas fimi* (Nahalka and Nidetzky, 2007; Choi et al., 2011), a variant of the maltose-binding protein (MBP), known as MalE31, from *E. coli* (Arié et al., 2006) and the pyruvate oxidase (PoxB) from *Paenibacillus polymyxa* (Park et al., 2012). These IB-tags have been used to produce catalytic IBs composed of amylases, alkaline phosphatases, and β-lactamases, among other enzymes. Interestingly, the IBs promoted by MalE31 are located at the periplasm of the *E. coli* cell, facilitating their isolation. Furthermore, the IBs formed by PoxB present an intrinsic enzymatic activity, allowing to obtain IBs with simultaneous and different catalytic activities depending on the appended enzyme.

Finally, the signal sequence of *E. coli* TorA (ssTorA) has been successfully used to promote the accumulation into IBs of two highly soluble proteins, the MBP and the thioredoxin-1 (TrxA), usually exploited as solubility-enhancing fusion tags (Jong et al., 2017), demonstrating the pro-aggregational potency of this IB-tag. Moreover, a subsequent study has demonstrated that the IBs formed by the fusion of ssTorA to ovalbumin (OVA)-derived epitopes can act as an antigenic vaccine formulation for T cell response (Schetters et al., 2020). Mutagenesis-screening of the ssTorA sequence has allowed the identification of improved versions of the tag with enhanced IB-formation properties (Jong et al., 2019).

COILED-COIL DOMAINS AS SCAFFOLDS FOR THE CREATION OF HIGHLY FUNCTIONAL IBs

A significant fraction of the aforementioned IB-tags promotes the formation of amyloid-based IBs (Cano-Garrido et al., 2013; Carrió et al., 2005; Choi et al., 2011; Morell et al., 2008; Wu et al., 2011) (**Figure 1**). In this supramolecular assembly, the amyloid architecture acts as a mechanically stable scaffold where the globular and functional protein is trapped (Cano-Garrido et al., 2013). However, these amyloid assemblies present two potential drawbacks: (I) A significant fraction of the target protein must be unfolded and inactivated to sustain the amyloid scaffold (Morell et al., 2008) and, (II) although IBs are assumed to be non-toxic entities, the release of toxic oligomeric species from the IBs cannot be entirely ruled out.

Different research groups have made use of the coiled-coil fold as a scaffold for the design of self-assembled protein nanostructures (Wu and Collier, 2017). The coiled-coil structure consists of two or more α-helices twisted around each other in a parallel or anti-parallel orientation (**Figure 1**). As conjectured by Crick in the 50's (Crick, 1953), the sidechains of the residues from the different helices interlock along the core of the structure with a defined periodicity (heptad, hendecad, or pentadecad repeats), creating a distinctive packing named the "knobs-into-holes" (Lupas and Bassler, 2017; Lupas et al., 2017). We know now that the coiled-coil fold comprises a vast diversity of periodicities and structures, allowing to perform a wide range of functions in nature (Truebestein and Leonard, 2016) and opening the avenue for the design of bioinspired nanomaterials such as nanotubes (Xu et al., 2013), nanocages (Ljubetič et al., 2017) and vaccines (Morris et al., 2019).

In this context, the use of the coiled-coil motif as IB-tag might allow avoiding the caveats of β-sheet enriched IBs, since the assembly of these α-helical structures is not expected to impact the fold of the adjacent proteins or result in potentially toxic species. The first coiled-coil domain used to obtain functional IBs

FIGURE 2 | Biotechnological and biomedical applications of functional IBs. Functional IBs can be exploited for biotechnological purposes as a source of pure and active protein and as reusable and immobilized catalysts. Regarding their use in biomedical applications, IBs can act as targeted-nanopills, as biocompatible scaffolds for tissue engineering, as immunostimulants to act against lethal infections or as antimicrobial agents.

was the tetramerization domain of the cell surface protein tetrabrachion (TDoT) of *Staphylothermus marinus*, a right-handed and parallel tetramer consisting of an 11-residues repeat, which has been extensively exploited for building up catalytic IBs (Stetefeld et al., 2000; Diener et al., 2016; Kloss et al., 2018a, Kloss et al., 2018b; Jäger et al., 2018, Jäger et al., 2019; Lamm et al., 2020). Another coiled-coil domain employed to create catalytically-active IBs was 3HAMP, derived from the oxygen sensor protein Aer2 of *Pseudomonas aeruginosa* (Kloss et al., 2018a; Jäger et al., 2019). This domain consists of a homodimeric structure formed by two parallel monomers harboring three successive helical regions connected by flexible linkers (Airola et al., 2010). More recently, Garcia-Fruitós and coworkers reported using two leucine zippers (Jun and Fos) as IB-tags to produce fluorescent and antimicrobial IBs with remarkable specific activity (Roca-Pinilla et al., 2020a; Roca-Pinilla et al., 2020b). Without any doubt, these α-helical tags can be used to form active IBs; however, except for Jun and Fos, the biophysical properties of the IBs were not characterized in detail, and the preservation of the native α-helical structure in the final assembly could not be corroborated.

To provide a suitable coiled-coil tag that preserves its native structure when embedded in IBs, our group performed a computational analysis of the biophysical properties of different IB-tags and selected the ZapB protein, a non-essential factor involved in the cell division process of *E. coli* (Ebersbach et al., 2008), as an optimal candidate for obtaining coiled-coil inspired IBs. ZapB is a homodimeric and anti-parallel coiled-coil domain able to self-assemble *in vivo*, and therefore, it allows the

production of α-helix-rich IBs when fused to a target protein, as demonstrated by circular dichroism and Fourier transformed infrared spectroscopy (Gil-Garcia et al., 2020). ZapB was fused to two different fluorescent proteins in our study, obtaining active and biocompatible nanoparticles displaying native-like fluorescent spectra and devoid of any amyloid character. As intended, these IBs were significantly more active than their amyloid-like IBs counterparts, indicating that the amount of inactivated target protein within ZapB-based IBs is low.

Although further biophysical studies of other coiled-coil-based IBs are needed to ascertain the generic non-amyloid character of these nanoparticles, the collected data suggest that this new class of IB-tags might substitute aggregation-prone sequences as a strategy for the production of high-quality functional IBs.

BIOMEDICAL AND BIOTECHNOLOGICAL APPLICATIONS OF FUNCTIONAL IBs: IS THERE ROOM FOR IMPROVEMENT?

From a biotechnological point of view, IBs can be used as a source of pure and active protein, since the functional target protein can be easily isolated using mild-solubilization methods (for example, using low concentrations of organic solvents or alkaline pH) (Singh et al., 2015; Singhvi et al., 2020) (**Figure 2**). A more industry-oriented application of functional IBs is their use as reusable biocatalysts. Enzyme immobilization is a common strategy for improving the stability, enantioselectivity, and

reusability of enzymes and the easy isolation of the reaction product (Mohamad et al., 2015). In this context, catalytic IBs fulfill the requirements of enzyme immobilization strategies, acting as porous and highly stable enzymatic reactors that can be separated from the reaction mixture by simple centrifugation (**Figure 2**). The use of catalytically active-IBs has been demonstrated for a wide range of different enzymes such as oxidases, reductases, and synthases, among others (Jäger et al., 2020), and different studies have reported striking recyclability for catalytic IBs (Koszagova et al., 2018), together with higher stability and activity under harsh conditions, like extreme pHs or the presence of organic solvents, than their soluble counterparts (Diener et al., 2016; Kloss et al., 2018a).

Apart from their biotechnological applications, IBs have been successfully applied in biomedicine acting as functional scaffolds in tissue engineering (García-Fruitós et al., 2010; Seras-Franzoso et al., 2013; Tatkiewicz et al., 2013), as targeted-delivery agents (Vázquez et al., 2012; Unzueta et al., 2018; Céspedes et al., 2019; Pesarrodona et al., 2019), or as immunostimulants (Schetters et al., 2020; Thwaite et al., 2018; Torrealba et al., 2016a, Torrealba et al., 2016b).

With regard to their application in tissue engineering, it has been demonstrated that the decoration of surfaces using IBs provides a mechanotransduction-mediated stimulation of cell adhesion and proliferation (via activation of the ERK pathway) (Seras-Franzoso et al., 2012) (**Figure 2**). This proliferative effect can also be achieved by the release of functional protein from IBs, as demonstrated for fibroblast growth factor-2 (Seras-Franzoso et al., 2014), with IB-decorated surfaces acting not only as a material to which cells can attach but also as proactive delivery agents. This property resembles that of the endocrine secretory granules, which store hormones in an amyloid-like conformation (Maji et al., 2009). Villaverde and coworkers have exploited this IBs property for delivering proteins to tumors, arresting their growth in mouse models of human colorectal (Céspedes et al., 2019) and breast (Pesarrodona et al., 2019) cancer (**Figure 2**). These IBs are constituted by a specific ligand for the CXCR4 receptor (overexpressed in metastatic cancer cells) and the PE24 toxin from *Pseudomonas aeruginosa*, or alternatively, by a specific ligand for the CD44 receptor and two anti-tumoral proteins, p31 and Omomyc.

Another intriguing biomedical application of functional IBs is the modulation of the immune system to fight lethal infections. IBs composed of cytokines such as TNFα displayed a prophylactic effect in zebrafish, protecting from infection by the pathogen *Pseudomonas aeruginosa*. This strategy overcomes the limitations associated with the low stability and short half-life of soluble cytokines (Torrealba et al., 2016a) (**Figure 2**).

Finally, the IBs technology can be exploited for the production of protein-based nanoparticles displaying antimicrobial activities. In this way, Roca-Pinilla et al. have been successful in the production of IBs composed of an antimicrobial polypeptide that combines different functional moieties, active against diverse Gram-negative and Gram-positive bacteria, including some multi-resistant strains (Roca-Pinilla et al., 2020b) (**Figure 2**).

The higher specific activity displayed by proteins when they are embedded in coiled-coil-based IBs allows forecasting that this type of assemblies would turn useful for biomedical applications, although their applicability in this context remains to be explored. First, the largest proportion of folded and active protein in these assemblies should permit reducing the critical concentrations for *in vivo* administration and the production costs. Additionally, lower doses would diminish undesired side effects, which, together with the lack of significant intrinsic toxicity demonstrated for coiled-coil based nanoparticles (Utterström et al., 2021), would make α-helix-rich IBs safer protein nanoparticles, mainly because they are expected to be devoid of potentially toxic β-sheet sustained conformations. Finally, because the native-like intermolecular interactions sustaining coiled-coil IBs are expected to be less robust than the strong amyloid-like contacts that glue β-sheet IBs (Carrió et al., 2005), these nanoparticles could display a higher release-efficiency upon their administration. However, an excessive release might be a drawback for their use as biocatalytic nanoparticles since it can negatively impact the product's purity and limit their recyclability.

CONCLUSIONS AND PERSPECTIVES

Initially considered waste products, IBs have opened as bio-nano-technologically relevant tools for a plethora of applications. The use of aggregation-inducing tags has allowed obtaining the desired target protein in an assembled state, independently of the protein identity, structure, or aggregation propensity.

In the last years, different strategies such as the use of lipopolysaccharide (LPS)-free bacterial strains have been implemented to create safe IBs devoid of bacterial toxic or immunocompromising components (Rueda et al., 2014; Gifre-Renom et al., 2018; Carratalá et al., 2021). An alternative to solve this problem is creating artificial IBs (ArtIBs) by assembling purified and initially soluble protein as IBs employing cell-free methods, such as the coordination of His-tags to divalent cations (Sánchez et al., 2020; Serna et al., 2020).

Considering the advantages of coiled-coil based-IBs and ArtIBs, we envision a promising avenue for the generation of a new class of functional IBs by mixing these two concepts since it has been demonstrated that the assembly of different coiled-coil proteins is dependent on the presence of divalent cations (Cristie-David and Marsh, 2019). In this way, a next generation of functional IBs can be potentially generated by creating ArtIBs assembled via coiled-coil interactions. On the one hand, the relatively high surface contact in this motif should render these nanoparticles highly stable, whereas the dependence on cations presence to keep their assembled state would allow tight and reversible control of their nanostructure.

All in all, the use of IBs as active nanoparticles is an emerging field that will continue cost-effectively providing new and unique applications, without the need for complex and harsh chemical reactions to assemble their active components; this environmentally friendly character is yet another advantage of this simple but effective technology.

AUTHOR CONTRIBUTIONS

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Airola, M. V., Watts, K. J., Bilwes, A. M., and Crane, B. R. (2010). Structure of Concatenated HAMP Domains Provides a Mechanism for Signal Transduction. *Structure* 18, 436–448. doi:10.1016/j.str.2010.01.013

Arié, J.-P., Miot, M., Sassoon, N., and Betton, J.-M. (2006). Formation of Active Inclusion Bodies in the Periplasm ofEscherichia Coli. *Mol. Microbiol.* 62, 427–437. doi:10.1111/j.1365-2958.2006.05394.x

Baneyx, F., and Mujacic, M. (2004). Recombinant Protein Folding and Misfolding in Escherichia coli. *Nat. Biotechnol.* 22, 1399–1408. doi:10.1038/nbt1029

Calcines-Cruz, C., Olvera, A., Castro-Acosta, R. M., Zavala, G., Alagón, A., Trujillo-Roldán, M. A., et al. (2018). Recombinant-phospholipase A2 Production and Architecture of Inclusion Bodies Are Affected by pH in Escherichia coli. *Int. J. Biol. Macromol.* 108, 826–836. doi:10.1016/j.ijbiomac.2017.10.178

Cano-Garrido, O., Rodríguez-Carmona, E., Díez-Gil, C., Vázquez, E., Elizondo, E., Cubarsi, R., et al. (2013). Supramolecular Organization of Protein-Releasing Functional Amyloids Solved in Bacterial Inclusion Bodies. *Acta Biomater.* 9, 6134–6142. doi:10.1016/j.actbio.2012.11.033

Carratalá, J. V., Gifre-Renom, L., Roca-Pinilla, R., Villaverde, A., Arís, A., Garcia-Fruitós, E., et al. (2021). Selecting Subpopulations of High-Quality Protein Conformers Among Conformational Mixtures of Recombinant Bovine MMP-9 Solubilized from Inclusion Bodies. *Int. J. Mol. Sci.* 22, 3020. doi:10.3390/ijms22063020

Carrió, M., González-Montalbán, N., Vera, A., Villaverde, A., and Ventura, S. (2005). Amyloid-like Properties of Bacterial Inclusion Bodies. *J. Mol. Biol.* 347, 1025–1037. doi:10.1016/j.jmb.2005.02.030

Castellanos-Mendoza, A., Castro-Acosta, R. M., Olvera, A., Zavala, G., Mendoza-Vera, M., García-Hernández, E., et al. (2014). Influence of pH Control in the Formation of Inclusion Bodies during Production of Recombinant Sphingomyelinase-D in Escherichia coli. *Microb. Cel Fact* 13, 137. doi:10.1186/s12934-014-0137-9

Castillo, V., Graña-Montes, R., and Ventura, S. (2011). The Aggregation Properties of Escherichia coli Proteins Associated with Their Cellular Abundance. *Biotechnol. J.* 6, 752–760. doi:10.1002/biot.201100014

Céspedes, M. V., Cano-Garrido, O., Álamo, P., Sala, R., Gallardo, A., Serna, N., et al. (2019). Engineering Secretory Amyloids for Remote and Highly Selective Destruction of Metastatic Foci. *Adv. Mater.* 32, 1907348–1907349. doi:10.1002/adma.201907348

Choi, S.-L., Lee, S. J., Ha, J.-S., Song, J. J., Rhee, Y. H., and Lee, S.-G. (2011). Generation of Catalytic Protein Particles in Escherichia coli Cells Using the Cellulose-Binding Domain from Cellulomonas Fimi as a Fusion Partner. *Biotechnol. Bioproc. E* 16, 1173–1179. doi:10.1007/s12257-011-0336-8

Costa, S., Almeida, A., Castro, A., and Domingues, L. (2014). Fusion Tags for Protein Solubility, Purification and Immunogenicity in Escherichia coli: the Novel Fh8 System. *Front. Microbiol.* 5, 63. doi:10.3389/fmicb.2014.00063

Crick, F. H. C. (1953). The Packing of α-helices: Simple Coiled-Coils. *Acta Crystallogr.* 6, 689–697. doi:10.1107/s0365110x53001964

Cristie-David, A. S., and Marsh, E. N. G. (2019). Metal-dependent Assembly of a Protein Nano-cage. *Protein Sci.* 28, 1620–1629. doi:10.1002/pro.3676

Dasari, M., Espargaro, A., Sabate, R., Lopez del Amo, J. M., Fink, U., Grelle, G., et al. (2011). Bacterial Inclusion Bodies of Alzheimer's Disease β-Amyloid Peptides Can Be Employed to Study Native-like Aggregation Intermediate States. *ChemBioChem* 12, 407–423. doi:10.1002/cbic.201000602

De Groot, N. S., Aviles, F. X., Vendrell, J., and Ventura, S. (2006). Mutagenesis of the central Hydrophobic Cluster in Abeta42 Alzheimer's Peptide. Side-Chain Properties Correlate with Aggregation Propensities. *FEBS J.* 273, 658–668. doi:10.1111/j.1742-4658.2005.05102.x

de Groot, N. S., and Ventura, S. (2006). Effect of Temperature on Protein Quality in Bacterial Inclusion Bodies. *FEBS Lett.* 580, 6471–6476. doi:10.1016/j.febslet.2006.10.071

de Groot, N. S., and Ventura, S. (2010). Protein Aggregation Profile of the Bacterial Cytosol. *PloS One* 5, e9383. doi:10.1371/journal.pone.0009383

De Marco, A., Ferrer-Miralles, N., Garcia-Fruitós, E., Mitraki, A., Peternel, S., Rinas, U., et al. (2019). Bacterial Inclusion Bodies Are Industrially Exploitable Amyloids. *FEMS Microbiol. Rev.* 43, 53–72. doi:10.1093/femsre/fuy038

Díaz-Caballero, M., Navarro, S., and Ventura, S. (2020). Soluble Assemblies in the Fibrillation Pathway of Prion-Inspired Artificial Functional Amyloids Are Highly Cytotoxic. *Biomacromolecules* 21, 2334–2345. doi:10.1021/acs.biomac.0c00271

Diener, M., Kopka, B., Pohl, M., Jaeger, K.-E., and Krauss, U. (2016). Fusion of a Coiled-Coil Domain Facilitates the High-Level Production of Catalytically Active Enzyme Inclusion Bodies. *ChemCatChem* 8, 142–152. doi:10.1002/cctc.201501001

Ebersbach, G., Galli, E., Møller-Jensen, J., Löwe, J., and Gerdes, K. (2008). Novel Coiled-Coil Cell Division Factor ZapB Stimulates Z Ring Assembly and Cell Division. *Mol. Microbiol.* 68, 720–735. doi:10.1111/j.1365-2958.2008.06190.x

García-Fruitós, E., González-Montalbán, N., Morell, M., Vera, A., Ferraz, R., Arís, A., et al. (2005). Aggregation as Bacterial Inclusion Bodies Does Not Imply Inactivation of Enzymes and Fluorescent Proteins. *Microb. Cel Fact* 4, 27. doi:10.1186/1475-2859-4-27

García-Fruitós, E., Rodríguez-Carmona, E., Díez-Gil, C., Ferraz, R. M., Vázquez, E., Corchero, J. L., et al. (2009). Surface Cell Growth Engineering Assisted by a Novel Bacterial Nanomaterial. *Adv. Mater.* 21, 4249–4253. doi:10.1002/adma.200900283

García-Fruitós, E., Seras-Franzoso, J., Vazquez, E., and Villaverde, A. (2010). Tunable Geometry of Bacterial Inclusion Bodies as Substrate Materials for Tissue Engineering. *Nanotechnology* 21, 205101. doi:10.1088/0957-4484/21/20/205101

García-Fruitós, E., Vázquez, E., Díez-Gil, C., Corchero, J. L., Seras-Franzoso, J., Ratera, I., et al. (2012). Bacterial Inclusion Bodies: Making Gold from Waste. *Trends Biotechnol.* 30, 65–70. doi:10.1016/j.tibtech.2011.09.003

Gifre-Renom, L., Cano-Garrido, O., Fàbregas, F., Roca-Pinilla, R., Seras-Franzoso, J., Ferrer-Miralles, N., et al. (2018). A New Approach to Obtain Pure and Active Proteins from Lactococcus Lactis Protein Aggregates. *Sci. Rep.* 8, 13917. doi:10.1038/s41598-018-32213-8

Gil-Garcia, M., Bañó-Polo, M., Varejão, N., Jamroz, M., Kuriata, A., Díaz-Caballero, M., et al. (2018). Combining Structural Aggregation Propensity and Stability Predictions to Redesign Protein Solubility. *Mol. Pharma.* 15, 3846–3859. doi:10.1021/acs.molpharmaceut.8b00341

Gil-Garcia, M., Navarro, S., and Ventura, S. (2020). Coiled-coil Inspired Functional Inclusion Bodies. *Microb. Cel Fact* 19, 1–16. doi:10.1186/s12934-020-01375-4

Jäger, V. D., Kloss, R., Grünberger, A., Seide, S., Hahn, D., Karmainski, T., et al. (2019). Tailoring the Properties of (Catalytically)-active Inclusion Bodies. *Microb. Cel Fact* 18, 1–20. doi:10.1186/s12934-019-1081-5

Jäger, V. D., Lamm, R., Kloß, R., Kaganovitch, E., Grünberger, A., Pohl, M., et al. (2018). A Synthetic Reaction Cascade Implemented by Colocalization of Two Proteins within Catalytically Active Inclusion Bodies. *ACS Synth. Biol.* 7, 2282–2295. doi:10.1021/acssynbio.8b00274

Jäger, V. D., Lamm, R., Küsters, K., Ölçücü, G., Oldiges, M., Jaeger, K.-E., et al. (2020). Catalytically-active Inclusion Bodies for Biotechnology-General Concepts, Optimization, and Application. *Appl. Microbiol. Biotechnol.* 104, 7313–7329. doi:10.1007/s00253-020-10760-3

Jevševar, S., Gaberc-Porekar, V., Fonda, I., Podobnik, B., Grdadolnik, J., and Menart, V. (2005). Production of Nonclassical Inclusion Bodies from Which Correctly Folded Protein Can Be Extracted. *Biotechnol. Prog.* 21, 632–639. doi:10.1021/bp0497839

Jong, W. S. P., Ten Hagen-Jongman, C. M., Vikström, D., Dontje, W., Abdallah, A. M., de Gier, J.-W., et al. (2019). Mutagenesis-Based Characterization and Improvement of a Novel Inclusion Body Tag. *Front. Bioeng. Biotechnol.* 7, 442. doi:10.3389/fbioe.2019.00442

Jong, W. S. P., Vikström, D., Houben, D., van den Berg van Saparoea, H. B., de Gier, J.-W., and Luirink, J. (2017). Application of an E. coli Signal Sequence as a Versatile Inclusion Body Tag. *Microb. Cel Fact* 16, 50. doi:10.1186/s12934-017-0662-4

Kloss, R., Karmainski, T., Jäger, V. D., Hahn, D., Grünberger, A., Baumgart, M., et al. (2018a). Tailor-made Catalytically Active Inclusion Bodies for Different Applications in Biocatalysis. *Catal. Sci. Technol.* 8, 5816–5826. doi:10.1039/c8cy01891j

Kloss, R., Limberg, M. H., Mackfeld, U., Hahn, D., Grünberger, A., Jäger, V. D., et al. (2018b). Catalytically Active Inclusion Bodies of L-Lysine Decarboxylase from E. coli for 1,5-diaminopentane Production. *Sci. Rep.* 8, 1–11. doi:10.1038/s41598-018-24070-2

Koszagova, R., Krajcovic, T., Palencarova-Talafova, K., Patoprsty, V., Vikartovska, A., Pospiskova, K., et al. (2018). Magnetization of Active Inclusion Bodies: Comparison with Centrifugation in Repetitive Biotransformations. *Microb. Cel Fact* 17, 139. doi:10.1186/s12934-018-0987-7

Krauss, U., Jäger, V. D., Diener, M., Pohl, M., and Jaeger, K.-E. (2017). Catalytically-active Inclusion Bodies-Carrier-free Protein Immobilizates for Application in Biotechnology and Biomedicine. *J. Biotechnol.* 258, 136–147. doi:10.1016/j.jbiotec.2017.04.033

Lamm, R., Jäger, V. D., Heyman, B., Berg, C., Cürten, C., Krauss, U., et al. (2020). Detailed Small-Scale Characterization and Scale-Up of Active YFP Inclusion Body Production with Escherichia coli Induced by a Tetrameric Coiled Coil Domain. *J. Biosci. Bioeng.* 129, 730–740. doi:10.1016/j.jbiosc.2020.02.003

Ljubetič, A., Lapenta, F., Gradišar, H., Drobnak, I., Aupič, J., Strmšek, Ž., et al. (2017). Design of Coiled-Coil Protein-Origami Cages that Self-Assemble *In Vitro* and *In Vivo. Nat. Biotechnol.* 35, 1094–1101. doi:10.1038/nbt.3994

Luo, J., Leeman, M., Ballagi, A., Elfwing, A., Su, Z., Janson, J.-C., et al. (2006). Size Characterization of green Fluorescent Protein Inclusion Bodies in E. coli Using Asymmetrical Flow Field-Flow Fractionation-Multi-Angle Light Scattering. *J. Chromatogr. A* 1120, 158–164. doi:10.1016/j.chroma.2005.11.048

Lupas, A. N., and Bassler, J. (2017). Coiled Coils - A Model System for the 21st Century. *Trends Biochem. Sci.* 42, 130–140. doi:10.1016/j.tibs.2016.10.007

Lupas, A. N., Bassler, J., and Dunin-Horkawicz, S. (2017). "The Structure and Topology of α-Helical Coiled Coils," in *Fibrous Proteins: Structures and Mechanisms.* Editors D. A. D. Parry and J. M. Squire (Cham: Springer International Publishing), 95–129. doi:10.1007/978-3-319-49674-0_4

Maji, S. K., Perrin, M. H., Sawaya, M. R., Jessberger, S., Vadodaria, K., Rissman, R. A., et al. (2009). Functional Amyloids as Natural Storage of Peptide Hormones in Pituitary Secretory Granules. *Science* 325, 328–332. doi:10.1126/science.1173155

Margreiter, G., Messner, P., Caldwell, K. D., and Bayer, K. (2008). Size Characterization of Inclusion Bodies by Sedimentation Field-Flow Fractionation. *J. Biotechnol.* 138, 67–73. doi:10.1016/j.jbiotec.2008.07.1995

Mohamad, N. R., Marzuki, N. H. C., Buang, N. A., Huyop, F., and Wahab, R. A. (2015). An Overview of Technologies for Immobilization of Enzymes and Surface Analysis Techniques for Immobilized Enzymes. *Biotechnol. Biotechnol. Equip.* 29, 205–220. doi:10.1080/13102818.2015.1008192

Morell, M., Bravo, R., Espargaró, A., Sisquella, X., Avilés, F. X., Fernàndez-Busquets, X., et al. (2008). Inclusion Bodies: Specificity in Their Aggregation Process and Amyloid-like Structure. *Biochim. Biophys. Acta Mol. Cel Res.* 1783, 1815–1825. doi:10.1016/j.bbamcr.2008.06.007

Morris, C., Glennie, S. J., Lam, H. S., Baum, H. E., Kandage, D., Williams, N. A., et al. (2019). A Modular Vaccine Platform Combining Self-Assembled Peptide Cages and Immunogenic Peptides. *Adv. Funct. Mater.* 29, 1807357. doi:10.1002/adfm.201807357

Nahalka, J., and Nidetzky, B. (2007). Fusion to a Pull-Down Domain: a Novel Approach of producingTrigonopsis variabilisD-Amino Acid Oxidase as Insoluble Enzyme Aggregates. *Biotechnol. Bioeng.* 97, 454–461. doi:10.1002/bit.21244

Park, S.-Y., Park, S.-H., and Choi, S.-K. (2012). Active Inclusion Body Formation Using Paenibacillus Polymyxa PoxB as a Fusion Partner in Escherichia coli. *Anal. Biochem.* 426, 63–65. doi:10.1016/j.ab.2012.04.002

Pesarrodona, M., Jauset, T., Díaz-Riascos, Z. V., Sánchez-Chardi, A., Beaulieu, M. E., Seras-Franzoso, J., et al. (2019). Targeting Antitumoral Proteins to Breast Cancer by Local Administration of Functional Inclusion Bodies. *Adv. Sci.* 6, 1900849. doi:10.1002/advs.201900849

Peternel, Š., Grdadolnik, J., Gaberc-Porekar, V., and Komel, R. (2008). Engineering Inclusion Bodies for Non Denaturing Extraction of Functional Proteins. *Microb. Cel Fact* 7, 34. doi:10.1186/1475-2859-7-34

Roberts, C. J. (2014). Therapeutic Protein Aggregation: Mechanisms, Design, and Control. *Trends Biotechnol.* 32, 372–380. doi:10.1016/j.tibtech.2014.05.005

Roca-Pinilla, R., Fortuna, S., Natalello, A., Sánchez-Chardi, A., Ami, D., Arís, A., et al. (2020a). Exploring the Use of Leucine Zippers for the Generation of a New Class of Inclusion Bodies for Pharma and Biotechnological Applications. *Microb. Cel Fact* 19, 175. doi:10.1186/s12934-020-01425-x

Roca-Pinilla, R., López-Cano, A., Saubi, C., Garcia-Fruitós, E., and Arís, A. (2020b). A New Generation of Recombinant Polypeptides Combines Multiple Protein Domains for Effective Antimicrobial Activity. *Microb. Cel Fact* 19, 122. doi:10.1186/s12934-020-01380-7

Rueda, F., Cano-Garrido, O., Mamat, U., Wilke, K., Seras-Franzoso, J., García-Fruitós, E., et al. (2014). Production of Functional Inclusion Bodies in Endotoxin-free Escherichia coli. *Appl. Microbiol. Biotechnol.* 98, 9229–9238. doi:10.1007/s00253-014-6008-9

Rueda, F., Gasser, B., Sánchez-Chardi, A., Roldán, M., Villegas, S., Puxbaum, V., et al. (2016). Functional Inclusion Bodies Produced in the Yeast Pichia pastoris. *Microb. Cel Fact* 15, 166. doi:10.1186/s12934-016-0565-9

Sanchez de Groot, N., Gomes, R. A., Villar-Pique, A., Babu, M. M., Coelho, A. V., and Ventura, S. (2015). Proteome Response at the Edge of Protein Aggregation. *Open Biol.* 5, 140221. doi:10.1098/rsob.140221

Sánchez, J. M., López-Laguna, H., Álamo, P., Serna, N., Sánchez-Chardi, A., Nolan, V., et al. (2020). Artificial Inclusion Bodies for Clinical Development. *Adv. Sci.* 7, 1902420. doi:10.1002/advs.201902420

Sanchez-Garcia, L., Martín, L., Mangues, R., Ferrer-Miralles, N., Vázquez, E., and Villaverde, A. (2016). Recombinant Pharmaceuticals from Microbial Cells: a 2015 Update. *Microb. Cel Fact* 15, 33. doi:10.1186/s12934-016-0437-3

Schetters, S. T. T., Jong, W. S. P., Kruijssen, L. J. W., van den Berg van Saparoeavan den, H. B., Engels, S., Unger, W. W. J., et al. (2020). Bacterial Inclusion Bodies Function as Vehicles for Dendritic Cell-Mediated T Cell Responses. *Cell Mol. Immunol.* 17, 415–417. doi:10.1038/s41423-019-0298-x

Seras-Franzoso, J., Díez-Gil, C., Vazquez, E., García-Fruitós, E., Cubarsi, R., Ratera, I., et al. (2012). Bioadhesiveness and Efficient Mechanotransduction Stimuli Synergistically provided by Bacterial Inclusion Bodies as Scaffolds for Tissue Engineering. *Nanomedicine* 7, 79–93. doi:10.2217/nnm.11.83

Seras-Franzoso, J., Peebo, K., García-Fruitós, E., Vázquez, E., Rinas, U., and Villaverde, A. (2014). Improving Protein Delivery of Fibroblast Growth Factor-2 from Bacterial Inclusion Bodies Used as Cell Culture Substrates. *Acta Biomater.* 10, 1354–1359. doi:10.1016/j.actbio.2013.12.021

Seras-Franzoso, J., Peebo, K., Luis Corchero, J., Tsimbouri, P. M., Unzueta, U., Rinas, U., et al. (2013). A Nanostructured Bacterial Bioscaffold for the Sustained Bottom-Up Delivery of Protein Drugs. *Nanomedicine* 8, 1587–1599. doi:10.2217/nnm.12.188

Serna, N., Cano-Garrido, O., Sánchez, J. M., Sánchez-Chardi, A., Sánchez-García, L., López-Laguna, H., et al. (2020). Release of Functional Fibroblast Growth Factor-2 from Artificial Inclusion Bodies. *J. Control. Release* 327, 61–69. doi:10.1016/j.jconrel.2020.08.007

Singh, A., Upadhyay, V., Upadhyay, A. K., Singh, S. M., and Panda, A. K. (2015). Protein Recovery from Inclusion Bodies of Escherichia coli Using Mild Solubilization Process. *Microb. Cel Fact* 14, 41. doi:10.1186/s12934-015-0222-8

Singhvi, P., Saneja, A., Srichandan, S., and Panda, A. K. (2020). Bacterial Inclusion Bodies: A Treasure Trove of Bioactive Proteins. *Trends Biotechnol.* 38, 474–486. doi:10.1016/j.tibtech.2019.12.011

Stetefeld, J., Kammerer, R. A., Jenny, M., Schulthess, T., Landwehr, R., and Engel, J. (2000). Crystal Structure of a Naturally Occurring Parallel Right-Handed Coiled Coil Tetramer. *Nat. Struct. Biol.* 7, 772–776. doi:10.1038/79006

Tartaglia, G. G., Pechmann, S., Dobson, C. M., and Vendruscolo, M. (2007). Life on the Edge: a Link between Gene Expression Levels and Aggregation Rates of Human Proteins. *Trends Biochem. Sci.* 32, 204–206. doi:10.1016/j.tibs.2007.03.005

Tatkiewicz, W. I., Seras-Franzoso, J., García-Fruitós, E., Vazquez, E., Ventosa, N., Peebo, K., et al. (2013). Two-Dimensional Microscale Engineering of Protein-Based Nanoparticles for Cell Guidance. *ACS Nano* 7, 4774–4784. doi:10.1021/nn400907f

Thwaite, R., Ji, J., Torrealba, D., Coll, J., Sabés, M., Villaverde, A., et al. (2018). Protein Nanoparticles Made of Recombinant Viral Antigens: A Promising Biomaterial for Oral Delivery of Fish Prophylactics. *Front. Immunol.* 9, 1652. doi:10.3389/fimmu.2018.01652

Tokatlidis, K., Dhurjati, P., Millet, J., Béguin, P., and Aubert, J.-P. (1991). High Activity of Inclusion Bodies Formed in Escherichia coli Overproducing Clostridium Thermocellum Endoglucanase D. *FEBS Lett.* 282, 205–208. doi:10.1016/0014-5793(91)80478-l

Torrealba, D., Parra, D., Seras-Franzoso, J., Vallejos-Vidal, E., Yero, D., Gibert, I., et al. (2016a). Nanostructured Recombinant Cytokines: A Highly Stable Alternative to Short-Lived Prophylactics. *Biomaterials* 107, 102–114. doi:10.1016/j.biomaterials.2016.08.043

Torrealba, D., Seras-Franzoso, J., Mamat, U., Wilke, K., Villaverde, A., Roher, N., et al. (2016b). Complex Particulate Biomaterials as Immunostimulant-Delivery Platforms. *PLoS one* 11, e0164073. doi:10.1371/journal.pone.0164073

Truebestein, L., and Leonard, T. A. (2016). Coiled-coils: The Long and Short of it. *BioEssays* 38, 903–916. doi:10.1002/bies.201600062

Unzueta, U., Cespedes, M. V., Sala, R., Alamo, P., Sánchez-Chardi, A., Pesarrodona, M., et al. (2018). Release of Targeted Protein Nanoparticles from Functional Bacterial Amyloids: A Death star-like Approach. *J. Controlled Release* 279, 29–39. doi:10.1016/j.jconrel.2018.04.004

Utterström, J., Naeimipour, S., Selegård, R., and Aili, D. (2021). Coiled Coil-Based Therapeutics and Drug Delivery Systems. *Adv. Drug Deliv. Rev.* 170, 26–43. doi:10.1016/j.addr.2020.12.012

Vázquez, E., Corchero, J. L., Burgueño, J. F., Seras-Franzoso, J., Kosoy, A., Bosser, R., et al. (2012). Functional Inclusion Bodies Produced in Bacteria as Naturally Occurring Nanopills for Advanced Cell Therapies. *Adv. Mater.* 24, 1742–1747. doi:10.1002/adma.201104330

Vecchi, G., Sormanni, P., Mannini, B., Vandelli, A., Tartaglia, G. G., Dobson, C. M., et al. (2020). Proteome-wide Observation of the Phenomenon of Life on the Edge of Solubility. *Proc. Natl. Acad. Sci. USA* 117, 1015–1020. doi:10.1073/pnas.1910444117

Ventura, S., and Villaverde, A. (2006). Protein Quality in Bacterial Inclusion Bodies. *Trends Biotechnol.* 24, 179–185. doi:10.1016/j.tibtech.2006.02.007

Villar-Piqué, A., Espargaró, A., Sabaté, R., de Groot, N. S., and Ventura, S. (2012). Using Bacterial Inclusion Bodies to Screen for Amyloid Aggregation Inhibitors. *Microb. Cel Fact* 11, 55. doi:10.1186/1475-2859-11-55

Villaverde, A., Corchero, J. L., Seras-Franzoso, J., and Garcia-Fruitós, E. (2015). Functional Protein Aggregates: Just the Tip of the Iceberg. *Nanomedicine* 10, 2881–2891. doi:10.2217/nnm.15.125

Villaverde, A., and Mar Carrió, M. (2003). Protein Aggregation in Recombinant Bacteria: Biological Role of Inclusion Bodies. *Biotechnol. Lett.* 25, 1385–1395. doi:10.1023/a:1025024104862

Walsh, G. (2018). Biopharmaceutical Benchmarks 2018. *Nat. Biotechnol.* 36, 1136–1145. doi:10.1038/nbt.4305

Wang, L., Maji, S. K., Sawaya, M. R., Eisenberg, D., and Riek, R. (2008). Bacterial Inclusion Bodies Contain Amyloid-like Structure. *Plos Biol.* 6, e195. doi:10.1371/journal.pbio.0060195

Wang, X., Zhou, B., Hu, W., Zhao, Q., and Lin, Z. (2015). Formation of Active Inclusion Bodies Induced by Hydrophobic Self-Assembling Peptide GFIL8. *Microb. Cel Fact* 14, 88. doi:10.1186/s12934-015-0270-0

Worrall, D. M., and Goss, N. H. (1989). The Formation of Biologically Active Beta-Galactosidase Inclusion Bodies in Escherichia coli. *Aust. J. Biotechnol.* 3, 28–32.

Wu, W., Xing, L., Zhou, B., and Lin, Z. (2011). Active Protein Aggregates Induced by Terminally Attached Self-Assembling Peptide ELK16 in Escherichia coli. *Microb. Cel Fact* 10, 9. doi:10.1186/1475-2859-10-9

Wu, Y., and Collier, J. H. (2017). α-Helical Coiled-Coil Peptide Materials for Biomedical Applications. *WIREs Nanomed. Nanobiotechnol.* 9, e1424. doi:10.1002/wnan.1424

Xu, C., Liu, R., Mehta, A. K., Guerrero-Ferreira, R. C., Wright, E. R., Dunin-Horkawicz, S., et al. (2013). Rational Design of Helical Nanotubes from Self-Assembly of Coiled-Coil Lock Washers. *J. Am. Chem. Soc.* 135, 15565–15578. doi:10.1021/ja4074529

CHAPTER 2

Combining Structural Aggregation Propensity and Stability Predictions to Redesign Protein Solubility

Marcos Gil-Garcia, Manuel Bañó-Polo, Nathalia Varejão, Michal Jamroz, Aleksander Kuriata, Marta Díaz-Caballero, Jara Lascorz, Bertrand Morel, Susanna Navarro, David Reverter, Sebastian Kmiecik and Salvador Ventura

Combining Structural Aggregation Propensity and Stability Predictions To Redesign Protein Solubility

Marcos Gil-Garcia,[†] Manuel Bañó-Polo,[†] Nathalia Varejão,[†] Michal Jamroz,[‡] Aleksander Kuriata,[‡] Marta Díaz-Caballero,[†] Jara Lascorz,[†] Bertrand Morel,[§] Susanna Navarro,[†] David Reverter,[†] Sebastian Kmiecik,[‡] and Salvador Ventura*[,†]

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

[‡]Biological and Chemical Research Centre, Faculty of Chemistry, University of Warsaw, 00-927 Warsaw, Poland

[§]Departamento de Química Física e Instituto de Biotecnología, Facultad de Ciencias, Universidad de Granada, 18071 Granada, Spain

S *Supporting Information*

ABSTRACT: The aggregation propensity of each particular protein seems to be shaped by evolution according to its natural abundance in the cell. The production and downstream processing of recombinant polypeptides implies attaining concentrations that are orders of magnitude above their natural levels, often resulting in their aggregation; a phenomenon that precludes the marketing of many globular proteins for biomedical or biotechnological applications. Therefore, there is a huge interest in methods aimed to increase the proteins solubility above their natural limits. Here, we demonstrate that an updated version of our AGGRESCAN 3D structural aggregation predictor, that now takes into account protein stability, allows for designing mutations at specific positions in the structure that improve the solubility of proteins without compromising their conformation. Using this approach, we have designed a highly soluble variant of the green fluorescent protein and a human single-domain VH antibody displaying significantly reduced aggregation propensity. Overall, our data indicate that the solubility of unrelated proteins can be easily tuned by in silico-designed nondestabilizing amino acid changes at their surfaces.

KEYWORDS: protein aggregation, protein stability, protein structure, green fluorescent protein, Aβ peptide, single-domain antibodies
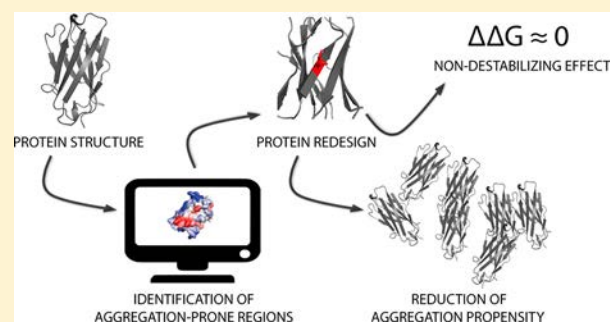
■ INTRODUCTION

Proteins are molecules involved in a wide range of molecular functions. To perform their functions, globular proteins have to acquire a defined 3D structure, skipping misfolding and side aggregation reactions.[1] Protein aggregation involves the establishment of intermolecular interactions, and therefore, it is exquisitely sensitive to the protein concentration.[2] Indeed, it appears that the aggregation propensities of natural proteins correlate with their abundance in the cell.[3] This suggests that proteins have been shaped to remain soluble at those concentrations required to sustain function under physiological conditions, according to the "living on the edge" hypothesis.[4] Thus, many proteins might be only marginally soluble above their natural levels.[5]

Despite their evident economic interest, there are challenging barriers in the production and manufacturing of therapeutic protein-based drugs that preclude the arrival of many of these molecules to the market.[6] Among the various degradation reactions that a protein might experience during its biotechnological production, aggregation is the most common.[7] This is not surprising, since therapeutic applications require protein concentrations that are several orders of magnitude beyond the cellular levels. Aggregation can occur at every phase of protein-based drug development, from recombinant expression to packaging and storage of the final product and administration.[8] Aggregation leads to a decrease in protein activity and might elicit an immunological response, having also regulatory implications.[9] Thus, pharmaceutical and biotechnological industries spend large efforts to prevent the eventual aggregation of their protein products,[10] in many cases using trial and error assays with different buffers, salts, or additives to increase solubility. These approaches are costly and time-consuming, and the adequate conditions for increased solubility may never be found.

In principle, one can modify the amino acid sequence of a globular protein in order to increase its solubility. However, this is not a straightforward task, mainly because the physicochem-

ical properties that allow a globular protein to fold into a stable structure overlap with those triggering aggregation.[11] Indeed, the most aggregation-prone regions in a globular protein are usually found at the hydrophobic core, whose correct packing is a crucial determinant of native structures.[12] This makes it difficult to decrease a protein's aggregation propensity without impacting at the same time its stability and conformation.[11,13]

On the one hand, mutations that decrease the thermodynamic and/or kinetic stability of a globular protein often facilitate aggregation, favoring the population of partially folded states, which expose the solvent hydrophobic regions that are otherwise protected in the native folded conformation.[11,13,14] On the other hand, proteins can also aggregate from the native or native-like conformations without the need of extensive destabilization and unfolding;[15,16] this is especially true when these polypeptides are very concentrated, as it occurs in most protein biopharmaceutics formulations.[10] In these cases, the key to redesign the protein solubility consists in identifying specific positions in protein structures for which mutations that uncouple aggregation from thermodynamic stability can be introduced. Novel algorithms able to analyze the aggregation propensity of folded protein structures constitute privileged tools to identify such potential amino acid substitutions in a cost- and time-effective manner.[9,17−19]

In general, mutations at the protein surface are better tolerated than the residue changes in the protein interior[20] and, accordingly, surface residues evolve faster than core ones.[21] Surface mutations are more frequent because of their weaker destabilizing effects.[22] In addition, in folded globular proteins, the solubility is strongly influenced by the identity of surface residues.[23] Thus, the identification of point mutations that decrease surface aggregation propensity without destabilizing the overall protein structure appears as a promising avenue to optimize protein solubility. We have recently developed AGGRESCAN 3D (A3D), a structure-based aggregation prediction tool that seems optimal to implement such an approach, since it already predicts the aggregation properties of globular protein surfaces.[18] Here, we coupled these predictions with calculations of protein stability changes upon mutation by incorporating stability predictions using the FoldX empirical force field[24] in the A3D pipeline. The value of this combined prediction scheme is experimentally demonstrated by redesigning the solubility of two unrelated protein folds of biotechnological interest.

■ EXPERIMENTAL SECTION

In Silico Analysis. A3D predictions and virtual mutations were made with the algorithm default parameters: static mode and a distance of aggregation analysis of 10 Å. The "Total Score" value results from the sum of the A3D scores of all individual residues. The "Average Score" results from normalizing the "Total Score" relative to the number of total amino acids in the protein.[25] For both scores, the more negative the value, the higher the solubility. We used as inputs the folding reporter green fluorescent protein (GFP) (fr-GFP) X-ray structure (PDB: 2B3Q) and a 3D model generated with SwissModel[26] (based on PDB: 5I19) in the case of the DP47 antibody. The impact of mutations on the protein stability was initially analyzed using the FoldX force field.[24] Protein structures and their surfaces were represented with Pymol (DeLano Scientific, LLC). The prediction of protein stability was implemented in the A3D web server,[18] and it is automatically invoked when the user introduces mutation(s). A3D calculates the stability for the input and mutated structure (using FoldX[24]) and reports the energy difference between the input and mutated protein in kcal/mol. Therefore, negative and positive values indicate stabilization and destabilization, respectively, with respect to the input structure. The report appears under the 'Project Details" tab of the A3D web server[18] (see example output in Figure S9).

Protein Engineering, Expression, and Purification. fr-GFP[27] and Aβ42-GFP[28] were cloned in pET21a and pET28a vectors (Addgene, USA), respectively. Mutations were introduced using the Quickchange mutagenesis kit (Stratagene, USA). *E. coli* BL21 (DE3) competent cells were transformed with the correspondent plasmids.

The DP47 VH antibody[29] was cloned in pET12a (Novagen, USA), mutagenesis was performed as above, and *E. coli* BL21-Gold competent cells (Stratagene, USA) were transformed with correspondent plasmids.

For the production of fr-GFP and Aβ42-GFP and their variants, transformed *E. coli* BL21 (DE3) cells were grown aerobically in LB medium supplemented with 50 μg/mL of kanamycin and 100 μg/mL of ampicillin, respectively. Protein expression was induced with 1 mM IPTG, and cells were grown at 25 °C for 12 h. For protein purification, cells expressing fr-GFP, GFP/KKK, and GFP/DDD were harvested by centrifugation (13 000 rpm for 15 min). After cell lysis by sonication and fractionation, the soluble fraction was collected and injected in a Histrap FF 5 mL column (GE Healthcare, USA) using an ÄKTA (GE Healthcare, USA). The proteins were repurified by size-exclusion chromatography in a HiLoad 26/600 Superdex 75 pg column (GE Healthcare, USA), and native buffer (50 mM Tris, 100 mM NaCl pH 8.4) was used as a mobile phase. The purity of the proteins was checked by 15% SDS-PAGE.

Antibody fragments were expressed using a periplasmic expression system in *E. coli* BL21-Gold at 30 °C in 2XTY medium supplemented with 100 μg/mL of ampicillin and tetracycline. Protein expression was induced with 1 mM IPTG for 48 h at 30 °C. Filtered supernatant was added to the protein A resin column (GE Healthcare, USA). The proteins were eluted with PBS pH 3.2 and neutralized with PBS pH 9.0 to a pH of 7.4. The purity of the proteins was checked by 18% SDS-PAGE.

GFP Fluorescence. Excitation and emission spectra of fr-GFP, GFP/KKK, and GFP/DDD were analyzed in a Jasco FP-8200 fluorescence spectrofluorometer (Jasco Corporation, Japan) using 0.5 mg/mL of protein in the native buffer at pH 8. Three spectra were accumulated at 25 °C, with slit widths of 5 nm, a 0.5 nm interval, and a 1000 nm/min scan rate. Emission spectra were obtained recording the emission between 500 and 600 nm. Excitation spectra were obtained by exciting in a 450−600 nm range. Fluorescence of fr-GFP and GFP/KKK purified proteins was recorded using the excitation and emission wavelengths of 485 and 515 nm, respectively.

For thermal denaturation, GFP fluorescence was recorded in a range of 25−90 °C with an increasing heat rate of 1 °C/min, slit widths of 5 nm, a 0.5 nm interval, and a 1000 nm/min scan rate. Proteins at 10 μM were excited at 485 nm, and emission was recorded at 515 nm each 0.5 °C.

For chemical denaturation, fr-GFP, GFP/KKK, and GFP/DDD proteins were mixed with increasing concentrations of guanidinium hydrochloride, from 0 to 6 M, keeping a constant protein concentration of 10 μM. Samples were left to stand for 48 h at 25 °C in order to equilibrate. Then, fluorescence was analyzed by exciting at 485 nm and recording the emission in a

500−600 nm range, with a bandwidth of 1 nm. Experiments were performed in duplicate.

For refolding kinetics, protein samples at 0.5 mg/mL in the native buffer pH 8 were unfolded at 95 °C for 5 min and allowed to refold at 25 °C for 30 min. The recovery of GFP fluorescence was monitored by exciting at 485 nm and recording the emission fluorescence at 515 nm every 5 s during 30 min.

For analyzing the dependence of the fluorescence emission on the pH, fr-GFP, GFP/KKK, and GFP/DDD proteins were diluted in different buffers ranging from pH 2 to 13, at a final protein concentration of 10 μM, and incubated for 2 h. Experiments were performed in duplicate.

In Vivo GFP Fluorescence. Induced *E. coli* cells expressing the Aβ42-GFP variants were harvested by centrifugation, washed with phosphate-buffered saline (PBS buffer pH 7.0), and adjusted to an OD of 0.2. GFP fluorescence of intact cells was measured on a Jasco FP-8200 fluorescence spectrofluorometer (Jasco Corporation, Japan), with slit widths of 5 nm, a 0.5 nm interval, and 1000 nm/min scan rate, by exciting the sample at 485 nm and recording the average of the three accumulated emission spectra from 500 to 600 nm. Experiments were performed in triplicate.

Fluorescence Microscopy. Induced Aβ42-GFP variant cultures were centrifuged and washed in PBS buffer, and 10 μL of cells was deposited on top of microscopy glass slides. Images were obtained in a Leica fluorescence microscope (Leica Microsystems, Germany) under UV light with a filter for GFP excitation (450−500 nm) and an emission filter (515−560 nm).

Immunoblotting. Ten microliters of fractioned cell extracts of Aβ42-GFP was loaded on a 15% SDS-PAGE gel containing 0.1% SDS. Proteins were transferred to a polyvinilidene difluoride membrane (PVDF) (MerckMillipore, Germany) with a Mini Trans-Blot Cell (Bio-Rad, USA) using the transfer buffer (0.5 M Tris, 0.04 M glycine, 0.04% SDS, 20% methanol) at 100 V for 1 h. Membranes were blocked in the blocking solution (0.05 M Tris-HCl, pH 7.6, 0.15 NaCl, 0.1% Tween 20, 5% milk) and incubated with 1:1000 anti-GFP antibody (Roche, Switzerland). The blot was then incubated with 1:2000 dilution of goat anti-mouse IgG conjugated with peroxidase for 1 h at room temperature. Detection was carried out using an enhanced chemiluminescence (ECL) SuperSignal kit (Pierce Protein Biology, USA).

Flow Cytometry Analysis. Flow cytometry was performed using a BD FACSCanto flow cytometer (BD Biosciences,USA) software equipped with 488 and 635 nm lasers. Cells were first gated (P1) for forward scatter (FSC) and side scatter (SCC), both set on logarithmic amplification. A threshold was adjusted on FSC to reduce electronic background noise. Cells in P1 were then regated (P2) for the FITC-channel to acquire GFP emission with a 530/30 nm band-pass filter. A total of 20 000 events were acquired. Noninduced cells were used as GFP signal controls. Data were collected with the FACSDiva Software (BD Biosciences, USA), and data analysis was performed with the FlowJo software. Experiments were performed in triplicate.

Protein Crystallization and X-ray Structure Determination. GFP/KKK was crystallized on MRC2 commercial crystallization plates at 18 °C, using the sitting drop vapor diffusion method; mixing 300 nL of protein (22.5 mg/mL) with 300 nL of precipitant (0.1 M MMT buffer, 25% PEG 1500, pH 4.0) (PACT premier HT-96, Molecular Dimensions), which was allowed to equilibrate against a well containing 50 μL of precipitant. Single crystals appeared after 2 weeks. Crystals were

cryo-protected in a reservoir buffer containing 10% glycerol and flash-frozen in liquid nitrogen prior to diffraction analysis.

Diffraction data were collected from cryo-cooled crystals (100 K) on ALBA synchrotron (BL13-XALOC beamline), Barcelona, Spain.[30]

Data were integrated and merged using XDS[31] and scaled, reduced, and further analyzed using CCP4.[32] The structure of GFP/KKK was determined from the X-ray data by molecular replacement using the wtGFP crystal structure (PDB: 1GFL) as a starting model with phaser.phenix. Refinements and model rebuilding of the GFP/KKK crystal structure were performed using PHENIX[33] and Coot.[34] All structure figures were generated using PyMOL (DeLano Scientific, LLC). The data-collection and refinement statistics are summarized in the Table S1. The atomic coordinates and structure factors for the GFP/KKK structure have been deposited in the Protein Data Bank (PDB entry 6FWW).

Circular Dichroism Spectroscopy. Thermal denaturation experiments were performed using a JASCO J-715 spectropolarimeter (Jasco Corporation, Japan) equipped with a Peltier system for temperature control. The protein concentration was 10 μM in PBS pH 7.4. The CD signal at 202 nm was recorded from 25 to 90 °C, with a heating rate of 1 °C/min. Far-UV spectra were recorded before and after heating. Five accumulations were averaged for each sample.

Light Scattering. Aggregation kinetics of GFP and DP47 variants were followed on a Jasco FP-8200 fluorescence spectrofluorometer (Jasco Corporation, Japan) by exciting at 330 nm and recording the orthogonal light scattering signal in a 320−340 nm range, with slit widths of 5 nm, a 0.5 nm interval, and 1000 nm/min scan rate.

For temperature induced aggregation, 0.5 mg/mL of GFP and GFP/KKK in the native buffer, pH 8, were incubated at 80 °C. Measurements were made every 10 min during the 70 min.

For chemical induced aggregation, 0.5 mg/mL of GFP and GFP/KKK were incubated in the presence of 40% TFE at 25 °C. The signal was acquired every 10 min for a total of 90 min in the same conditions as above.

DP47 and DP47/KKK aggregation kinetics were analyzed at 37 °C with an agitation of 300 rpm for 72 h. The protein concentration was 10 μM in PBS at pH 7.4.

Intrinsic Tryptophan Fluorescence. Tryptophan intrinsic fluorescence of DP47 and DP47/KKK was analyzed at 25 °C in a Jasco FP8200 spectrofluorometer (Jasco Corporation, Japan). Three averaged spectra were accumulated by using an excitation wavelength of 280 nm and recording the emission from 300 to 400 nm, with slit widths of 5 nm, a 0.5 nm interval, and 1000 nm/min scan rate.

Electrophoretic Mobility Assay. The EMSA reaction with (FITC)-labeled 50b random ssDNA was prepared with a constant DNA concentration of 0.2 μM, and various concentrations of purified protein in the binding buffer (10 mM HEPES−KOH pH 7.5, 2 mM 2-ME). After a 30 min incubation at 30 °C and the addition of 10% (v/v) glycerol, free DNA and DNA−protein mixtures were resolved by electrophoresis for 90 min at 10 V on 0.5% (w/v) TAE-agarose gels at 4 °C. Gels with FITC-labeled short ssDNA were detected directly on a Versadoc system (BioRad) with excitation at 488 nm.

Size-Exclusion Chromatography. The protein solutions of fr-GFP and GFP/KKK (500 μL) at 10 μM in the native buffer pH 8.0 were loaded onto a Superdex 75 10/300 GL (GE Healthcare, USA) column coupled to an ÄKTA-fast protein liquid chromatography device (GE Healthcare, USA), pre-

viously equilibrated with the native buffer pH 8. Protein elution was monitored by following absorbance at 280 nm.

Differential Scanning Calorimetry (DSC). Samples were dialyzed in 20 mM phosphate buffer, 100 mM NaCl pH 7.95. Sample concentrations were adjusted to 15 μM. Experiments were carried out on a VP-DSC capillary-cell microcalorimeter from MicroCal (Malvern Panalytical Ltd., UK) at a scan rate of 120 °C/h from 5 to 105 °C. Calorimetric cells (operating volume 0.134 mL) were kept under pressure (60 psi) to prevent sample degassing. Several buffer−buffer baselines were obtained before each protein run to ascertain proper equilibration of the instrument. Reheating showed that the unfolding process was irreversible.

■ RESULTS

A3D Description. The A3D algorithm uses the protein 3D structure as an input, which is subsequently energetically minimized using the FoldX force field.[24] Then, an aggregation propensity score is calculated for all the spheres with a 10 Å radius in the protein structure. The variables that contribute to the A3D score are (i) the experimentally derived individual amino acid aggregation propensities,[35] (ii) the surface area exposure of the amino acids in the sphere, and (iii) the effective distance between adjacent residues and the central amino acid in the sphere. Therefore, the A3D score is structurally corrected and, in contrast to sequence-based aggregation predictors, provides information on the structural aggregation propensity (STAP) of globular proteins in their functional folded states.[18]

Rational Design of an Aggregation-Resistant GFP. We initiated our redesign exercise trying to generate a more soluble variant of an already folding-optimized GFP version: the fr-GFP,[27] a folding-enhanced GFP containing the "cycle-3" mutations[36] and the "enhanced GFP" (EGFP) mutations[37] (Figure 1A,B). The cycle-3 mutations F99S, M153T, and V163A reduced the hydrophobicity of the protein, rendering it significantly more resistant to aggregation than the wild-type GFP,[38] whereas the EGFP mutations F64L and S65T modified the chromophore for the simple absorption spectrum and strong fluorescence.[39] fr-GFP was the variant on top of which the superfolder-GFP was evolved upon four rounds of DNA shuffling and paneling.[40]

The structure of fr-GFP (PDB: 2B3Q) was analyzed using the A3D algorithm. Only three aggregation-prone residues were identified in the fr-GFP A3D profile: V11, Y39, and L221 (Figure 1C). In order to obtain a more soluble version of fr-GFP, we exploited the so-called aggregation gatekeeper residues. They correspond to either charged amino acids (Asp, Glu, Arg, or Lys) or Pro, all shown to contribute to modulate the aggregation properties of natural proteins.[3,41,42] The three fr-GFP aggregation-prone residues were virtually and individually mutated by gatekeepers and the theoretical most soluble variants selected according to their average and A3D total scores. Mutations to Asp or Lys were top ranked. Accordingly, variants containing one, two, or three aggregation-prone residues mutated either to Asp or to Lys were analyzed with A3D. Predictions indicated that the successive introduction of charged residues at the fr-GFP surface would potentially abrogate the presence of aggregation-prone exposed patches and accordingly decrease the STAP of the redesigned proteins (Table 1 and Figure 2).

Even when a particular residue substitution at the protein surface is predicted to increase the protein solubility, this positive effect might be canceled in case the mutation impacts
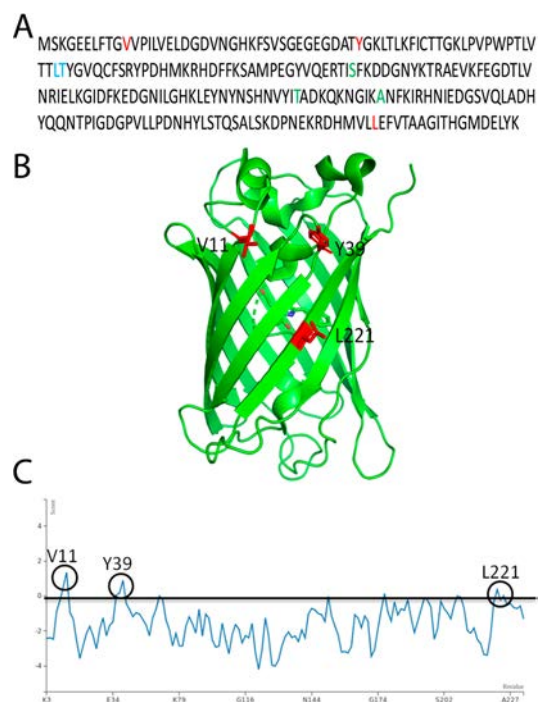
Figure 1. Sequence, structure, and aggregation profile of GFP. (A) Amino acid sequence of fr-GFP. Residues shared with cycle-3 and EGFP variants are indicated in green and blue, respectively. (B) 3D structure of fr-GFP (PDB: 2B3Q). C) Aggregation profile of fr-GFP according to A3D. In (A) and (B), the predicted aggregation-prone residues are indicated in red, in (C), they are circled.

Table 1. A3D and FoldX Analysis of fr-GFP and Its Variants[a]

GFP mutations	name	average score	total score	$\Delta\Delta G$ (kcal/mol)
	fr-GFP	−1.026	−231.814	
V11K	GFP/K	−1.078	−243.631	−0.065
V11K/L221K	GFP/KK	−1.102	−248.993	−0.026
V11K/Y39K/L221K	GFP/KKK	−1.132	−255.922	0.017
V11D	GFP/D	−1.09	−246.239	1.08
V11D/L221D	GFP/DD	−1.107	−250.147	2.555
V11D/Y39D/L221D	GFP/DDD	−1.140	−257.73	3.545

[a]More negative average and total A3D scores indicate higher solubility. Negative and positive $\Delta\Delta G$ values correspond to over- and de-stabilizing mutations, respectively.

the thermodynamic stability, since it would facilitate protein fluctuations and the exposure of aggregation-prone regions previously protected in the structure.[43] This might render the mutant variant even more aggregation-prone than the original protein. We calculated the impact of the designed mutations on the fr-GFP stability using the FoldX force field.[24] FoldX is among the most popular stability predictors and was already implemented in the original version of A3D to minimize the input PDB files. Despite this, computational predictions of stability changes upon mutation are still far from being perfect, FoldX has been described as an accurate stability predictor[44] with a low prediction bias.[45] Moreover, FoldX has already been used to explore the connection between the protein stability and aggregation.[3,46] Mutations to Asp were predicted to be destabilizing ($\Delta\Delta G$ > 1 kcal/mol), whereas mutations to Lys were predicted to be neutral (0.5 > $\Delta\Delta G$ > −0.5 kcal/mol). This
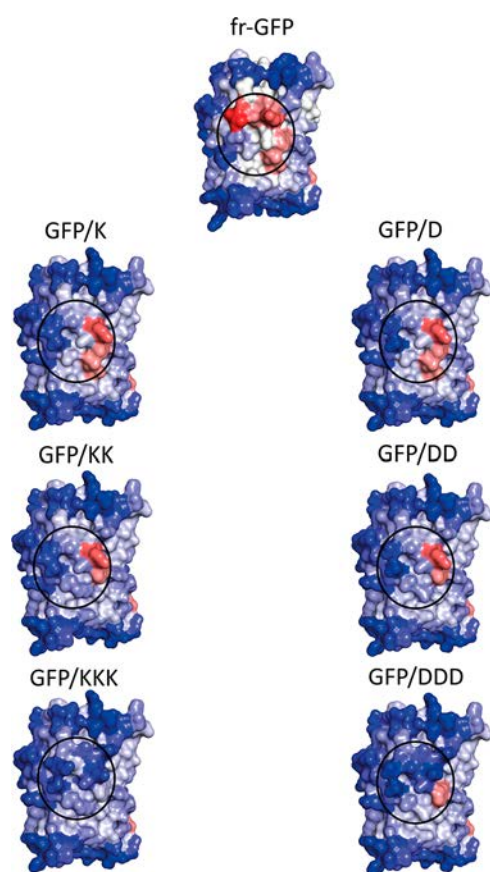
Figure 2. A3D structural predictions for fr-GFP and its variants. 3D structures of fr-GFP and its mutational variant structures. The protein surface is colored according to the A3D score in a gradient from blue (high-predicted solubility) to white (negligible impact on protein aggregation) to red (high-predicted aggregation propensity).

provided us a unique opportunity to evaluate the interplay between the predicted STAP and stability in determining the fr-GFP solubility and, more generally, on the eventual benefit of incorporating stability predictions in our A3D pipeline.

Designed Lysine Mutations Increase Aβ42-GFP Solubility upon Overexpression. The already optimized folding and solubility properties of fr-GFP[27] make the selection of a more robust variant a challenging exercise. We reasoned that this can be achieved by fusing the above six designed mutants with an N-terminal aggregation-prone bait that would favor the aggregation of the fusion protein. To this aim, we used a fusion of the amyloidogenic Aβ Alzheimer's peptide (Aβ42) with fr-GFP (Aβ42-GFP) or its variants. We have used this kind of fusion extensively to analyze the impact of Aβ42 mutations on aggregation-propensity,[28] toxicity,[47] or cellular aging,[48] by mutating specific residues in the Aβ42 moiety, while leaving intact the GFP domain. Independently of the Aβ42 mutation, Aβ42-GFP was shown to accumulate always in the insoluble fraction as inclusion bodies.[28] However, the presence of active GFP in such aggregates differed significantly, displaying a high correlation with the intrinsic aggregation propensity of the Aβ42 mutants.[49] This is due to the existence of an in vivo kinetic competition between the attainment of the GFP functional structure and the aggregation of the fusion protein, directed by the Aβ moiety;[50] highly aggregation-prone Aβ variants resulting in low fluorescence and the other way around.

In this work, we proceeded in the opposite way, testing if mutations in the GFP domain can somehow counteract the high aggregation-propensity of wild-type Aβ42. In this system, it is expected that the more soluble fr-GFP variants would result in higher levels of active GFP upon intracellular overexpression, by decreasing the overall aggregation rate of the protein fusion. Therefore, we measured the GFP fluorescence of intact cells overexpressing Aβ42 fused to fr-GFP (Aβ42-GFP) as well as to our six fr-GFP variants using spectrofluorimetry. As shown in Figure 3A, a single Lys mutation at position 11 (Aβ42-GFP/K) did not suffice to significantly increase GFP emission, relative to cells expressing the original fusion, despite it being predicted to
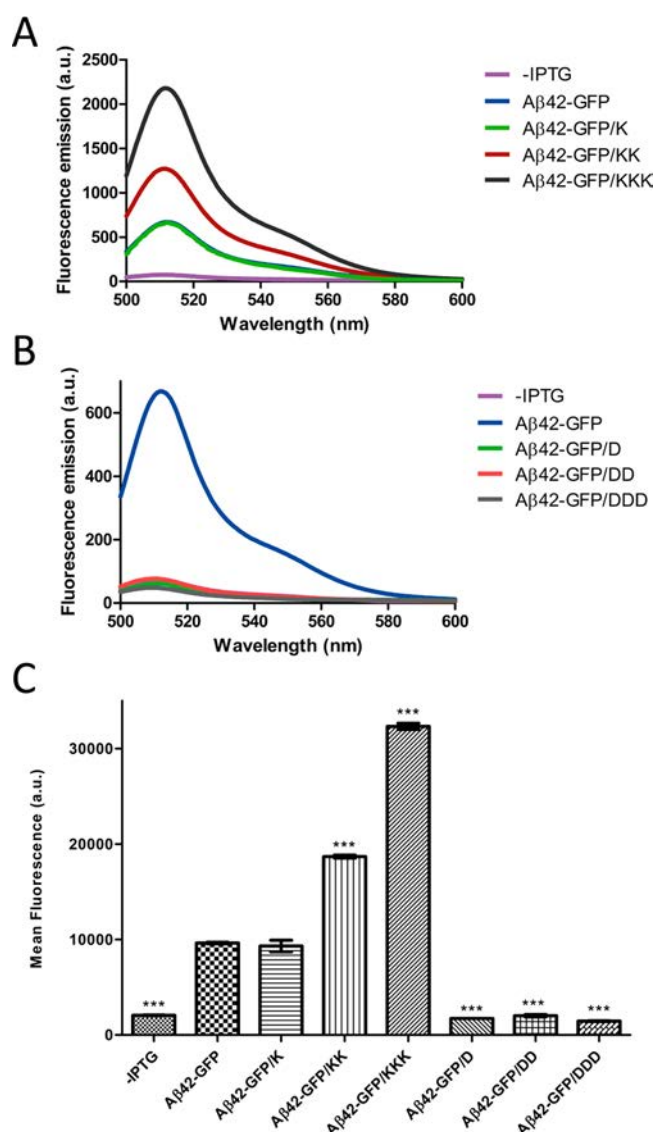
Figure 3. GFP fluorescence of intact cells expressing Aβ42-GFP variants. (A) Quantification of the fluorescence emission of different lysine variants in *E. coli* cells measured by fluorescence spectroscopy. (B) Quantification of the fluorescence emission of different aspartic acid variants in *E. coli* cells measured by fluorescence spectroscopy. (C) Histogram of Aβ42-GFP variants showing the average value of fluorescence intensity as detected by FC. All experiments correspond to triplicates, and the statistical values were obtained using an unpaired student *t*-test. Statistically significant values as compared with Aβ42-GFP are indicated as *p*-value <0.001 (***). Error bars correspond to SEM.
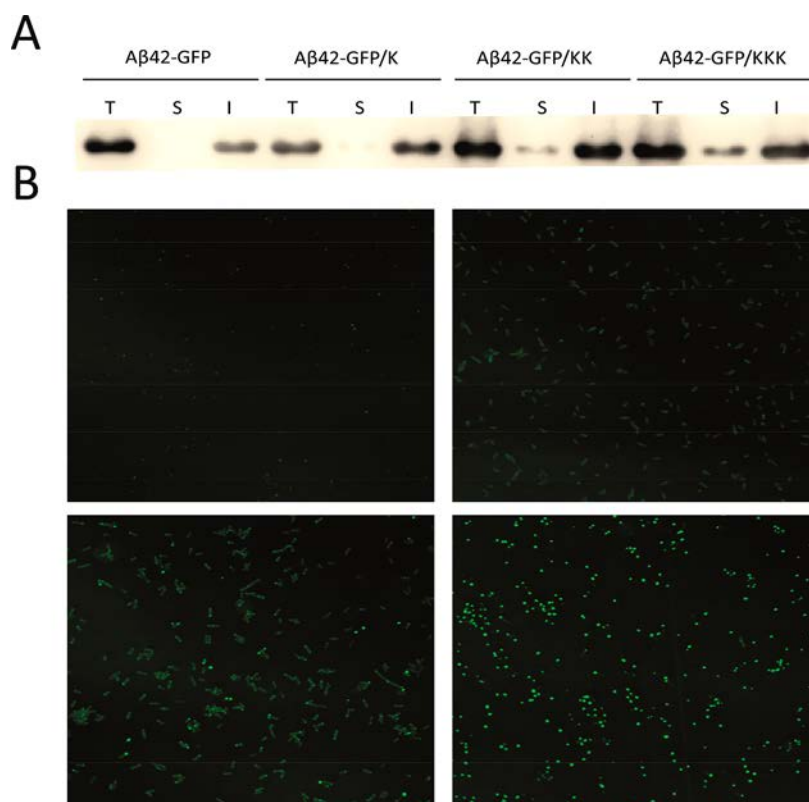
Figure 4. Cellular distribution of Aβ42-GFP fusions. (A) Western blot of Aβ42-GFP in the total (T), soluble (S), and insoluble (I) fractions of *E. coli* cells. (B) Visualization of GFP fluorescence in intact *E. coli* cells expressing Aβ42-GFP (top, left), Aβ42-GFP/K (top, right), Aβ42-GFP/KK (bottom, left), and Aβ42-GFP/KKK (bottom, right).

be the individual Lys mutation with the largest impact on protein solubility. However, the introduction of two and three Lys point mutations in fr-GFP (Aβ42-GFP/KK and Aβ42-GFP/KKK) promoted a significant increase in the fluorescence emitted by the respective cells (Figure 3A). The need for several simultaneous mutations to attain a significant solubilizing effect has been evidenced in different computer-assisted protein solubility redesign exercises.[3,17,42,51] Cells expressing Aβ42-GFP/KKK were 3-fold more fluorescent than those expressing Aβ42-GFP. We also assayed the GFP fluorescence of the different cell populations using flow cytometry (FC) (Figure 3C). The ability of FC to analyze a large number of cells (20.000 cells/sample) averages the experimental variability and provides more quantitative results.[47] The analysis showed that, independently of the considered variant, a large majority of the cells were somehow fluorescent, indicating that they were expressing the correspondent protein fusions (data not shown). The mean fluorescence values obtained for the different variants were in excellent agreement with those obtained by spectro-fluorimetry, confirming that cells expressing Aβ42-GFP/KKK were three times more fluorescent than the ones expressing the original fusion.

The distribution of the recombinant protein fusions between the soluble and insoluble cellular fractions was analyzed by Western blotting (Figure 4A). As previously reported,[48,49] all Aβ42-GFP resided in the insoluble fraction, as was the case for Aβ42-GFP/K. In contrast, despite a majority of Aβ42-GFP/KK and Aβ42-GFP/KKK fusions still being located in the insoluble fraction, increasing amounts of soluble protein were detected for these two variants. It is worth emphasizing here that, in our previous studies, we only observed the Aβ42-GFP fusion in the soluble fraction when at least two highly disruptive mutations were introduced in the Aβ42 moiety.[47]

Finally, we used fluorescence microscopy to identify the cellular location of the detected GFP emission in *E. coli* cells. As expected, the cells become more fluorescent as the number of Lys in fr-GFP increases (Figure 4B). In good agreement with fractionation analysis, the fluorescence was confined mainly in inclusion bodies at the poles of the cells. The high fluorescence of Aβ42-GFP/KKK inclusions indicates that the higher solubility of this fr-GFP variant competes with the aggregation of the Aβ42 peptide, allowing for a higher amount of fluorophore formation before the intracellular aggregation occurs.

Interestingly enough, despite A3D STAP predictions for Asp variants being, overall, equivalent to those obtained for the Lys counterparts (Table 1 and Figure 2), both spectrofluorimetry and FC data indicated that bacterial cells expressing Aβ42-GFP/D, Aβ42-GFP/DD, and Aβ42-GFP/DDD fusion proteins displayed fluorescence intensities equivalent to that of non-induced cultures and thus, significantly lower than that of the induced Aβ42-GFP (Figure 3B,C), even if they were expressed at similar levels to that of the original fusion (Figure S1). This highlighted the important role played by protein stability on effective protein solubility and advised us on the convenience of incorporating stability in our A3D prediction scheme.

GFP/KKK Is Folded and Stable. Once in cell analysis suggested GFP/KKK being the more soluble among the generated variants, we wanted to asses first that, once purified, it was well-folded and stable. To this aim, the mutant GFP domain alone was His-tagged, recombinantly produced, and purified using affinity and size exclusion chromatography. The
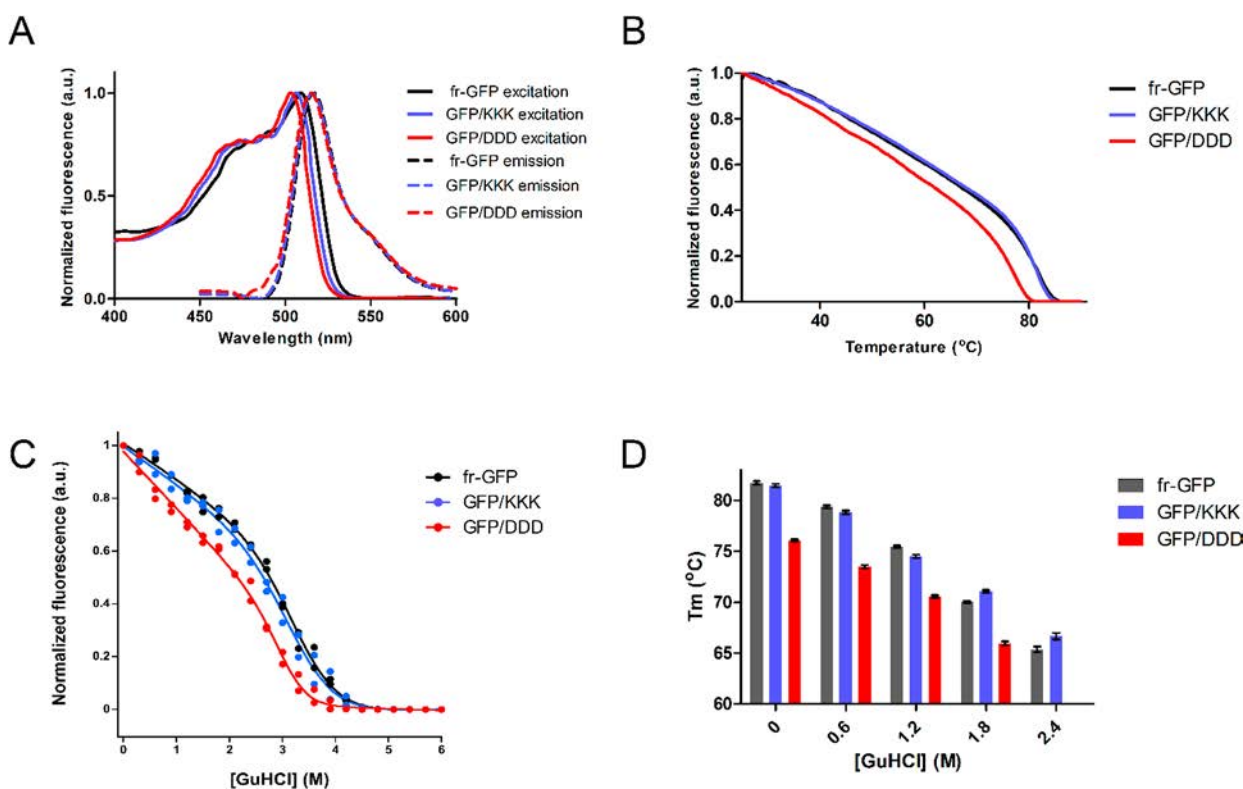
Figure 5. Spectral properties and stability of fr-GFP, GFP/KKK, and GFP/DDD. (A) Normalized excitation and emission spectra of different GFP variants. (B) Thermal unfolding curves followed by GFP fluorescence from 25 to 90 °C with an increasing rate of 1 °C/min. (C) Chemical denaturation of variants at 10 μM with increasing concentrations of GuHCl. Samples were equilibrated for 48 h. (D) Melting temperature values in the presence of increasing GuHCl concentrations. Experiments were performed in duplicate. Error bars correspond to SEM.

same procedure was used for the original fr-GFP protein. We also produced and purified GFP/DDD in order to assess if, as predicted by FoldX, this variant was destabilized relative to both GFP/KKK and fr-GFP. Both GFP/DDD and Aβ42-GFP/DDD exhibited anomalous electrophoretic mobility (Figures S1 and S2), a property already described for other proteins in which mutations to acidic residues have been introduced.[52]

The fluorescence emission spectra of fr-GFP, GFP/KKK, and GFP/DDD were superimposable (Figure 5A). The fluorescence excitation of GFP/KKK and GFP/DDD were slightly blue-shifted relative to fr-GFP, an effect that was also observed for superfolder-GFP.[40] At pH 8.0, the emission fluorescence intensities of GFP/KKK and GFP/DDD at 515 nm were 91 and 87% of that exhibited by fr-GFP, respectively.

In order to confirm that the introduced mutations did not significantly impact protein stability, as predicted, fr-GFP and GFP/KKK were submitted to thermal and chemical denaturation. The thermal stability of the proteins was analyzed at pH 8.0 by monitoring the changes in the GFP fluorescence emission. We obtained cooperative, superimposable denaturation curves with $T_m = 81.72 \pm 0.17$ °C and $T_m = 81.47 \pm 0.16$ °C for fr-GFP and GFP/KKK, respectively (Figure 5B). As predicted by FoldX, the two variants displayed a T_m higher than that of GFP/DDD, with a $T_m = 76.06 \pm 0.10$ °C (Figure 5B). The heat denaturation of all of the variants in these conditions was irreversible, which precluded the calculation of the thermodynamic constants. To confirm that fr-GFP and GFP/KKK differ from GFP/DDD in terms of thermal stability, we performed thermal denaturation assays for the three variants in the presence of increasing concentrations of guanidinium hydrochloride (GuHCl) (Figure S3). As it can be deduced from the calculated

T_ms in the different conditions (Figure 5D), fr-GFP and GFP/KKK display a similar stability, both being more stable than GFP/DDD, for which the denaturation curve in the presence of 2.4 M GuHCl was no longer cooperative. To validate that fr-GFP and GFP/KKK have similar thermal stability, we used DSC (Figure S4). We calculated $T_m = 84.6 \pm 0.1$ °C and $T_m = 84.0 \pm 0.1$ °C for fr-GFP and GFP/KKK, respectively. For GFPs, the calculated T_ms depend significantly on the scan rates;[53] this would explain the difference between the T_ms obtained for both variants in fluorimetry and DSC analysis. In any case, all of the results converge to indicate that, in contrast to GFP/DDD, GFP/KKK is very similar to fr-GFP in terms of thermal stability.

The resistance against chemical denaturation with guanidinium hydrochloride (GuHCl) was also monitored by recording the changes in GFP fluorescence in the presence of increasing concentrations of denaturant after equilibration for 48 h. The obtained curves were cooperative and displayed a midpoint of denaturation of 3.19 ± 0.06, 3.08 ± 0.08, and 2.56 ± 0.10 M (Figure 5C). It is worth noting that the conditions of the assay are of pseudoequilibrium, since the equilibrium of GFP is not attained even after several weeks,[54] and that the reaction is not reversible due to the hysteresis caused by the denaturation of the chromophore during the unfolding reaction,[55] both factors impeding the calculation of thermodynamic parameters.

Overall, as predicted in GFP/KKK, the introduced mutations did not compromise GFP function or stability, whereas the GFP/DDD variant is significantly less stable.

Finally, we calculated the dependence of the chromophore fluorescence on the pH for the three variants (Figure S5). The three proteins remained inactive below pH 5.0. fr-GFP and GFP/DDD exhibited a very similar profile, displaying maximum

fluorescence emission at pH 9.0. GFP/KKK is more fluorescent at alkaline pH, reaching a maximum at pH 11.0. Upon incubation at pH 12.0 and above, all of the proteins become inactive.

GFP/KKK Is Highly Resistant to Aggregation. In order to confirm that the increased solubility of the Aβ42-GFP/KKK inside cells, relative to Aβ42-GFP, can be assigned to a more soluble GFP moiety, we compared the aggregation properties of purified GFP/KKK and fr-GFP.

First, we monitored the kinetics of aggregation for the two variants by following the increase in light scattering when the proteins were heated at 80 °C. The aggregation rate was significantly lower for the redesigned protein (k_1= 0.017 ± 0.005 min^{-1}) than for fr-GFP (k_1= 0.030 ± 0.009 min^{-1}), as well as the scattering at the end of the reaction (Figure 6A). Next, we tested
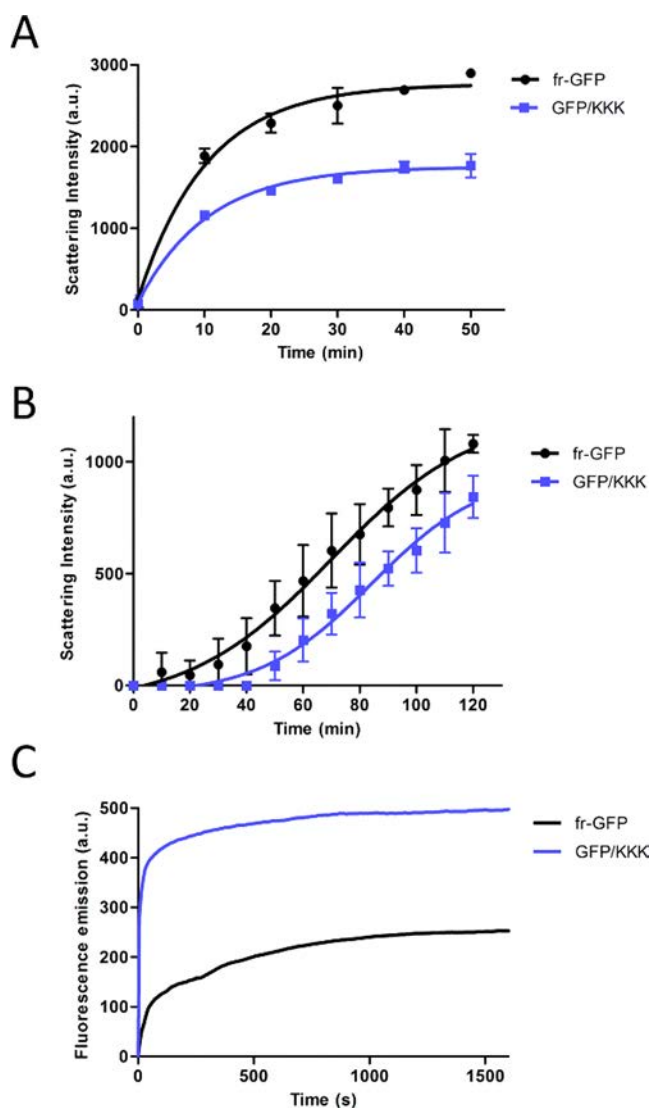
Figure 6. Comparison of the aggregation properties of fr-GFP and GFP/KKK. (A) Evolution of the light scattering signal upon incubation at 80 °C. (B) Evolution of the light scattering signal upon incubation in 40% TFE. Both experiments were performed in duplicate. (C) Refolding kinetics of both proteins at 25 °C after thermal unfolding at 95 °C, as measured by monitoring the recovery of GFP fluorescence emission. All the experiments were performed at a 0.5 mg/mL concentration in the native buffer pH 8.0 and in duplicate. Error bars correspond to SEM.

the proteins resistance to aggregation in the presence of 40% of TFE, a strongly denaturing condition known to promote the aggregation of β-sheet rich proteins.[23,56] We recorded the light scattering signal at different time points after the addition of the cosolvent. As shown in Figure 6B, fr-GFP already started to aggregate after 10 min, whereas for GFP/KKK, the detection of the initial aggregates occurred only after 50 min, exhibiting lower scattering intensity than the original protein at any analyzed time point.

Finally, we analyzed the ability of both proteins to refold into the native state upon thermal denaturation. fr-GFP and GFP/KKK were incubated at 95 °C for 5 min and allowed to refold at 25 °C, and the recovery of GFP fluorescence along time was monitored (Figure 6C). The refolding reactions were adjusted to double-exponential functions ($R \geq 0.99$). GFP/KKK folds more than 3-fold faster (k_1= 0.175 ± 0.004 min^{-1}) than fr-GFP (k_1= 0.045 ± 0.001 min^{-1}). In addition, GFP/KKK recovers two times more fluorescence than fr-GFP at the end of the reaction.

Overall, the data in this section demonstrate how the combined prediction of STAP and the stability allowed us to design a GFP variant with significantly improved solubility and folding properties by mutating only three residues at the protein surface. This result is remarkable if we take into account that fr-GFP was already designed to have optimal properties as a reporter of protein folding.

X-ray Structure of GFP/KKK. To investigate if the inserted mutations in GFP/KKK impact the 3D structure of GFP, we solved the crystal structure of this mutant at 1.13 Å resolution (PDB: 6FWW). The side chains of the newly introduced Lys residues are well-defined in the electron density maps, all being exposed to the solvent (Figure 7A). The chromophore formed from residue T65-Y66-G67 (annotated as CRO in the PDB file) is correctly formed and buried in the middle of the central helix (Figure 7A,B). The Cα backbone trace of wild-type GFP (1GFL) and GFP/KKK could be superimposed with a root-mean-square deviation (rmsd) of 0.52 Å (Figure 7C), which is comparable to the rmsd of 0.50 Å between the fr-GFP and GFP/KKK. Thus, as expected, the three-surface mutations did not change the overall GFP fold, indeed no evident structural rearrangements were observed around the mutated regions (Figure 7D). An analysis of the GFP/KKK structure with A3D confirms that, in contrast to fr-GFP, this variant lacks any surface exposed aggregation-prone region (Figures 7E,F).

Although purified fr-GFP and GFP/KKK proteins were monomeric (Figure S6), it has been reported that fr-GFP crystallizes as an antiparallel dimer (PDB:2B3Q), similar to other reported GFP crystal structures, likely because GFP has a certain tendency to dimerize in solution.[40] In contrast, GFP/KKK crystallizes as a monomer. Using the dimeric structure of fr-GFP, we generated a model of a GFP/KKK dimer. The L221K residue of each monomer protrudes into the symmetry-related monomer facing the equivalent L221K residue, potentially generating repulsion and hindering dimerization (Figure S7). Calculation of the impact of this single mutation on the stability of the dimeric interface with FoldX[24] confirms that it is highly disruptive ($\Delta\Delta G$ = 4.96 kcal/mol). We found out that Lippincott-Schwartz and co-workers introduced exactly the same single mutation in GFP to generate a variant that could not dimerize.[57,58] This effect might contribute to the higher solubility of this re-engineered variant. In addition, in the crystal structure of GFP/KKK, the side chains of the other two mutated residues, Lys 11 and Lys 39, are at less than a 5 Å distance (Table S2), resulting in a local clustering of positive charges, which will
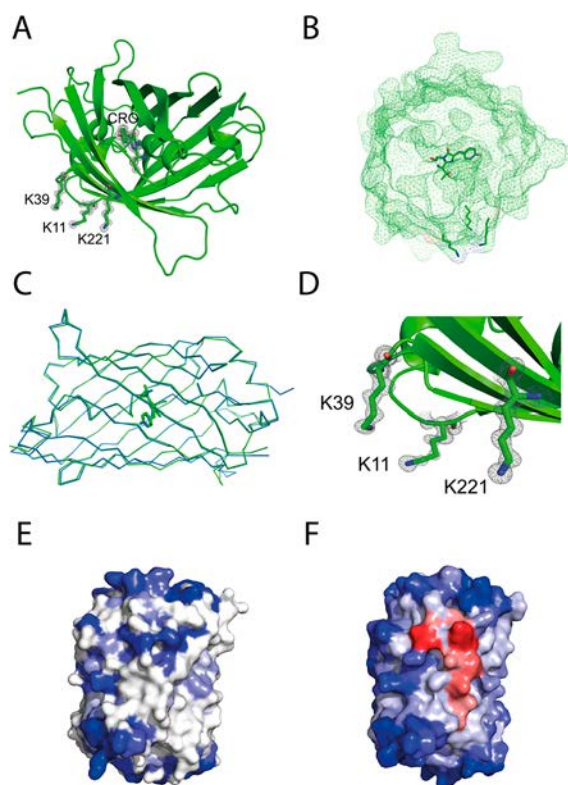
Figure 7. 3D structure and A3D structural prediction of GFP/KKK. (A) Crystal structure of GFP/KKK (PDB: 6FWW) mutant showing the differential map (2mFo-Fc contoured at 1.2σ) surrounding the mutated lysine residues 11, 39, and 221 and the fluorescent cromophore, which are depicted in mesh and stick. (B) Surface representation of the superposed GFP structures showing the perfect overlay of the fluorescent chromophore (blue and green sticks for wtGFP and GFP/KKK mutant, respectively). (C) Superposition of wtGFP (blue, PDB: 1GFL) and GFP/KKK (green) monomers. The overall rmsd for the 226 superposed Cα atoms is 0.52 Å with 96.9% identity. (D) Zoom of GFP/KKK structure around the mutated lysines. (E, F) A3D of GFP/KKK and fr-GFP crystal structures, respectively. The protein surface is colored according to the A3D score in a gradient from blue (high-predicted solubility) to white (negligible impact on protein aggregation) to red (high-predicted aggregation propensity).

likely cause the electrostatic repulsion between the GFP/KKK molecules.

It has been shown that a positively supercharged version of GFP binds strongly to nucleic acids, which compromises its use in a cellular context.[59] The electrophoretic mobility shift assays showed that increasing concentrations of fr-GFP or GFP/KKK did not reduce the mobility of a random FITC-labeled 50 base long ssDNA in the native agarose gel, even when the protein concentration exceeds 50 molar times over the DNA (Figure S8). This result indicates that the insertion of the three mutated Lys is not sufficient to promote the binding of GFP/KKK to DNA.

Incorporating Stability Predictions in A3D. The above-described data clearly indicated the convenience of skipping the experimental validation of mutations that, despite the optimizing STAP, negatively impacted the protein stability. To this aim, we implemented the FoldX empirical force field[24] in the A3D web server for automated protein stability calculations if a user introduces any mutation(s) (see the In Silico Analysis section). An example of the output is presented in Figure S9B. We wanted to prevent A3D users from producing redesigned

proteins that ultimately will aggregate at the same or higher levels than the original protein only because they are destabilized. Therefore, we took a conservative criterion for structural integrity, in which a mutation resulting in a $\Delta\Delta G \geq 1$ kcal/mol is considered disruptive. Indeed, all fr-GFP Asp mutants overpass this destabilization limit. Accordingly, a warning message is now displayed in the "Project Details" screen of A3D when sequence changes are predicted to destabilize the protein ≥ 1 kcal/mol (Figure S9C).

Rational Design of a VH Antibody with Increased Aggregation Resistance. We decided to test the novel version of A3D in the redesign of a therapeutically relevant scaffold. There is a trend in the industry toward the use of small antibody formats for imaging and tumor targeting purposes. They include human single chain fragments (scFv) and human single domain antibodies.[29] However, the biophysical properties of these human proteins are poor, and their aggregation propensity is even higher than the one of complete antibodies, due to the lack of interdomain stabilization. Here we tried to improve the solubility of a variable heavy (VH) segment of the human antibody germline (DP47).[29] DP47 VH is a monomer consisting of 3 α-helices, 10 β-strands, 1 disulfide bond, and 3 complementarity-determining regions (CDRs) essential for antigen binding (Figure 8A,B). DP47 VH was selected because
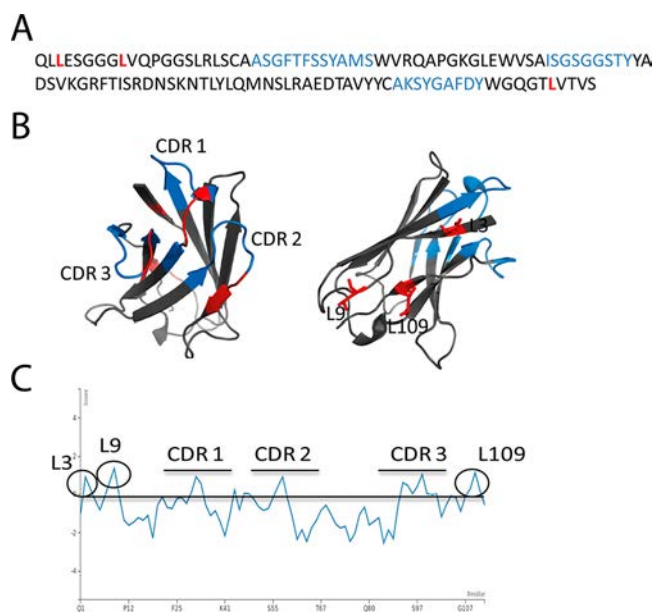
Figure 8. Sequence, structure, and aggregation profile of DP47. (A) Amino acid sequence of DP47. (B) 3D structure of DP47 model obtained from SwissModel. (C) Aggregation profile of DP47 model according to A3D. In (A) and (B), the predicted aggregation-prone residues are indicated in red and CDR residues in blue, in (C), they are circled and underlined, respectively.

the evolution of more soluble variants of this specific antibody has been previously addressed using phage display.[29] In this work, CDR3 was excluded from mutation to limit effects on the antigen binding. Despite 6 mutations that increased the solubility of the molecule being identified, all of them were clustered at CDR1, while mutations at other positions had little effect.[60] This was expected, because CDRs tend to be sticky in order to recognize antigens. Nevertheless, mutations at CDRs can significantly compromise binding and, accordingly, they are

not a preferred strategy to generate more soluble antibody variants.

Here, we aimed at generating a DP47 variant with increased aggregation resistance without changing any of its CDRs residues. No structure is available for wild-type DP47, and the version with mutated CDRs was crystallized as a domain-swapped dimer (PDB: 3UPC).[29] Therefore, no suitable DP47 monomeric structure was available for direct A3D analysis. Accordingly, we created a homology model on top of sequentially related available structures (PDB:5I19) using Swiss-Model.[25] This model has a GMQE (global model quality estimation) of 0.99 and a QMEAN Z-score of 0.6 (Figure 8B). As expected, A3D predicted a high aggregation propensity for the three DP47 CDRs. Apart from them, only three other residues were predicted as aggregation-prone, Leu3 and Leu9 at the N-terminus and Leu109 at the C-terminus (Figure 8C). These residues were virtually mutated by the different gatekeeper residues using the new A3D algorithm, and again, a triple Lys mutant (DP47/KKK) appeared as the version that best combined the predicted solubility and stability (Table 2 and Figure 9), whereas an equivalent DP47/DDD version was again predicted to be destabilized (Table S3).

Table 2. A3D Analysis of DP47 and Its Variants[a]

DP47 protein	average score	total score	$\Delta\Delta G$ (kcal/mol)
DP47	−0.48	−54.248	
DP47/K	−0.6021	−68.0402	−0.5196
DP47/KK	−0.6993	−79.0154	−0.014
DP47/KKK	−0.758	−85.681	−0.427

[a]More negative averages and total A3D scores indicate higher solubility. Negative $\Delta\Delta G$ values correspond to over-stabilizing mutations.
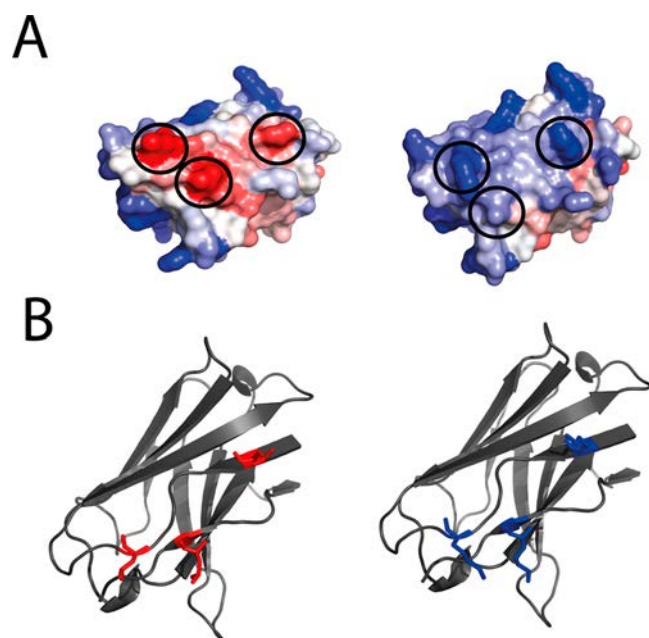
Figure 9. A3D structural predictions for DP47 and its variant. 3D Structures of DP47 (left) and its mutational variant structure (right). (A) The protein surface is colored according to the A3D score in a gradient from blue (high-predicted solubility) to white (negligible impact on protein aggregation) to red (high-predicted aggregation propensity). (B) Ribbon representation of the two models indicating the wild-type Leu residues in red and the mutated Lys residues in blue.

DP47/KKK Is Folded and Stable. Wild-type DP47 and DP47/KKK were recombinantly expressed in *E. coli* and purified to homogeneity from the extracellular medium by affinity on a protein A column. Then we compared the biophysical properties of both single-domain antibodies.

The circular dichroism (CD) spectra of both proteins showed the characteristic β-sheet minimum at 218 nm and overlapping, indicating that they essentially shared the same secondary structure content (Figure 10A). DP47 contains three Trp
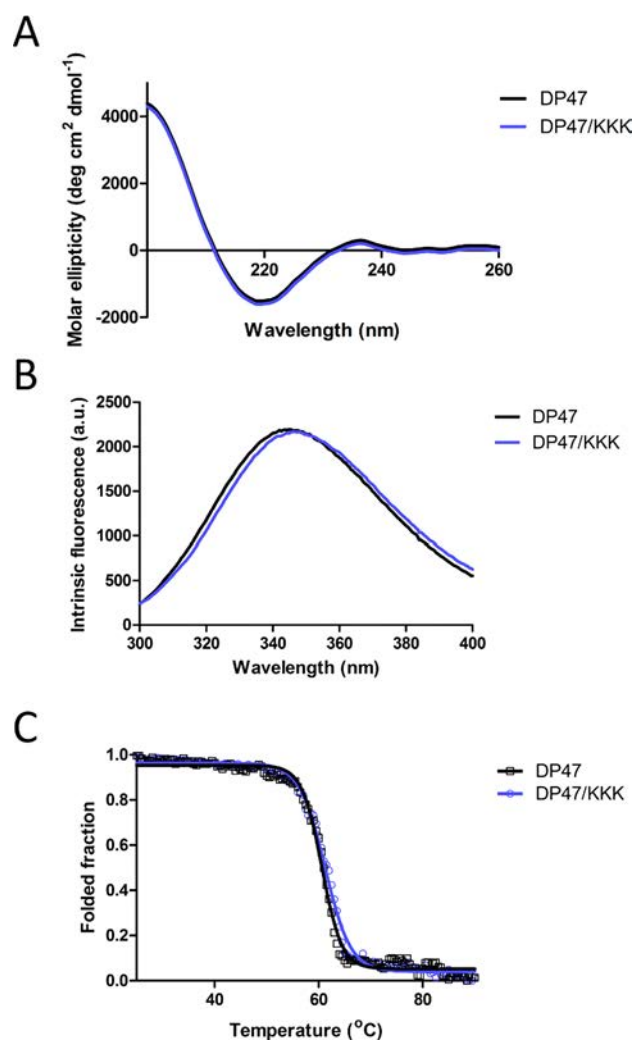
Figure 10. Biophysical characterization of DP47 variants. (A) Far-UV CD spectra with a minimum value at 218 nm. (B) Tryptophan intrinsic fluorescence. (C) Thermal unfolding curves followed by Far-UV from 25 to 90 °C with a heating rate of 1 °C/min. The experiments were performed at 10 μM protein in PBS pH 7.4.

residues at positions 34, 45, and 104, with W34 being inside the structure, and W45 and W104 being exposed to the solvent. Its intrinsic fluorescence spectra in the folded state exhibits a maximum at 345.5 nm. The Trp intrinsic fluorescence spectra of DP47/KKK shows the same intensity as that of DP47, although the maximum is slightly red-shifted (maximum at 346.5 nm) (Figure 10B), suggesting that, despite the two antibodies sharing a similar globular tertiary structure, they can exhibit certain local conformational differences.

Finally, in order to confirm that, as predicted, the introduced mutations did not significantly impact the antibody stability

(predicted $\Delta\Delta G$ = −0.43 kcal/mol), DP47 and DP47/KKK were submitted to thermal denaturation. The thermal stability of the proteins was analyzed by monitoring the changes in the CD signal at 202 nm. Both proteins cooperatively unfolded, displaying a very similar T_m of 60.0 ± 0.1 and 60.4 ± 0.1 °C for DP47 and DP47/KKK, respectively (Figure 10C). The reaction was not reversible, and thermodynamic constants could not be calculated.

DP47/KKK Displays Increased Aggregation Resistance. To address if, as intended, the three Leu to Lys mutations increased the resistance of the antibody against aggregation in close to physiological conditions, DP47 and DP47/KKK were incubated in PBS at pH 7.4. and 37 °C for 3 days, and we monitored the changes in light scattering. Figure 11 illustrates
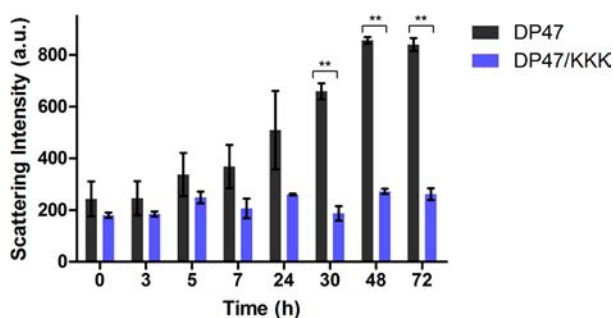
Figure 11. Comparison of the aggregation properties of DP47 variants. The two antibodies were incubated at 37 °C with an agitation of 300 rpm for 72 h. The protein concentration was 10 μM in PBS at pH 7.4. The light scattering signal was recorded along time. Statistical values were obtained using an unpaired student *t*-test. Statistically significant values are indicated as *p*-value < 0.01 (**). Error bars correspond to SEM.

how the scattering of the DP47 solution steadily increases with time reaching a maximum after 2 days. In contrast, the scattering of the DP47/KKK solution remains constant and is lower than that of the wild-type protein at any time point. Despite the final scattering signal attained in these conditions being low compared with the one obtained after heat denaturation, the differences between the proteins are statistically significant, indicating that the redesign has succeeded in reducing the inherent high aggregation propensity of this particular antibody.

DISCUSSION

Protein aggregation remains a major factor limiting the biotechnological and therapeutic use of many proteins.[10] When globular proteins are employed for such applications, they should remain soluble and active in artificial conditions for which they have not been optimized by evolutionary selection. Therefore, it is not surprising that they tend to aggregate, since they are expressed and manufactured at concentrations that are far beyond their natural limits and stored for periods of time that are much longer than their cellular lifetimes, both factors favoring the establishment of intermolecular interactions that can nucleate the aggregation reaction, especially in the absence of the protective mechanisms provided in vivo by the protein quality surveillance machinery. This implies that the use of proteins above their cellular concentrations would require certain sequential adaptations; however, it is not known what the margin of improvement is, if any, for such an approach.[61]

Protein aggregation not only affects protein production but also the downstream processes, like purification, storage,

administration, and activity.[10] In this context, effective computational approaches able to rank protein therapeutics according to their solubility and improve this particular property upstream in the industrial pipeline hold the promise to result in significant savings in time and money, since, in principle, they can sift the properties of hundreds to thousands protein variants before performing any experiment.

Redesigning globular proteins to reduce their natural aggregation propensity is, however, not an easy task, since the more potent aggregation-promoting regions lie at the hydrophobic core, and it becomes difficult to mutate them without disrupting the protein structure and function.[3] Because mutations at the surface are, in general, less disruptive, an alternative is to mutate those exposed residues that contribute significantly to the aggregation properties of globular structures. The identification of such an amino acid requires a novel generation of aggregation-prediction algorithms that work on top of protein structures and not on protein sequences. SAP, CamSol, and A3D are examples of these kinds of programs.[9,17,18] An important feature of these computational tools is that they allow for the detection of aggregating regions that are relevant only in the native state. These regions are generally formed by residues that, despite being scattered in the sequence, come together in the structure upon folding to form an aggregation-prone surface. Indeed, the three aggregation-prone residues predicted by A3D and mutated here in both fr-GFP and DP47 lie in the same face of the protein, despite some of them being N- and others C-terminal in the respective sequences. Nevertheless, in their present versions, none of these programs take into account the protein stability, and thus they do not allow for the assessment of the impact of the designed mutations in the protein structural integrity. This is an important issue, because mutations that negatively impact the thermodynamic and/or kinetic stability would likely facilitate aggregation, increasing the exposure of the hydrophobic regions to the solvent, irrespective of whether they are predicted to increase or decrease the protein surface aggregation propensity in the native state. In fact, all these programs predict the fr-GFP Lys and Asp mutational series to be similarly solubilizing, whereas, in contrast, our results indicate that this is not the case. This stresses the advantage of integrating protein stability predictions in A3D, since it allows for further experimental validation of potentially destabilized or not-folding protein variants to be discarded.

Solubis is a recent program that combines aggregation and stability predictions with remarkable success.[3,19,42] However, in contrast to A3D, it looks for contiguous aggregation-prone residues in the sequence and not in the structure, implementing the TANGO sequential prediction algorithm.[61] It aims to introduce solubilizing mutations in the middle of these sequence stretches without disturbing the protein structure. V11 and Y39 residues in fr-GFP do not lie inside an aggregation-prone region (APR) according to TANGO. L221 maps in a C-terminal APR, but TANGO predicts that its mutation to Lys would increase the aggregation potential of the sequence (Figure S10). Therefore, the gain in solubility observed for GFP/KKK cannot be predicted by Solubis. In the case of DP47, only Leu109 lies in a TANGO-predicted APR, comprising residues 107-GTLVTV-112 (Figure S10), illustrating again how, despite in certain cases sequential and structural APRs might overlap, this is not generally the case. Thus, the new version of A3D incorporates features that are not combined in any pre-existent algorithm.

Our data demonstrate that there is room to increase the solubility of globular proteins by introducing a small number of

mutations that reduce the STAP while preserving the structure. GFP/KKK is clearly more-aggregation resistant than the folding reporter version of the fluorescent protein,[27] being able to provide intracellular solubility to the highly aggregation-prone Aβ42 peptide. GFP/KKK only shares with superfolder-GFP a mutation at position 39 (Y39N), indicating that experimental direct evolution and computational redesign can arrive at different solutions in order to optimize protein solubility. This is best illustrated in the case of antibodies, where the overlap between the regions responsible for epitope binding and those accounting for their aggregation potential critically compromise their redesign. In fact, for DP47, phage display screening could only identify solubilizing mutations mapping in the CDRs, without any other surface exposed residue impacting aggregation.[29] In contrast, A3D detected three of these residues, whose mutations to Lys effectively increased the antibody aggregation-resistance, allowing for the avoidance of the modification of amino acids involved in antigen recognition.

As in our designs, the aggregation of several proteins has been shown to be reduced by introducing charged residues.[3,42,60] It has been argued that this effect results from an increase in colloidal stability by increasing the net charge of the protein, and therefore that it is rather unspecific.[62] However, the introduction of three Lys residues in fr-GFP reduces the net charge of the protein from −7 to −4, whereas in the nonfunctional GFP/DDD variant, the net charge increases to −10. This suggests that, in our study, the observed impact on the solubility results from structural gatekeeping and preservation of the protein conformation and not from the mere change in the physicochemical characteristics of the protein sequence.

Overall, we developed an improved prediction tool (accessible at http://biocomp.chem.uw.edu.pl/A3D/) that should allow for the solubility to be redesigned through protein engineering without impacting the native structure and stability of the target protein. Despite the functionality of the redesigned human antibody still needing to be confirmed, with an increasing number of protein-based therapeutics under development,[63] this algorithm might become a powerful tool to assist the design of well folded and soluble versions of these drugs, and it might compete or, at least, complement experimental screening efforts.

■ ASSOCIATED CONTENT

ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.molpharmaceut.8b00341.

> X-ray data collection and refinement statistics spectra; GFP Asp mutants' expression gels; distances between residues in crystal structures; model of GFP/KKK dimer; EMSA assay; screenshots of A3D predictions; table of predicted A3D scores for Asp mutants of DP47; and TANGO predictions for fr-GFP, DP47, and KKK mutants (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*Salvador Ventura Tel: 93 58689 56; Fax: 93 5812011, E- mail: salvador.ventura@uab.cat.

ORCID Ⓞ

Marta Díaz-Caballero: 0000-0001-9032-9819
Salvador Ventura: 0000-0002-9652-6351

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Chiti, F.; Dobson, C. M. Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade. *Annu. Rev. Biochem.* **2017**, *86* (1), 27−68.

(2) Ellis, R. J. Macromolecular Crowding: Obvious but Underappreciated. *Trends Biochem. Sci.* **2001**, *26* (10), 597−604.

(3) Ganesan, A.; Siekierska, A.; Beerten, J.; Brams, M.; Van Durme, J.; De Baets, G.; Van der Kant, R.; Gallardo, R.; Ramakers, M.; Langenberg, T.; Wilkinson, H.; De Smet, F.; Ulens, C.; Rousseau, F.; Schymkowitz, J. Structural Hot Spots for the Solubility of Globular Proteins. *Nat. Commun.* **2016**, *7*, 10816.

(4) Tartaglia, G. G.; Pechmann, S.; Dobson, C. M.; Vendruscolo, M. Life on the edge: A link between gene expression levels and aggregation rates of human proteins. *Trends Biochem. Sci.* **2007**, *32* (5), 204−206.

(5) Ciryam, P.; Kundra, R.; Morimoto, R. I.; Dobson, C. M.; Vendruscolo, M. Supersaturation Is a Major Driving Force for Protein Aggregation in Neurodegenerative Diseases. *Trends Pharmacol. Sci.* **2015**, *36* (2), 72−77.

(6) Roberts, C. J. Therapeutic Protein Aggregation: Mechanisms, Design, and Control. *Trends Biotechnol.* **2014**, *32* (7), 372−380.

(7) Hamrang, Z.; Rattray, N. J. W.; Pluen, A. Proteins Behaving Badly: Emerging Technologies in Profiling Biopharmaceutical Aggregation. *Trends Biotechnol.* **2013**, *31* (8), 448−458.

(8) Wang, W.; Nema, S.; Teagarden, D. Protein Aggregation-Pathways and Influencing Factors. *Int. J. Pharm.* **2010**, *390* (2), 89−99.

(9) Chennamsetty, N.; Voynov, V.; Kayser, V.; Helk, B.; Trout, B. L. Design of Therapeutic Proteins with Enhanced Stability. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (29), 11937−11942.

(10) Roberts, C. J. Protein Aggregation and Its Impact on Product Quality. *Curr. Opin. Biotechnol.* **2014**, *30*, 211−217.

(11) Espargaró, A.; Castillo, V.; de Groot, N. S.; Ventura, S. The in Vivo and in Vitro Aggregation Properties of Globular Proteins Correlate With Their Conformational Stability: The SH3 Case. *J. Mol. Biol.* **2008**, *378* (5), 1116−1131.

(12) Ventura, S.; Vega, M. C.; Lacroix, E.; Angrand, I.; Spagnolo, L.; Serrano, L. Conformational Strain in the Hydrophobic Core and Its Implications for Protein Folding and Design. *Nat. Struct. Biol.* **2002**, *9* (6), 485−493.

(13) Castillo, V.; Espargaró, A.; Gordo, V.; Vendrell, J.; Ventura, S. Deciphering the Role of the Thermodynamic and Kinetic Stabilities of SH3 Domains on Their Aggregation inside Bacteria. *Proteomics* **2010**, *10* (23), 4172−4185.

(14) Jahn, T. R.; Radford, S. E. The Yin and Yang of Protein Folding. *FEBS J.* **2005**, *272* (23), 5962−5970.

(15) Bemporad, F.; Chiti, F. Native-like Aggregation" of the Acylphosphatase from Sulfolobus Solfataricus and Its Biological Implications. *FEBS Lett.* **2009**, *583* (16), 2630−2638.

(16) Garcia-Pardo, J.; Graña-Montes, R.; Fernandez-Mendez, M.; Ruyra, A.; Roher, N.; Aviles, F. X.; Lorenzo, J.; Ventura, S. Amyloid Formation by Human Carboxypeptidase D Transthyretin-like Domain under Physiological Conditions. *J. Biol. Chem.* **2014**, *289* (49), 33783−33796.

(17) Sormanni, P.; Aprile, F. a.; Vendruscolo, M. The CamSol Method of Rational Design of Protein Mutants with Enhanced Solubility. *J. Mol. Biol.* **2015**, *427* (2), 478−490.

(18) Zambrano, R.; Jamroz, M.; Szczasiuk, A.; Pujols, J.; Kmiecik, S.; Ventura, S. AGGRESCAN3D (A3D): Server for Prediction of Aggregation Properties of Protein Structures. *Nucleic Acids Res.* **2015**, *43* (W1), W306−W313.

(19) De Baets, G.; Van Durme, J.; Van Der Kant, R.; Schymkowitz, J.; Rousseau, F. Solubis: Optimize Your Protein. *Bioinformatics* **2015**, *31* (15), 2580−2582.

(20) Franzosa, E. a.; Xia, Y. Structural Determinants of Protein Evolution Are Context-Sensitive at the Residue Level. *Mol. Biol. Evol.* **2009**, *26* (10), 2387−2395.

(21) Sasidharan, R.; Chothia, C. The Selection of Acceptable Protein Mutations. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (24), 10080−10085.

(22) Tokuriki, N.; Stricher, F.; Schymkowitz, J.; Serrano, L.; Tawfik, D. S. The Stability Effects of Protein Mutations Appear to Be Universally Distributed. *J. Mol. Biol.* **2007**, *369* (5), 1318−1332.

(23) Lawrence, M. S.; Phillips, K. J.; Liu, D. R. Supercharging Proteins Can Impart Unusual Resilience. *J. Am. Chem. Soc.* **2007**, *129* (33), 10110−10112.

(24) Schymkowitz, J.; Borg, J.; Stricher, F.; Nys, R.; Rousseau, F.; Serrano, L. The FoldX Web Server: An Online Force Field. *Nucleic Acids Res.* **2005**, *33*, W382−W388.

(25) Pujols, J.; Pena-Diaz, S.; Ventura, S. AGGRESCAN3D: Toward the Prediction of the Aggregation Propensities of Protein Structures. *Methods Mol. Biol.* **2018**, *1762*, 427−443.

(26) Schwede, T.; Kopp, J.; Guex, N.; Peitsch, M. C. SWISS-MODEL: An Automated Protein Homology-Modeling Server. *Nucleic Acids Res.* **2003**, *31* (13), 3381−3385.

(27) Waldo, G. S.; Standish, B. M.; Berendzen, J.; Terwilliger, T. C. Rapid Protein-Folding Assay Using Green Fluorescent Protein. *Nat. Biotechnol.* **1999**, *17* (7), 691−695.

(28) De Groot, N. S.; Aviles, F. X.; Vendrell, J.; Ventura, S. Mutagenesis of the Central Hydrophobic Cluster in Aβ42 Alzheimer's Peptide: Side-Chain Properties Correlate with Aggregation Propensities. *FEBS J.* **2006**, *273* (3), 658−668.

(29) Dudgeon, K.; Rouet, R.; Kokmeijer, I.; Schofield, P.; Stolp, J.; Langley, D.; Stock, D.; Christ, D. General Strategy for the Generation of Human Antibody Variable Domains with Increased Aggregation Resistance. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (27), 10879−10884.

(30) Juanhuix, J.; Gil-Ortiz, F.; Cuní, G.; Colldelram, C.; Nicolás, J.; Lidón, J.; Boter, E.; Ruget, C.; Ferrer, S.; Benach, J. Developments in Optics and Performance at BL13-XALOC, the Macromolecular Crystallography Beamline at the Alba Synchrotron. *J. Synchrotron Radiat.* **2014**, *21* (4), 679−689.

(31) Kabsch, W. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66* (2), 125−132.

(32) Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G. W.; McCoy, A.; McNicholas, S. J.; Murshudov, G. N.; Pannu, N. S.; Potterton, E. A.; Powell, H. R.; Read, R. J.; Vagin, A.; Wilson, K. S. Overview of the CCP4 Suite and Current Developments. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2011**, *67* (4), 235−242.

(33) Adams, P. D.; Afonine, P. V.; Bunkóczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L. W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. PHENIX: A Comprehensive Python-Based System for Macromolecular Structure Solution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66* (2), 213−221.

(34) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and Development of Coot. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66* (4), 486−501.

(35) 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 Bioinf.* **2007**, *8* (1), 65.

(36) Crameri, A.; Whitehorn, E. a.; Tate, E.; Stemmer, W. P. C. Improved Green Fluorescent Protein by Molecular Evolution Using DNA Shuffling. *Nat. Biotechnol.* **1996**, *14* (3), 315−319.

(37) Patterson, G. H.; Knobel, S. M.; Sharif, W. D.; Kain, S. R.; Piston, D. W. Use of the Green Fluorescent Protein and Its Mutants in Quantitative Fluorescence Microscopy. *Biophys. J.* **1997**, *73* (5), 2782−2790.

(38) Fukuda, H.; Arai, M.; Kuwajima, K. Folding of Green Fluorescent Protein and the Cycle3Mutant. *Biochemistry* **2000**, *39* (39), 12025−12032.

(39) Yang, T. T.; Cheng, L.; Kain, S. R. Optimized Codon Usage and Chromophore Mutations Provide Enhanced Sensitivity with the Green Fluorescent Protein. *Nucleic Acids Res.* **1996**, *24* (22), 4592−4593.

(40) Pédelacq, J.-D.; Cabantous, S.; Tran, T.; Terwilliger, T. C.; Waldo, G. S. Engineering and Characterization of a Superfolder Green Fluorescent Protein. *Nat. Biotechnol.* **2006**, *24* (1), 79−88.

(41) Beerten, J.; Schymkowitz, J.; Rousseau, F. Aggregation Prone Regions and Gatekeeping Residues in Protein Sequences. *Curr. Top. Med. Chem.* **2013**, *12* (22), 2470−2478.

(42) Van der Kant, R.; Karow-Zwick, A. R.; Van Durme, J.; Blech, M.; Gallardo, R.; Seeliger, D.; Aβfalg, K.; Baatsen, P.; Compernolle, G.; Gils, A.; Studts, J. M.; Schulz, P.; Garidel, P.; Schymkowitz, J.; Rousseau, F. Prediction and Reduction of the Aggregation of Monoclonal Antibodies. *J. Mol. Biol.* **2017**, *429* (8), 1244−1261.

(43) Chiti, F.; Dobson, C. M. Amyloid Formation by Globular Proteins under Native Conditions. *Nat. Chem. Biol.* **2009**, *5* (1), 15−22.

(44) Buß, O.; Rudat, J.; Ochsenreither, K. FoldX as Protein Engineering Tool: Better Than Random Based Approaches? *Comput. Struct. Biotechnol. J.* **2018**, *16*, 25−33.

(45) Usmanova, D. R.; Bogatyreva, N. S.; Ariño Bernad, J.; Eremina, A. a.; Gorshkova, A. a.; Kanevskiy, G. M.; Lonishin, L. R.; Meister, A. V.; Yakupova, A. G.; Kondrashov, F. a.; Ivankov, D. N. Self-Consistency Test Reveals Systematic Bias in Programs for Prediction Change of Stability upon Mutation. *Bioinformatics* **2018**, 1−6.

(46) Kumar, V.; Rahman, S.; Choudhry, H.; Zamzami, M. a.; Sarwar Jamal, M.; Islam, A.; Ahmad, F.; Hassan, M. I. Computing Disease-Linked SOD1Mutations: Deciphering Protein Stability and Patient-Phenotype Relations Article. *Sci. Rep.* **2017**, *7* (1), 1−13.

(47) Navarro, S.; Villar-Piqué, A.; Ventura, S. Selection against Toxic Aggregation-Prone Protein Sequences in Bacteria. *Biochim. Biophys. Acta, Mol. Cell Res.* **2014**, *1843* (5), 866−874.

(48) Villar-Pique, A.; De Groot, N. S.; Sabaté, R.; Acebrón, S. P.; Celaya, G.; Fernàndez-Busquets, X.; Muga, A.; Ventura, S. The Effect of Amyloidogenic Peptides on Bacterial Aging Correlates with Their Intrinsic Aggregation Propensity. *J. Mol. Biol.* **2012**, *421* (2−3), 270−281.

(49) De Groot, N. S.; Pallarés, I.; Avilés, F. X.; Vendrell, J.; Ventura, S. Prediction of "Hot Spots" of Aggregation in Disease-Linked Polypeptides. *BMC Struct. Biol.* **2005**, *5*, 18.

(50) De Groot, N. S.; Ventura, S. Protein Activity in Bacterial Inclusion Bodies Correlates with Predicted Aggregation Rates. *J. Biotechnol.* **2006**, *125* (1), 110−113.

(51) Camilloni, C.; Sala, B. M.; Sormanni, P.; Porcari, R.; Corazza, A.; De Rosa, M.; Zanini, S.; Barbiroli, A.; Esposito, G.; Bolognesi, M.; et al. Rational Design of Mutations That Change the Aggregation Rate of a Protein While Maintaining Its Native Structure and Stability. *Sci. Rep.* **2016**, *6*, 25559.

(52) Varejão, N.; Ibars, E.; Lascorz, J.; Colomina, N.; Torres-rosell, J.; Reverter, D. DNA Activates the Nse 2/Mms 21 SUMO E 3 Ligase in the Smc 5/6 Complex. *EMBO J.* **2018**, *37*, e98306.

(53) Melnik, T.; Povarnitsyna, T.; Solonenko, H.; Melnik, B. Studies of Irreversible Heat Denaturation of Green Fluorescent Protein by Differential Scanning Microcalorimetry. Thermochim. *Thermochim. Acta* **2011**, *512* (1−2), 71−75.

(54) Jackson, S. E.; Craggs, T. D.; Huang, J. R. Understanding the Folding of GFP Using Biophysical Techniques. Expert Rev. *Expert Rev. Proteomics* **2006**, *3* (5), 545−559.

(55) Hsu, S. T. D.; Blaser, G.; Jackson, S. E. The Folding, Stability and Conformational Dynamics of B-Barrel Fluorescent Proteins. *Chem. Soc. Rev.* **2009**, *38* (10), 2951−2965.

(56) Pallarès, I.; Vendrell, J.; Avilés, F. X.; Ventura, S. Amyloid Fibril Formation by a Partially Structured Intermediate State of A-Chymotrypsin. *J. Mol. Biol.* **2004**, *342* (1), 321−331.

(57) Patterson, G. H. A. Photoactivatable GFP for Selective Photolabeling of Proteins and Cells. *Science (Washington, DC, U. S.)* **2002**, *297* (5588), 1873−1877.

(58) Zacharias, D. a.; Violin, J. D.; Newton, A. C. Partitioning of Lipid-Modified Monomeric CFPs into Membrane Microdomains of Live Cells. *Science* **2002**, *296*, 913−916.

(59) Snapp, E. L.; Hegde, R. S.; Francolini, M.; Lombardo, F.; Colombo, S.; Pedrazzini, E.; Borgese, N.; Lippincott-Schwartz, J. Formation of Stacked ER Cisternae by Low Affinity Protein Interactions. *J. Cell Biol.* **2003**, *163* (2), 257−269.

(60) Perchiacca, J. M.; Bhattacharya, M.; Tessier, P. M. Mutational Analysis of Domain Antibodies Reveals Aggregation Hotspots within and near the Complementarity Determining Regions. *Proteins: Struct., Funct., Genet.* **2011**, *79* (9), 2637−2647.

(61) Fernandez-Escamilla, A.-M.; Rousseau, F.; Schymkowitz, J.; Serrano, L. Prediction of Sequence-Dependent and Mutational Effects on the Aggregation of Peptides and Proteins. *Nat. Biotechnol.* **2004**, *22* (10), 1302−1306.

(62) Kramer, R. M.; Shende, V. R.; Motl, N.; Pace, C. N.; Scholtz, J. M. Toward a Molecular Understanding of Protein Solubility: Increased Negative Surface Charge Correlates with Increased Solubility. *Biophys. J.* **2012**, *102* (8), 1907−1915.

(63) Elgundi, Z.; Reslan, M.; Cruz, E.; Sifniotis, V.; Kayser, V. The State-of-Play and Future of Antibody Therapeutics. *Adv. Drug Delivery Rev.* **2017**, *122*, 2−19.

CHAPTER 3

Coiled-Coil Inspired Functional Inclusion Bodies

Marcos Gil-Garcia, Susanna Navarro and Salvador Ventura

Microbial Cell Factories

Coiled-coil inspired functional inclusion bodies

Marcos Gil-Garcia, Susanna Navarro and Salvador Ventura*

Abstract

Background: Recombinant protein expression in bacteria often leads to the formation of intracellular insoluble protein deposits, a major bottleneck for the production of soluble and active products. However, in recent years, these bacterial protein aggregates, commonly known as inclusion bodies (IBs), have been shown to be a source of stable and active protein for biotechnological and biomedical applications. The formation of these functional IBs is usually facilitated by the fusion of aggregation-prone peptides or proteins to the protein of interest, leading to the formation of amyloid-like nanostructures, where the functional protein is embedded.

Results: In order to offer an alternative to the classical amyloid-like IBs, here we develop functional IBs exploiting the coiled-coil fold. An in silico analysis of coiled-coil and aggregation propensities, net charge, and hydropathicity of different potential tags identified the natural hom -dimeric and anti-parallel coiled-coil ZapB bacterial protein as an optimal candidate to form assemblies in which the native state of the fused protein is preserved. The protein itself forms supramolecular fib illar networks exhibiting only α-helix secondary structure. This non-amyloid self-assembly propensity allows generating innocuous IBs in which the recombinant protein of interest remains folded and functional, as demonstrated using two different fluo escent proteins.

Conclusions: Here, we present a proof of concept for the use of a natural coiled-coil domain as a versatile tool for the production of functional IBs in bacteria. This α-helix-based strategy excludes any potential toxicity drawback that might arise from the amyloid nature of β-sheet-based IBs and renders highly active and homogeneous submicrometric particles.

Keywords: Coiled-coil protein, Fusion tag, Functional inclusion bodies, Fluorescent proteins, Protein engineering

Background

Biotechnological and pharmaceutical industries exploit microorganisms as cell factories in order to produce their biological products, including therapeutic proteins, such as hormones, enzymes for replacement therapies, or antibodies [1, 2]. However, the production of these molecules in their soluble and functional states faces significant barriers [3, 4]. Proteins have been shaped by natural selection to remain soluble and functional under physiological conditions, according to the "living on the edge" hypothesis [5]. The heterologous expression of these molecules in bacteria leads to intracellular concentrations that are several times above their natural solubility limits. As a result, these proteins might establish non-native intermolecular interactions, which would facilitate their aggregation into inclusion bodies (IBs) in the bacterial cytosol [6].

Traditionally IBs were thought to be formed by misfolded conformations and thus devoid of any functionality. However, data is accumulating to indicate that, at least for specific proteins and production conditions, IBs might exhibit significant activity [7, 8]. IBs are easy to purify and can be stored for long periods, thanks

*Correspondence: salvador.ventura@uab.cat
Institut de Biotecnologia i de Biomedicina i Departament de Bioquímica i Biologia Molecular, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain

Gil-Garcia *et al. Microb Cell Fact*   (2020) 19:117

Page 2 of 16

to their inherent stability [9]. They have a nanometric size (50–1000 nm) [9–12], and ~ 90% of them are composed of the target protein [13–15]. These properties, combined, make them active nanoparticles, which are finding increasing applications in biotechnology and biomedicine. In this way, the ability to immobilize enzymes in IBs has been exploited to build up reusable catalysts [16, 17], and IBs have been used as nanocarriers or/and nanopills to deliver antitumoral polypeptides in the body [18, 19].

We have shown that, generically, a significant proportion of the protein contacts that sustain IBs have an amyloid-like nature [20–22]. When embedded in IBs, proteins exhibit a significant increase in β-sheet content, relative to their soluble counterparts, and, often, become able to bind typical amyloid dyes. For specific proteins or protein fusions, this amyloid scaffold coexists with functional conformations [23, 24]. However, because the process of protein aggregation into non-native intermolecular β-sheet structures necessarily involves the population of misfolded species, a fraction of the recombinant protein is necessarily inactivated to build up the amyloid structure that sustains the IB [22].

IBs have been assimilated to natural functional amyloids [25], which are non-toxic for their host cells or organisms [26–28]. However, different studies indicate that this is not because functional amyloids are intrinsically non-cytotoxic [29], but instead because, in these specific cases, nature has evolved dedicated mechanisms to prevent amyloid-associated toxicity [27, 30, 31]. The artificial formation of IBs lack these natural control mechanisms and, although bacterial aggregates have been assumed to be innocuous, it cannot be entirely discarded that toxic β-sheet conformations can be incorporated or released from these amyloid-like inclusions.

In order to overcome the two above-described limitations, we introduce here an IBs production strategy that exploits a natural coiled-coil protein to promote non-amyloid supramolecular interactions. In the last years, the use and design of coiled-coil domains as building blocks in protein assemblies have attracted significant attention [32, 33]. The self-organization capacity of these structures has been exploited to create different nanostructures, such as nanofibers [34] and nanocages [35, 36]. Furthermore, two different coiled-coil domains have been previously used to generate active IBs: the tetramerization domain of the tetrabrachion protein (TDoT) from *Staphylothermus marinus* [16, 37–40] and the 3HAMP coiled-coil, which was derived from the oxygen sensor protein Aer2 from *Pseudomonas aeruginosa* [37, 41]. In this work, we apply this strategy to build up functional IBs using ZapB, a non-essential *Escherichia coli* (*E. coli*) protein consisting of two anti-parallel α-helices,

involved in Z-ring formation during the bacterial cell division process [42, 43].

We first show that the ZapB self-assembles to form α-helix-based fibrillar networks, and afterward, we demonstrate how this property allows its use as a tag to form non-amyloid and non-toxic IBs which preserve the activity of the attached polypeptides.

Results and discussion

Selection of a polar and non-aggregating coiled-coil protein for the production of functional IBs

A wide range of fusion tags has been used to induce IBs formation. They comprise small artificial peptides [44], and aggregation-prone natural proteins or domains [22, 37, 45], which are fused to functional globular proteins. A characteristic property of most of these tags is that they promote the formation of aggregates sustained by collective intermolecular β-sheet interactions. Alternatively, two different coiled-coil domains have been used for the production of fluorescent and/or catalytically IBs [37]. Nevertheless, the biophysical properties of these active aggregates were not assessed, and thus it is not known if the coiled-coil encoding sequences keep their native helical structure in the IBs or they had just transitioned to a conventional amyloid-like assembly. Indeed, the formation of β-sheet-rich amyloid fibrils by aggregation-prone coiled-coil sequences is behind the onset of several neurodegenerative disorders [46]. A similar transition in coiled-coil-tagged proteins during IBs formation might turn these aggregates potentially toxic.

We selected the *E. coli* protein ZapB as a scaffold to obtain functional IBs. ZapB is an 81 residues-long protein whose 3D-structure (PDB: 2JEE) consists of two α-helical polypeptide chains arranged in anti-parallel orientation to form a dimeric coiled-coil of 116 Å (PDB: 2JEE) [42]. In the crystal structure, individual coiled-coils interact close to their termini, which already suggested that, under appropriate conditions, these helical modules might self-assemble into supramolecular structures [42].

The propensity to form a stable coiled-coil assembly in solution is encoded in the protein sequence. The higher the coiled-coil propensy, the lowest the probability to transition into an aggregated β-sheet structure since stable α-helices protect against aggregation [47, 48]. We calculated the coiled-coil propensity of ZapB and compared it with that of the two coiled-coil domains used as IBs formation tags in previous studies (3HAMP and TDoT) using four different algorithms: COILS [49], PCoils [50], MARCOIL [51] and DeepCoil [52]. Additional file 1: Figures S1–S3 show the coiled-coil probability profiles for ZapB, 3HAMP and TDoT. The four algorithms coincide to predict a very high coiled-coil propensity along the complete ZapB sequence. In the case of 3HAMP, the

Gil-Garcia *et al. Microb Cell Fact*   (2020) 19:117

Page 3 of 16

programs identify a region of high propensity close to the N-terminus and two additional stretches with low to moderate propensity. This is consistent with the homodimeric 3HAMP structure, in which parallel monomers exhibit three successive domains (HAMP1, 2, and 3), each about 50 residues long and bridged by flexible linkers. For TDoT, only DeepCoil is able to identify a significant coiled-coil propensity in the central part of the sequence. This makes sense, since TDoT is a parallel and right-handed coiled-coil tetramer, which is based on the 11-residue repeat, and COILS, PCoils and MARCOIL were trained to identify canonical heptad repeats, where DeepCoil was aimed to identify both kinds of periodicities. Therefore, we used this last algorithm to compare the average coiled-coil probabilities of ZapB, TDoT and 3HAMP primary sequences. As it can be seen in Fig. 1 ZapB seems to be a better coiled-coil former than the proteins it has been compared to.

Not only the coiled-coil propensity accounts for the ability to maintain the native state in the multimeric state, but also the protein's intrinsic aggregation propensity, which facilitates the conformational shift to aggregated β-sheet states. We analyzed this property for ZapB, TDoT, 3HAMP, and other three non-coiled-coil sequences used previously as IB-tags, namely the amyloid β-peptide (Aβ42) [53, 54], the viral capsid peptide VP1 [45, 55], and the signal sequence of *E. coli* TorA (ssTorA) [56, 57]. In order to do that, we used two of the most popular sequence-based aggregation prediction servers, Aggrescan [58] and TANGO [59]. As it can be seen in Fig. 2a, b, ZapB was predicted as the least aggregation-prone sequence in this polypeptide set. Aggrescan (Fig. 2a), predicts ZapB to be the most soluble sequence with a significant difference, relative to the other proteins or domains. In the case of TANGO (Fig. 2b), ZapB remains as the less aggregation-prone sequence, in this case, together with ssTorA.

Fig. 1 Coiled-coil probability for ZapB, TDoT, and 3HAMP. The coiled-coil propensity is represented according to DeepCoil predictor
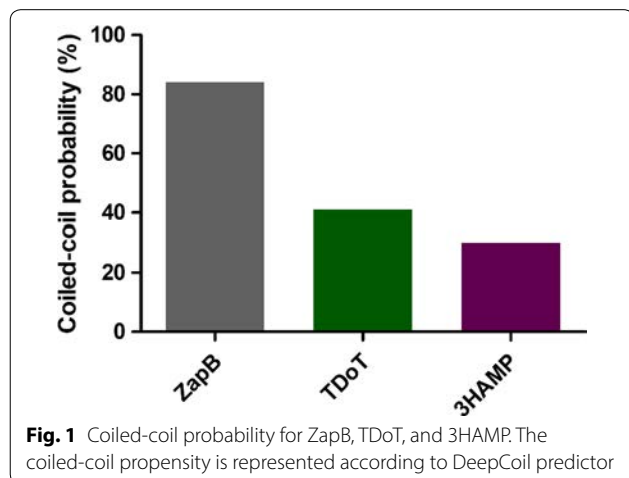
Polarity and hydrophobicity are critical negative and positive contributors to protein sequence aggregation propensity, respectively. Their role is crucial at the nucleation step of the aggregation reaction [60, 61]. These biophysical properties were analyzed by calculating the Grand Average of Hydropathicity (GRAVY score) provided by the ProtParam server, according to the Kyte-Doolittle scale [62]. Positive GRAVY scores indicate higher hydrophobicity, whereas negative values correspond to polar sequences. As it is shown in Fig. 2c, ZapB has the more negative GRAVY score (−0.980), being this value three times higher than the secondly ranked sequence (VP1, GRAVY score=−0.307), thus indicating that ZapB is a highly polar protein, relative to compared sequences.

We also checked the secondary structure propensity of the six tags. Cryptic regions of significant β-sheet propensity might exist even in sequences that usually fold into α-helices [63]. These stretches might favor aggregation into amyloid-like structures upon coiled-coil unfolding or once the polypeptide chain emerges from the ribosome [64]. To this aim, we used the PSIPRED [65] and GOR [66] servers. ZapB is predicted to be completely α-helix, with an extremely high propensity for this secondary structure; meanwhile, 3HAMP and specially TDoT are predicted to have a significantly lower α-helical propensity and predicted β-strand segments are identified at their sequences (Additional file 1: Figures S4, S5).

The above-described analyses converge to indicate that the sequence of ZapB is more polar, less aggregation-prone, devoid of cryptic β-sheet regions, and with highest α-helical and coiled-coil propensities than any previously used IBs tag. Each of these individual properties disfavors the potential aggregation of the ZapB sequence into β-sheet-rich aggregates. However, aggregation can still occur from the folded state of the proteins [67]. We have recently developed AGGRESCAN3D, an algorithm that allows predicting a protein's aggregation-propensity taking into account the structural context [68–70]. When we analyzed the published 3D structures of the three coiled-coil domains (ZapB, PDB:2JEE; TDoT, PDB:1FE6; and 3HAMP, PDB:3LNR), ZapB turned to be the less aggregation-prone structure, displaying a highly soluble surface (Additional file 1: Figure S6).

Finally, we assessed the net charge of the different IBs tagging sequences. This value is essential because if the high predicted solubility of ZapB comes together with a high positive charge, this will be a significant drawback for its use in nanomedical applications. Cationic sequences bind to negatively charged nucleic acids and cell membranes, which then become incorporated into IBs during their formation or purification, making them compositionally heterogeneous. The calculated net charge for the six tags is represented in
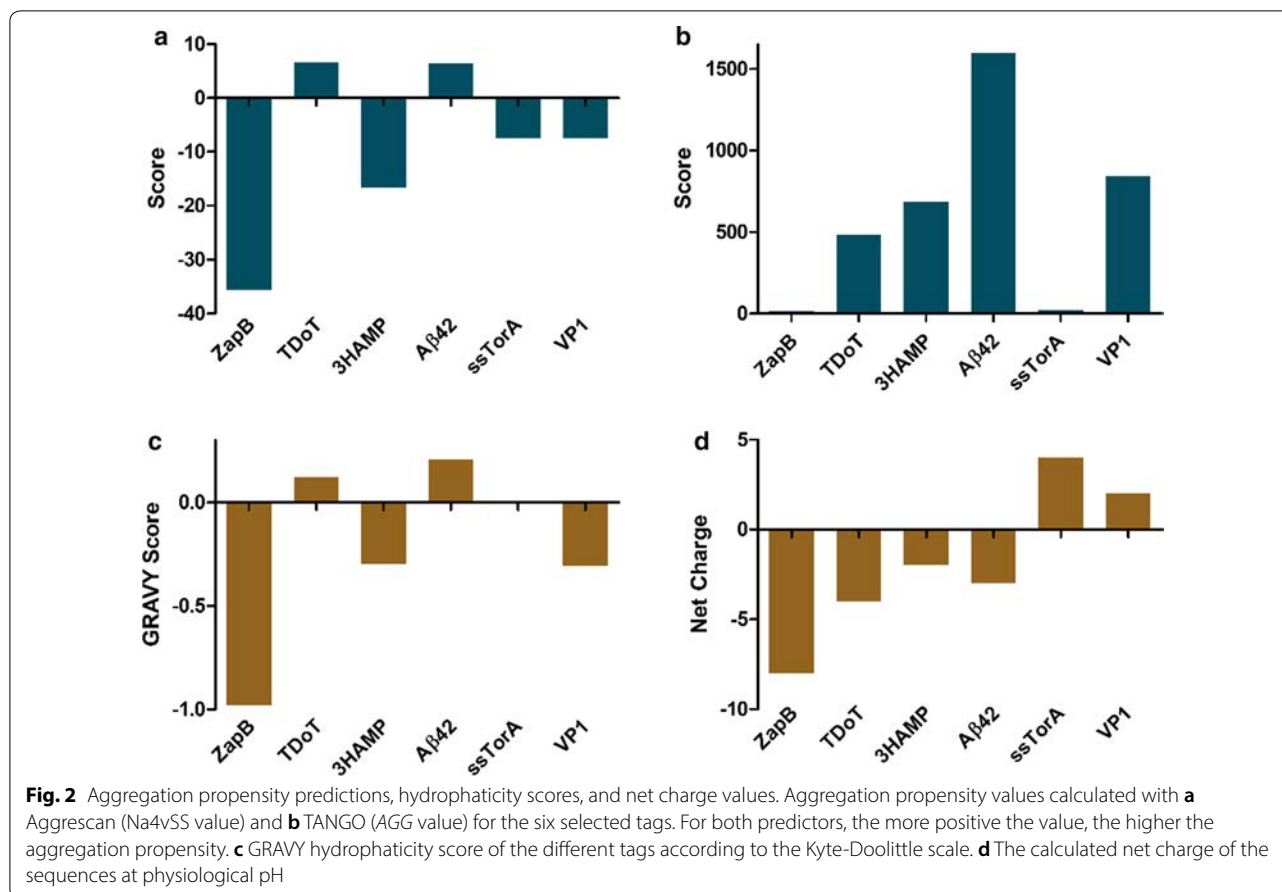
Fig. 2 Aggregation propensity predictions, hydrophaticity scores, and net charge values. Aggregation propensity values calculated with **a** Aggrescan (Na4vSS value) and **b** TANGO (*AGG* value) for the six selected tags. For both predictors, the more positive the value, the higher the aggregation propensity. **c** GRAVY hydrophaticity score of the different tags according to the Kyte-Doolittle scale. **d** The calculated net charge of the sequences at physiological pH

Fig. 2d. ZapB is the most acidic of the sequences (net charge −8), the net charge being two times higher than that of the second most anionic tag (TDoT, net charge −4). These tags differ in their lengths. In order to obtain a value independent of the protein size, we calculated the net charge per residue (NCPR) (Additional file 1: Figure S7) [71]. According to the NCPR values, ZapB is again the most acidic tag (NCPR: −0.099), followed by TDoT (NCPR: −0.077). Therefore, no interaction is expected between DNA, RNA, or membranes and ZapB.

Overall, we can conclude that, theoretically, ZapB fulfills all the requirements to work as a tag to promote the formation of non-amyloid-like functional IBs; still, to act as such, the domain should be able to self-assemble, despite its high predicted solubility in both the folded and unfolded states.

ZapB self-assembles into coiled-coil, non-amyloid, nanofibers

In order to test the ability of ZapB to form protein assemblies driven by interactions between natively folded coiled-coils, a His-tag was added at the C-terminus of the full-length protein, and it was recombinantly produced in *E.coli* at 30 °C. The protein was expressed at high yield (> 1 g/L culture) (Additional file 1: Figure S8) and purified from the soluble cell fraction by IMAC. This soluble fraction was significantly viscous, and an imidazole gradient should be used for ZapB homogeneous purification (Additional file 1: Figure S9). This unusual viscosity already suggested a certain degree of ZapB self-assembly, which was further corroborated by transmission electron microscopy (TEM) analysis of the purified protein. Abundant long protein nanofibers 20 ± 5 nm in width and displaying a regular striated pattern were observed (Fig. 3a), in excellent agreement with the nanostructures visualized in a previous study [42].

We analyzed the secondary structure content of self-assembled ZapB by circular dichroism (CD) spectroscopy. The spectrum is characteristic of an α-helical structure, with two minima at 208 nm and 222 nm (Fig. 3b). Some studies have revealed that the 222/208 nm ratio allows discriminating individual α-helices from coiled-coils, owing to the different periodicity of the two folds. A 222/208 nm ratio ≥ 1 is

Gil-Garcia *et al. Microb Cell Fact* (2020) 19:117

Page 5 of 16
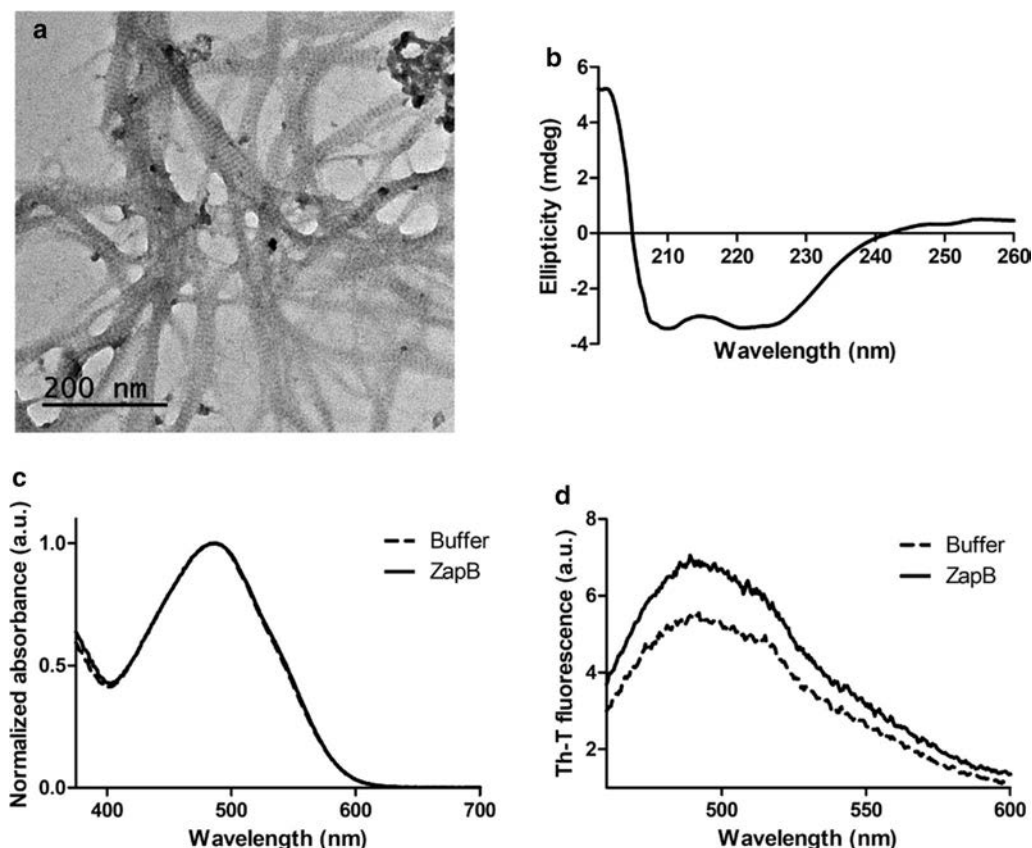
Fig. 3 ZapB coiled-coil assembly biophysical characterization. **a** TEM image of ZapB fibers upon negati e staining. **b** Far-UV circular dichroism spectrum. **c** Congo-Red absorbance spectra and **d** Th-T fluo escence emission spectra

indicative of coiled-coil structures, and ratios ≤ 0.86 can be attributed to individual α-helices in a protein structure [72, 73]. For ZapB, the 222/208 nm ratio is 0.99, indicating that, as anticipated by the protein crystal structure (PDB: 2JEE), ZapB keeps the coiled-coil fold in the macromolecular fibrillar assembly. The absence of a β-signature in the CD spectrum of self-assembled ZapB is consistent with a non-amyloid nature. This trait was confirmed using the Congo Red (CR) and Thioflavin-T (Th-T) amyloid dyes. Both analyses indicated that the ZapB nanofibers do not have an amyloid-like nature, since the spectra of both dyes in the presence and absence of the protein (buffer alone) are fairly similar (Fig. 3c, d).

ZapB IBs exhibit a coiled-coil conformation

In order to explore whether, like the nanofilaments purified from the soluble cell fraction, ZapB IBs are sustained by coiled-coil interactions, they were purified from the insoluble cell fraction (Additional file 1: Figure S10) and their secondary structure content analyzed by CD

spectroscopy and Fourier Transform Infrared Spectroscopy (FTIR).

The Far-UV CD spectrum of ZapB IBs resembles the one obtained for the nanofibrillar solution (Fig. 4a). The packing of ZapB into IBs seems to favor the preservation of the coiled-coil conformation since the 222/208 nm ratio of these aggregates is 1.14. We recorded the infrared spectra of ZapB IBs in the amide I region of the spectrum (1700–1600 cm^{-1}), corresponding to the absorption of the main chain carbonyl group and sensitive to protein conformation (Fig. 4b). It is important to note here, that in coiled-coils, the supercoil bending of the α-helices results in a spectral splitting of the α-helical IR amide I band and in a shift to lower wavenumbers [74]. FTIR measurements on the model coiled-coil GCN4 demonstrated the assignment of the solvated portion of the coiled-coil to a low helix frequency (1631 cm^{-1}); in contrast, the buried helix frequency (1651 cm^{-1}) is observed for residues in the interior of the coiled-coil. These two signals alone accounted for the 78% of the GCN4 spectral area [75, 76]. Similarly, the major contributors to ZapB IBs IR spectra are two signals at 1632 cm^{-1} and

Fig. 4 Conformational properties of ZapB IBs. **a** Far-UV circular dichroism spectrum. **b** FTIR absorbance spectrum in the amide I region (dashed line). The different component bands (solid lines) were obtained by deconvolution of the absorbance spectrum. **c** SEM image of a ZapB IB. **d** Congo-Red absorbance spectra. **e** Th-T fluo escence emission spectra
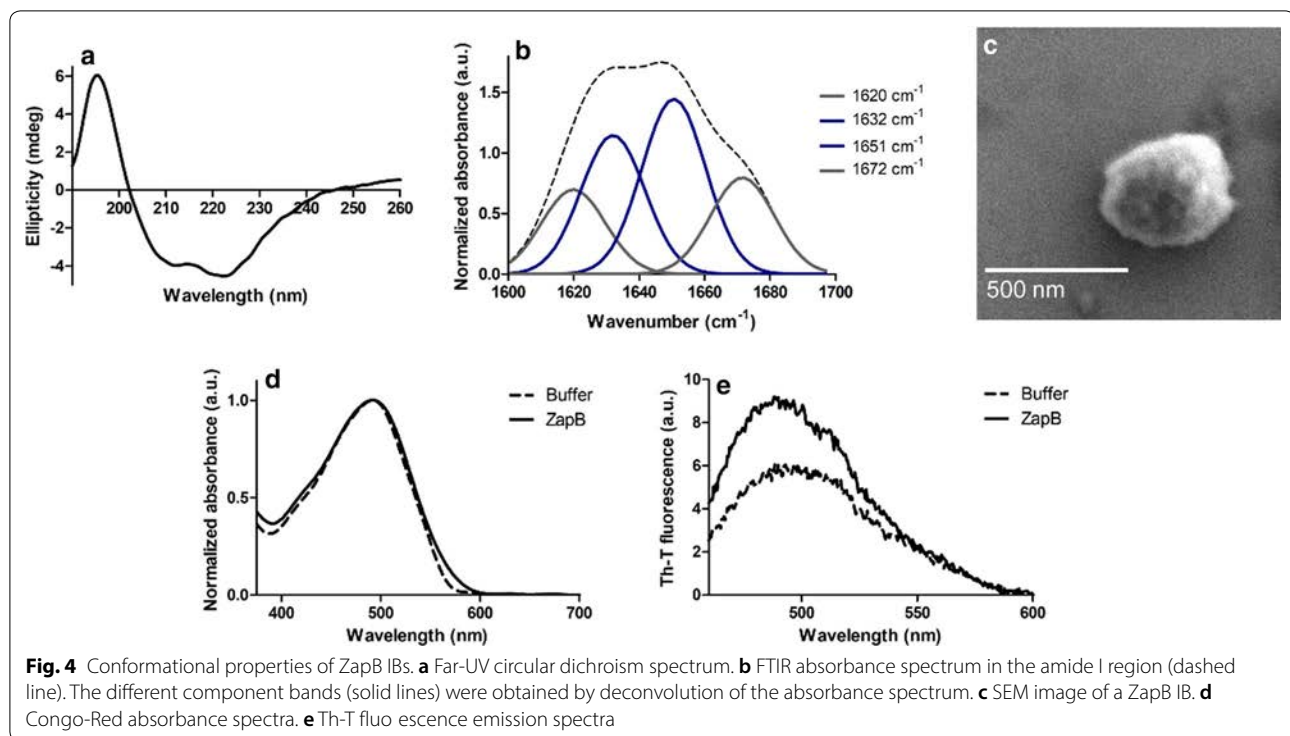
1651 cm^{-1} bands, accounting for the 28% and the 35% of the absorbance spectrum area, respectively. Additional bands at 1620 cm^{-1} and 1672 cm^{-1} were detected, the latest one being also present in the IR spectra of GCN4 and other coiled-coil proteins [75]. When observed using Scanning Electron Microscopy (SEM) (Fig. 4c), these IBs displayed a spherical shape. The non-amyloid nature of the IBs was confirmed using both CR and Th-T amyloid dyes, observing similar absorbance and fluorescence spectra for the IBs and the buffer alone, respectively (Fig. 4d, e).

Overall, the CD and FTIR data converge to indicate that ZapB IBs consist mostly of coiled-coil molecules.

ZapB-GFP and ZapB-mCherry proteins are produced as fluorescent IBs in *E. coli*

In order to test the ability of ZapB to assist the formation of functional IBs inside bacteria, ZapB was N-terminally fused to GFP and expressed in *E. coli*. After induction of protein expression, the soluble and the insoluble cellular fractions were separated by centrifugation and analyzed by SDS-PAGE. As shown in Fig. 5a, the ZapB-GFP fusion protein was mainly located in the insoluble fraction (~95%), whereas, non-tagged insoluble GFP accounts for ~30% of the recombinant protein (Additional file 1: Figure S11). Thus, the fusion of the two polypeptides facilitates GFP deposition. We used fluorescence confocal microscopy to localize the GFP fluorescence

emission in *E. coli* intact cells. As expected, the GFP signal was confined mainly in IBs placed at the poles of cells (Fig. 5b).

GFP is considered the default fluorescent protein for most applications, but when dealing with deep in vivo imaging, red-shifted fluorescent proteins are preferred since, at these wavelengths, light absorption by tissues is significantly lower [77], among them mCherry is one of the most used variants [78, 79]. We N-terminally fused ZapB to mCherry, in order to asses if we can obtain red fluorescent IBs with potential in vivo applications.

ZapB-mCherry was expressed in *E. coli*, and the soluble and the insoluble fraction were separated as above. In this case, ~55% of the fusion protein is present in the insoluble fraction (Fig. 6a), whereas in non-tagged mCherry only ~5% of the protein is insoluble (Additional file 1: Figure S12). The difference between the fraction of ZapB-GFP and ZapB-mCherry located in the respective insoluble fractions likely owes to the highest solubility of the mCherry structure when compared with GFP, as assessed using the AGGRESCAN3D algorithm (Additional file 1: Figure S13). When the location of the red fluorescence was monitored using confocal microscopy, highly fluorescent IBs become evident at the poles. However, their discretization was more difficult than in the case of ZapB-GFP, due to the soluble fusion protein fluorescent background (Fig. 6b).

Gil-Garcia *et al. Microb Cell Fact*   (2020) 19:117

Page 7 of 16

Fig. 5 Cellular distribution of ZapB-GFP IBs. **a** SDS-PAGE of intact cells, soluble, and insoluble cellular fractions. The arrow indicates the ZapB-GFP. **b** Visualization of ZapB-GFP IBs in intact *E. coli* cells by confocal microscopy. Some IBs located at the poles of cells are indicated
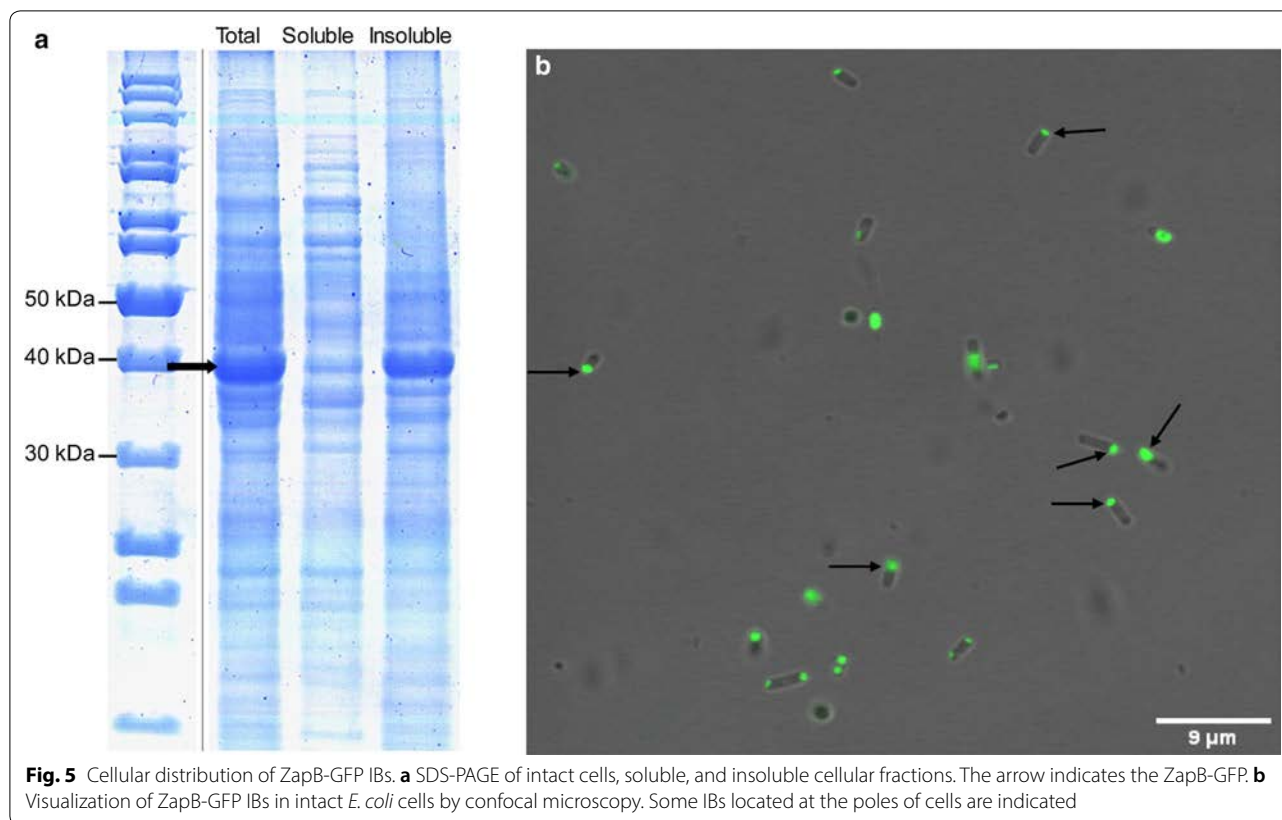
GFP and mCherry maintain native spectral properties in ZapB-based IBs

In order to evaluate the impact of the coiled-coil structure in the functionality of the attached fluorescent proteins when embedded in the IBs, we purified both IBs from the insoluble fraction (Additional file 1: Figures S14, S15). The analysis of the purified protein aggregates using an epifluorescence microscope and adequate filters allowed us to observe the presence of abundant green and red fluorescent particles for ZapB-GFP and ZapB-mCherry, respectively (Fig. 7a,b). We compared the spectral properties of the fluorescent proteins trapped in the IBs with those of their soluble and non-tagged counterparts. As it is shown in Fig. 7c, soluble GFP and ZapB-GFP IBs presented identical excitation and emission maxima at 495–496 and 512 nm, respectively. The same behavior was observed when comparing soluble mCherry with ZapB-mCherry IBs, both displaying excitation and emission maxima at 589 and 604 nm, respectively (Fig. 7d). The emission spectra of ZapB-GFP and ZapB-mCherry overlap perfectly with that of the respective soluble untagged fluorescent proteins, whereas, in both cases, the left side of the excitation spectrum is slightly red-shifted when the protein is located within the IBs, which likely respond to differences in crowding and

or mobility between soluble and assembled fluorescent proteins.

Overall, these results allow us to confirm that GFP and mCherry keep their activity and, likely, their native conformation inside ZapB-induced IBs.

ZapB-GFP inclusion bodies contain a coiled-coil conformation

We selected the ZapB-GFP fusion as a model system to further study the properties of ZapB promoted IBs. We analyzed the secondary structure content of these IBs using CD spectroscopy and FTIR. The far-UV CD spectrum of ZapB-GFP is of course influenced by the all β-sheet structure of GFP (Fig. 8a); still, the two main signatures of α-helices could be detected in the IBs, with the global minimum placed at 222 nm and inflection of the spectrum at 208 nm (Fig. 8a).

When ZapB-GFP IBs were analyzed using FTIR in the amide I region of the spectrum, we could detect, again, the characteristic helical band at 1651 cm^{-1}, accounting for 39% of the absorbance area. However, now the major signal is located at 1626 cm^{-1} (49% of the area). This band, likely results from the sum of the low-frequency coiled-coil signal and the β-sheet signal of the fused GFP β-barrel (Fig. 8b). An additional band at 1678 cm^{-1}, likely

Gil-Garcia *et al. Microb Cell Fact*   (2020) 19:117

Page 8 of 16

Fig. 6 Cellular distribution of ZapB-mCherry IBs. **a** SDS-PAGE of intact cells, soluble, and insoluble cellular fractions. The arrow indicates the ZapB-mCherry. **b** Visualization of ZapB-mCherry IBs in intact *E. coli* cells by confocal microscopy. Some IBs located at the poles of cells are indicated
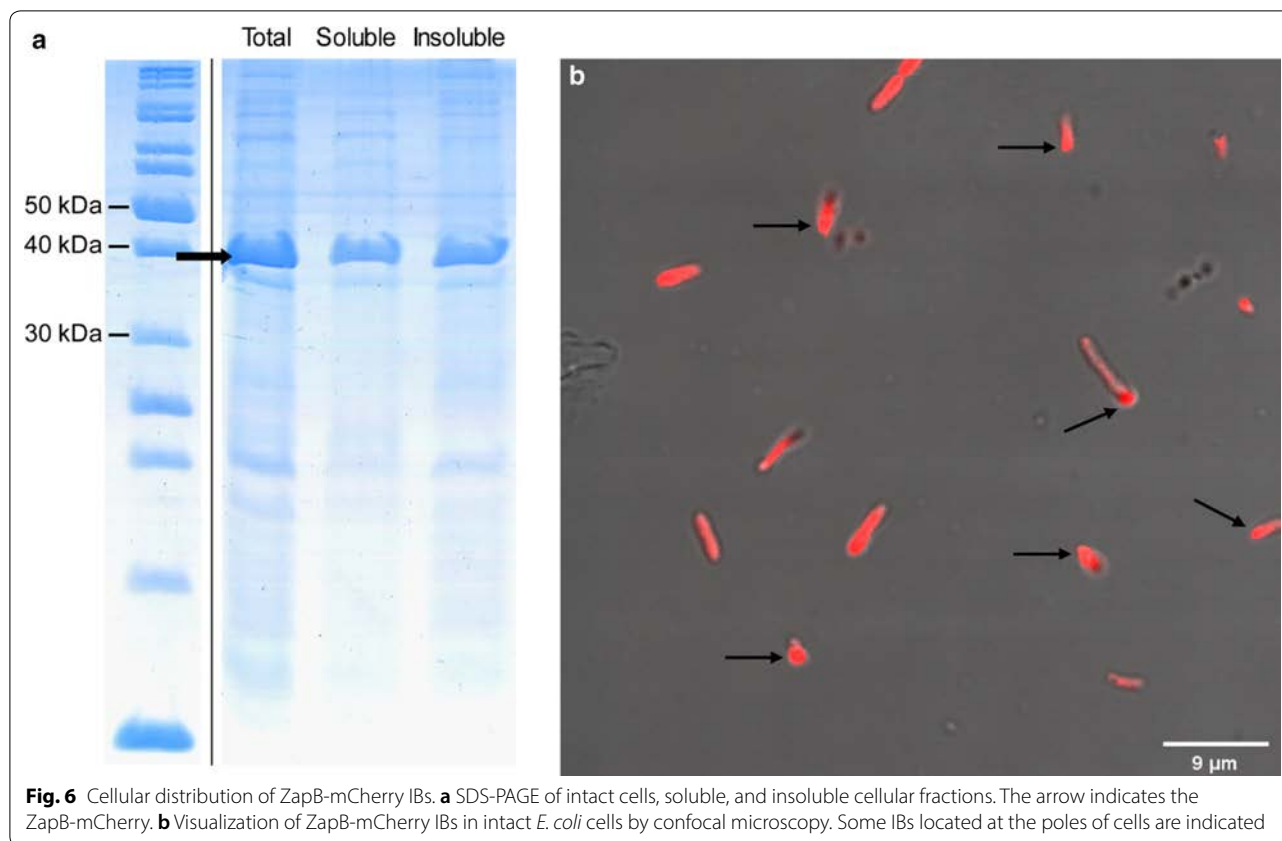
arising from the mixed contribution of the two structurally different domains in the fusion at high frequencies was also evident. Finally, the non-amyloid character of ZapB-GFP IBs was corroborated using the CR amyloid dye, observing similar absorbance spectra in the presence of ZapB-GFP IBs, soluble GFP or only buffer (Fig. 8c).

Overall, the secondary structure analyses of ZapB-GFP IBs suggest that they contain a significant proportion of coiled-coil conformations and do not have an amyloid-like nature. This is likely also the case for ZapB-mCherry IBs, since their binding to Th-T and CR are negligible (Additional file 1: Figure S16).

Comparison of ZapB-based and amyloid-like IBs

We wanted to compare the properties of the above described coiled-coil-based IBs with those of model amyloid-like IBs. To this aim, we selected fusion of the Alzheimer's related β-amyloid peptide (Aβ42) and GFP. In previous studies, we have characterized in detail the properties of Aβ42-GFP IBs [22, 80, 81]. The high amyloid propensity of the Aβ42 peptide drives the incorporation of the GFP moiety into β-sheet enriched and fluorescent IBs.

We expressed Aβ42-GFP (Additional file 1: Figure S17) and purified its IBs (Additional file 1: Figure S18) and

compared the spectral properties of GFP in these aggregates with those of ZapB-GFP. Figure 9a demonstrates that Aβ42-GFP and ZapB-GFP IBs share the same excitation and emission spectra, indicating that the active and properly folded GFP they contain is in a similar environmental context. The size of the IBs was analyzed using dynamic light scattering (DLS). Both IBs exhibited similar sizes, with moderately polydisperse distributions and calculated average diameters of 462.2 ± 69.51 nm and 463.3 ± 99.9 nm, for ZapB-GFP and Aβ42-GFP, respectively. The DLS data suggested that ZapB-GFP IBs are quite homogeneous in size (Additional file 1: Figure S19). Effectively, when these aggregates were imaged by Scanning Electron Microscopy (SEM), it was observed that they correspond to submicrometric spherical assemblies (Fig. 9b), and accordingly, they can be assimilated to protein nanoparticles.

Once confirmed that ZapB-GFP and Aβ42-GFP IBs share spectral properties and dimensions, we wondered if the GFP activity in both nanostructures was similar. To this aim, we analyzed the GFP fluorescence intensity of both IBs by fluorescence microscopy. The mean fluorescence intensity values as obtained from images (Additional file 1: Figure S20) quantification of 50 isolated fluorescent dots in each sample, using ImageJ, revealed

Gil-Garcia *et al. Microb Cell Fact*   (2020) 19:117

Page 9 of 16
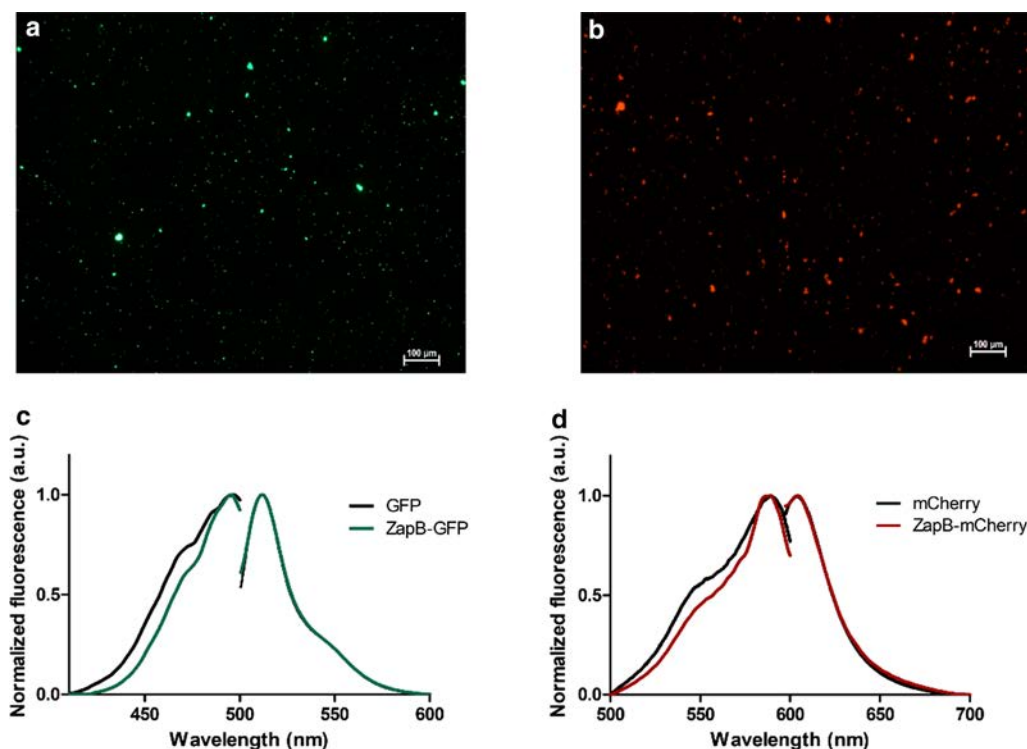
Fig. 7 Epifluoescence microscopy images and spectral properties of ZapB-GFP and ZapB-mCherry IBs. **a** Fluorescence microscopy image of purified Zap-GFP IBs. **b** Fluorescence microscopy image of purified Zap-mCherry IBs. **c** Excitation and emission spectra of ZapB-GFP IBs and soluble GFP. **d** Excitation and emission spectra of ZapB-mCherry IBs and soluble mCherry

that ZapB-GFP IBs exhibited two times higher activity (3713 ± 91.25 a.u.) than Aβ42-GFP IBs (1839 ± 25.23 a.u.) (Fig. 10). This observation is not surprising since the assembly of Aβ42-GFP IBs depends on an aberrant interaction between hydrophobic Aβ42 regions, which leads to a relatively rapid aggregation into amyloid-like structures, with the subsequent inactivation of a least a fraction of the attached globular domains, with their most aggregation-prone sequence stretches contributing to stabilize the aggregate through amyloid-like contacts [22]. Indeed, we have demonstrated that the activities of the IBs formed by 20 Aβ42-GFP variants, bearing different mutations in the Aβ42 moiety, inversely correlate with the aggregation propensities of the peptides [82]. Aβ42-GFP IBs inactivation was favored by increased β-sheet propensity and hydrophobicity and counteracted by increased net charge [54]. ZapB has a negligible β-sheet propensity, is polar, highly charged and has a very low aggregation propensity, compared with Aβ42, all these factors likely contributing to the higher activity of ZapB-GFP IBs. In addition, the inter- and supramolecular assembly of ZapB is directed by native interactions, and not by non-native contacts, as in Aβ42, which

are expected to interfere less with the folding and structure of the GFP moiety, and indeed, many coiled–coil domains naturally exist and function appended to globular domains [83].

ZapB-GFP IBs are innocuous for human cells

One of the potential advantages of coiled-coil inspired IBs, relative to amyloid-like IBs, is that in the absence of intermolecular β-sheet assemblies that might elicit cytotoxicity, these α-helical-based assemblies should be non-toxic for human cells. To confirm this extent, we incubated HeLa cells with increasing concentrations of ZapB-GFP IBs (from 0 to 12 μM) for 72 h and monitored their viability using PrestoBlue® fluorescent assay. As it can be observed in Fig. 11a, the IBs turned to be innocuous at any of the assayed concentrations, which should make them suitable for in vivo applications. This is contrast with Aβ42-GFP IBs, which exhibit a moderate and concentration-dependent toxicity for HeLa cells (Fig. 11b), in good agreement with the toxicity we described previously for Aβ42 IBs [84].

Gil-Garcia *et al. Microb Cell Fact* (2020) 19:117

Page 10 of 16

Fig. 8 Conformational properties of ZapB-GFP IBs. **a** Far-UV circular dichroism spectrum of ZapB-GFP IBs and soluble GFP. **b** FTIR absorbance spectrum of ZapB-GFP IBs in the amide I region (dashed line). The different component bands (solid lines) were obtained by deconvolution of the spectrum using PeakFit software. **c** Congo-Red absorbance spectra

finding amazing applications in biotechnology [86] and biomedicine [18, 19].

Conventionally, IBs are enriched in intermolecular β-sheet structure, and contain both amyloid-like and native protein conformations [24]. Unavoidably, a fraction of the target protein becomes inactivated to form the amyloid skeleton that provides mechanical stability to IBs. In order to endorse IBs with tailored functionalities, the particular protein of interest is usually fused to an aggregation-prone polypeptide that act as IB-forming tag [44, 87]. This allows to recruit otherwise soluble proteins into IBs. However, the high-aggregation potential of these tags compromises the folding of the target protein, a fraction of which establishes non-native contacts and becomes an integral part of the amyloid-like architecture, and thus inactive [22].

Because of their amyloidogenic nature, active IBs are assimilated to natural functional amyloids [88]. It is important to note here that, as pathogenic amyloids, functional amyloids found in nature are also inherently cytotoxic [29]; the difference being that nature has evolved dedicated mechanisms to prevent the toxicity of natural assemblies. In the absence of these control mechanisms it cannot can be completely discarded that amyloid-like IBs might contain or release toxic protein conformers.

A way to avoid protein inactivation by non-native interactions and potential toxicity is to run away from non-native-β-sheet based IBs and develop native-α-helix based IBs. We take here a step towards this direction by implementing ZapB based IBs. Catalytic coiled-coil based IBs have been described previously [16, 17, 37–41]. However, our computational analysis suggests that these sequences display lower α-helical and coiled-coil propensities and higher intrinsic aggregation propensities than ZapB, which suggests that a partial transition towards amyloid-like structures upon overexpression cannot be fully discarded, especially because the IBs they promoted were not conformationally characterized.

To avoid the above-mentioned α-helix to β-sheet transition, we have selected ZapB. A natural sequence with extremely high α-helical and coiled-coil propensities, a neglectable intrinsic aggregation propensity and a high polar and anionic character. This coiled-coil protein exhibits the ability to spontaneously assemble into α-helical macromolecular fibrils devoid of any amyloid character. We show how this property can be exploited to obtain innocuous, spherical, relatively homogeneous and highly active sub-micrometric coiled-coil inspired IBs. ZapB expands our existing toolbox to generate immobilized enzymes or biomedical nanocarriers, among other applications. However, the present work constitutes a proof-of-concept, and the ability of ZapB to facilitate the

Conclusions

IBs have been traditionally regarded as waste reservoirs containing only misfolded and thus non-active and useless proteins. However, it is now clear that at least some proteins can retain certain activity when embedded in these aggregates, turning them into functional sub-micron particles [23, 85]. These nanostructures are

Gil-Garcia *et al. Microb Cell Fact*   (2020) 19:117

Page 11 of 16

Fig. 9 Spectral properties and morphology of ZapB-GFP IBs. **a** GFP-Excitation and emission spectra of ZapB-GFP IBs, Aβ42-GFP IBs, and soluble GFP. **b** SEM image of ZapB-GFP IBs

Fig. 10 GFP fluo escence of ZapB-GFP and Aβ42-GFP inclusion bodies. Quantification of GFP flu   escence intensity from fluo escence microscopy images using ImageJ. The corresponding intensity of 50 individual fluo escent dots for each sample was analyzed and averaged. The statistical values were derived using the Mann–Whitney Test. A *p* value < 0.001 is indicated as ***. Error bars correspond to SEM

formation of functional IBs when fused to other protein folds should still be demonstrated.

Methods

In silico analysis

For the analysis of the coiled-coil tendency, four different predictors were used: Coils, DeepCoil, PCoils and MARCOIL. For the different predictions, the Coiled-coil probability (%) was calculated in order to obtain a quantitative value of the coiled-coil tendency of the analyzed sequences.

Aggregation propensity predictions were performed using Aggrescan and TANGO webservers. In the case of Aggrescan, the Normalized a4v Sequence Sum for 100 residues (Na^{4}vSS) was employed as the aggregation propensity value. In the case of TANGO, predictions were made using the default parameters and the *AGG* value was selected as the aggregation propensity value.

Fig. 11 Citotoxicity of ZapB-GFP and Aβ42-GFP IBs. HeLa cells viability was assessed after adding different concentrations (from 2 to 12 μM) of ZapB-GFP **a** and Aβ42-GFP **b** IBs for 72 h. PBS buffer was used as a control

Gil-Garcia *et al. Microb Cell Fact*   (2020) 19:117

Page 12 of 16
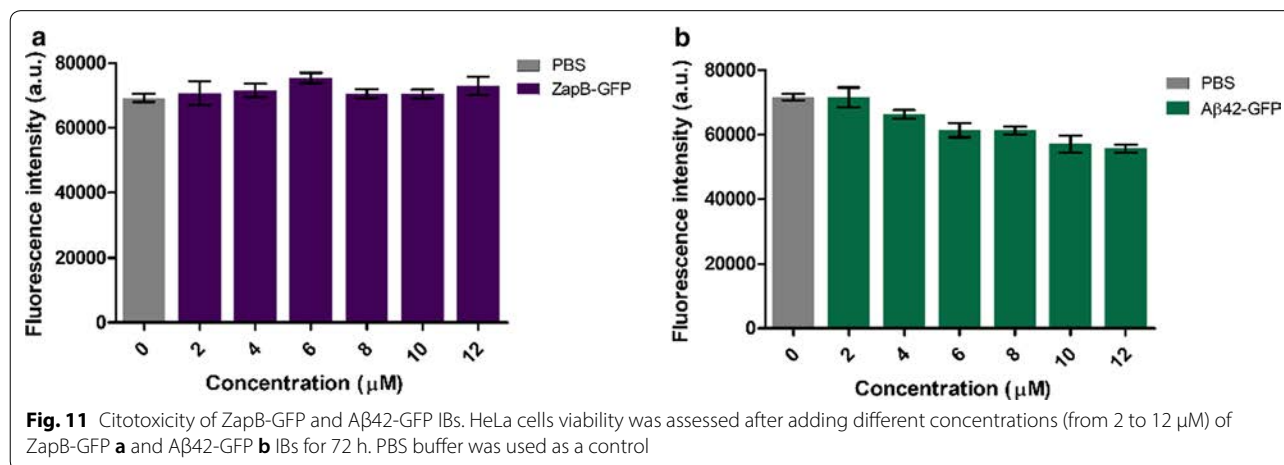
For the aggregation propensity predictions considering the 3D structure, AGGRESCAN3D webserver was used with the default parameters and using the following PDBs as an input: PDB: 2JEE for ZapB, PDB: 1FE6 for TDoT and PDB: 3LNR for 3HAMP, PDB: 2Y0G for GFP and PDB: 2H5Q for mCherry. In the case of ZapB and 3HAMP proteins, the dimeric structures were generated with PISA. Protein structures and their surfaces were represented with Pymol (DeLano Scientific, LLC).

The Grand Average of Hydropathicity (GRAVY score) and the net charge at physiological pH were calculated using the ProtParam tool. The net charge per residue (NCPR) was calculated dividing the net charge by the total number of residues.

Finally, secondary structure predictions were performed using GOR and PSIPRED webservers.

Protein production and purification

ZapB gene fragment (Additional file 1: DNA and amino acid sequences of ZapB protein) was cloned into a pET-21a vector between *NdeI* and *BamHI* restriction sites. The ZapB-GFP and ZapB-mCherry fusion proteins were generated inserting both target proteins after ZapB sequence, using a linker (SIPGA) and *BamHI* and *EcoRI* restriction sites.

For the production of soluble and His-tagged ZapB, GFP and mCherry proteins, transformed *E. coli* BL21 (Invitrogen, USA) cells were grown aerobically in Luria Broth (LB) medium supplemented with 100 μg/ml of ampicillin. Protein expression was induced with 1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) at 30 °C for 12 h. For protein purification, cells expressing the recombinant protein were harvested by centrifugation (5000*g* for 20 min). After cell lysis by sonication and fractionation, the soluble fraction was collected and injected in a Histrap FF 5 mL column (GE Healthcare, USA) using an ÄKTA (GE Healthcare, USA). After purification, proteins were dialyzed in buffer A (50 mM Tris, 100 mM NaCl, pH 7.4). The purity of these proteins was checked by SDS-PAGE.

For the production of ZapB, ZapB-GFP, ZapB-mCherry and Aβ42-GFP IBs, transformed *E. coli* BL21 (Invitrogen, USA) cells were grown aerobically in LB medium supplemented with 100 μg/mL of ampicillin for ZapB, ZapB-GFP and ZapB-mCherry transformed cells, and with 50 μg/mL of kanamycin for Aβ42-GFP transformed cells. Protein expression was induced with 1 mM IPTG at 30 °C for 12 h. For IBs purification, cells were centrifuged at 5000*g* for 20 min. After that, IBs were purified as described [89]. Briefly, pelleted cells were resuspended in 400 μL of buffer A (50 mM Tris, 100 mM NaCl, pH 7.4) plus 4 μL of 100 mM PMSF and 6 μL of 10 mg/mL lysozyme. After 1 h of incubation at 37 °C, the mixture

was cooled in ice and sonicated 3 min at 15% amplitude under 1 s cycles. After that, 4 μL of Nonidet P40 (NP-40) were added and the mixture incubated at 4 °C for 1 h. Then, 10 μL of 1 mg/mL DNase I and 12 μL of 1 M MgSO$_4$ were added and the mixture was further incubated at 37 °C for 45 min. IBs were collected by centrifugation at 15.000*g* for 15 min at 4 °C and washed with buffer A (50 mM Tris, 100 mM NaCl, pH 7.4) containing 0.5% Triton X-100. All incubations were done under gentle agitation. Finally, IBs were washed three times with buffer A (50 mM Tris, 100 mM NaCl, pH 7.4) to remove remaining detergent. The purity of the IBs was checked by SDS-PAGE. Protein concentration was estimated measuring the absorbance at 280 nm in a Specord 200 Plus spectrophotometer (Analytik Jena, Germany).

Circular dichroism (CD) spectroscopy

For the analysis of the secondary structure, ZapB protein, ZapB and ZapB-GFP IBs were exhaustively resuspended in buffer A (50 mM Tris, 100 mM NaCl, pH 7.4). Soluble GFP was diluted in buffer A (50 mM Tris, 100 mM NaCl, pH 7.4) at 5 μM. Sedimentation problems were avoided with thorough resuspension of the sample before the measurements. The correct resuspension of the sample was checked before and after of each measurement. Far-UV CD spectra of the different protein solutions were recorded using a 1 nm bandwith, a response time of 1 s, 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.

Transmission electron microscopy (TEM)

For TEM sample preparation, 10 μL of 10 μM of ZapB protein were deposited onto a carbon-coated copper grid for 10 min and the excess of liquid was removed with filter paper, followed by a negative stain with 10 μL of 2% (w/v) uranyl acetate for 1 min. 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 fibers was analyzed using ImageJ, averaging the measures of 5 individual fibers.

Scanning electron microscopy (SEM) and dynamic light scattering (DLS)

Scanning electron microscopy (SEM) was used in order to analyze the morphology of ZapB and ZapB-GFP IBs. To do that, 10 μL of sample resuspended in water were deposited on silicon wafers (Ted Pella Inc., USA), air-dried and observed using a SEM Merlin (Zeiss Merlin, Germany) operating at 2 kV.

Gil-Garcia *et al. Microb Cell Fact*   (2020) 19:117

Page 13 of 16

Dynamic light scattering (DLS) was used for a quantitative determination of ZapB-GFP and Aβ42-GFP IBs size. The size of these nanoparticles was determined using a Zetasizer Nano ZS (Malvern Instruments Limited, UK) at 25 °C. Three different measures of ten runs were recorded for each sample.

Cell fractionation

The distribution of the expressed ZapB, GFP, mCherry and the fusion proteins (ZapB-GFP, ZapB-mCherry and Aβ42-GFP) in *E. coli* cells was analyzed by SDS-PAGE. After protein expression at 30 °C for 12 h, cells were harvested by centrifugation (5000g for 20 min) and resuspended in buffer A (50 mM Tris, 100 mM NaCl, pH 7.4). After that, disrupted cells (total fraction) by sonication were centrifuged at 15.000g for 15 min at 4 °C, and supernatant (soluble fraction) was separated from pellet (insoluble fraction). The insoluble fraction was resuspended in the same volume of buffer A (50 mM Tris, 100 mM NaCl, pH 7.4) than the soluble fraction and the different fractions were heated at 98 °C for 10 min. After that, 10 μL of each fraction were loaded separately into SDS-PAGE. Band intensity quantification was performed using the ImageJ software in order to estimate the percentage of protein in each fraction.

Confocal microscopy

E. coli BL21 cells expressing ZapB-GFP and ZapB-mCherry proteins at 30 °C for 12 h were centrifuged and resuspended in PBS pH 7.4 to an OD of 0.1. 10 μL of resuspended cells were deposited on top of microscopy poly-L-lysine glass slides, covered with coverslips and observed in a Leica SP5 confocal fluorescence microscope (Leica Microsystems, Germany).

Epifluorescence microscopy

10 μL of the ZapB-GFP, ZapB-mCherry and Aβ42-GFP IB samples in buffer A (50 mM Tris, 100 mM NaCl, pH 7.4) were deposited on top of microscopy glass slides and covered with coverslips. The fluorescence was observed using an Eclipse Ts2R-FL inverted microscope (Nikon, Japan) using a C-LED470 filter for GFP fluorescence and a C-LED525 for mCherry fluorescence.

To determine the GFP-fluorescence intensity of ZapB-GFP and Aβ42-GFP IBs, images were analyzed using the ImageJ software. 50 fluorescent dots were selected in each image maintaining the same dimensions (height × width) for the different selections. After that, the intensity of these different fluorescent dots was calculated and the average and SEM values were estimated.

GFP and mCherry fluorescence spectra

Excitation and emission spectra of soluble GFP, ZapB-GFP IBs and Aβ42-GFP IBs in buffer A (50 mM Tris, 100 mM NaCl, pH 7.4) were analyzed in a Jasco FP-8200 fluorescence spectrofluorometer (Jasco Corporation, Japan). Emission spectra were obtained recording the emitted fluorescence between 500 and 600 nm. Excitation spectra were obtained by exciting the samples in a 400–500 nm range.

Excitation and emission spectra of soluble mCherry and ZapB-mCherry IBs in buffer A (50 mM Tris, 100 mM NaCl, pH 7.4) were analyzed using Jasco FP-8200 fluorescence spectrofluorometer (Jasco Corporation, Japan). Emission spectra were obtained recording the emitted fluorescence between 600 and 700 nm. Excitation spectra were obtained by exciting the samples in a 500–600 nm. Three spectra were accumulated at 25 °C with slit widths of 5 nm, a 0.5 nm Interval, and a 1000 nm/min scan rate for each sample.

Fourier transform infrared spectroscopy (FTIR)

Samples of ZapB and ZapB-GFP IBs were washed with H_2O to remove the presence of salts. Both 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}. Data were acquired and normalized using the OPUS MIR Tensor 27 software (Bruker Optics, USA). IR spectrum was fitted employing a nonlinear peak-fitting equation using PeakFit package v4.12 (Systat Software, USA). The area for each Gaussian curve was calculated in the amide I region from 1700 to 1600 cm^{-1} using second derivative deconvolution method in PeakFit package v4.12 (Systat Software, USA).

Toxicity assay

HeLa cells were cultured in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% Fetal Bovine Serum (FBS) and seeded into 96-well plates. ZapB-GFP and Aβ42-GFP IBs were resuspended in PBS pH 7.4 and added at a range from 2 to 12 μM. For control, the same volume of PBS pH 7.4 was added. Treated and control cells were incubated for 72 h at 37 °C, and then 10 μL of PrestoBlue® reagent (ThermoFisher Scientific, USA) was added and incubated for 10 min. To determine cell viability, fluorescence signal was measured by exciting at 560 nm and collecting at 590 nm in a Victor3 fluorescent plate reader (Perkin Elmer, USA).

Gil-Garcia *et al. Microb Cell Fact*   (2020) 19:117

Page 14 of 16

Thioflavin T (Th-T) and congo red (CR) binding

For the Th-T binding assay, ZapB, mCherry. ZapB and ZapB-mCherry IBs, were diluted in buffer A (50 mM Tris, 100 mM NaCl, pH 7.4) and incubated with 25 µM Th-T. Emission fluorescence was recorded using a Jasco FP-8200 spectrofluorometer (Jasco Corporation, Japan) in the 460–600 nm range, using an excitation wavelength of 440 nm and an emission bandwith of 5 nm. The same buffer with 25 µM Th-T and without protein was employed as a control. In the case of ZapB-GFP, Th-T binding assay was not performed due to an overlap between the fluorescence spectra of GFP and Th-T.

For the CR binding assay the different IBs and soluble proteins were diluted in buffer A (50 mM Tris, 100 mM NaCl, pH 7.4) and mixed with CR to a final concentration of 10 µM CR. Optical absorption spectrum was recorded in the range from 375 to 700 nm in a Specord 200 Plus spectrophotometer (Analytik Jena, Germany). Spectrum of protein alone was acquired to subtract protein scattering.

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s12934-020-01375-4.

Additional file 1: Figure S1. Coiled-coil predictions for ZapB. **Figure S2.** Coiled-coil predictions for 3HAMP. **Figure S3.** Coiled-coil predictions for TDoT. **Figure S4.** Secondary structure prediction by PSIPRED server. **Figure S5.** Secondary structure prediction by GOR server. **Figure S6.** AGGRESCAN3D structural aggregation propensity predictions for ZapB, TDoT and 3HAMP. **Figure S7.** Net charge per residue (NCPR) of the different tags. **Figure S8.** SDS-PAGE of the cellular distribution of ZapB. **Figure S9.** SDS-PAGE of ZapB purification y IMAC. **Figure S10.** SDS-PAGE of purified ZapB IB . **Figure S11.** SDS-PAGE of the cellular distribution of GFP. **Figure S12.** SDS-PAGE of the cellular distribution of mCherry. **Figure S13.** AGGRESCAN3D structural aggregation propensity predictions for GFP and mCherry. **Figure S14.** SDS-PAGE of purified Zap -GFP IBs. **Figure S15.** SDS-PAGE of purified Zap -mCherry IBs. **Figure S16.** Characterization of the non-amyloid nature of ZapB-mCherry IBs. **Figure S17.** SDS-PAGE of the cellular distribution of Aβ42-GFP. **Figure S18.** SDS-PAGE of purified Aβ42-GFP IBs. **Figure S19.** DLS spectra of ZapB-GFP and Aβ42-GFP IBs. **Figure S20.** Epifluo escence microscopy images of ZapB-GFP and Aβ42-GFP IBs. DNA and amino acid sequences of ZapB protein.

Availability of data and materials
All data generated and analyzed during this study are shown in this article and it Additional fil 1.

Ethics approval and consent to participate
Not applicable.

Consent for publication
We state that all the authors (Marcos Gil-Garcia, Susanna Navarro and Salvador Ventura) mutually agree that this research work should be submitted to Microbial Cell Factories; it is an original work of the above-mentioned authors; and that the article has not been published or submitted in any other peer-reviewed journal.

Competing interests
The authors declare that they have no competing interests.

References
1. Ferrer-Miralles N, Domingo-Espín J, Corchero J, Vázquez E, Villaverde A. Microbial factories for recombinant pharmaceuticals. Microb Cell Fact. 2009;8:1–8.
2. Rosano GL, Ceccarelli EA. Recombinant protein expression in *Escherichia coli*: advances and challenges. Front Microbiol. 2014;5:1–17.
3. Schrödel A, De Marco A. Characterization of the aggregates formed during recombinant protein expression in bacteria. BMC Biochem. 2005;6:1–11.
4. De Marco A. Recombinant polypeptide production in *E. coli*: towards a rational approach to improve the yields of functional proteins. Microb Cell Fact. 2013;12:1–8.
5. Tartaglia GG, Pechmann S, Dobson CM, Vendruscolo M. Life ond the edge: a link between gene expression levels and aggregation rates of human proteins. Trends Biochem Sci. 2007;32:204–6.
6. Villaverde A, Carrio MM. Protein aggregation in recombinant bacteria: biological role of inclusion bodies. Biotechnol Lett. 2003;25:1385–95.
7. Slouka C, Kopp J, Spadiut O, Herwig C. Perspectives of inclusion bodies for bio-based products: curse or blessing? Appl microbiol biotechnol. 2019;103:1143–53.
8. Rinas U, Garcia-Fruitós E, Corchero JL, Vázquez E, Seras-Franzoso J, Villaverde A. Bacterial inclusion bodies: discovering their better half. Trends Biochem Sci. 2017;42:726–37.
9. García-Fruitós E, Rodríguez-Carmona E, Díez-Gil C, Ferraz RM, Vázquez E, Corchero JL, et al. Surface cell growth engineering assisted by a novel bacterial nanomaterial. Adv Mater. 2009;21:4249–53.
10. García-Fruitós E, Seras-Franzoso J, Vazquez E, Villaverde A. Tunable geometry of bacterial inclusion bodies as substrate materials for tissue engineering. Nanotechnology. 2010;21:205101.
11. De Marco A, Ferrer-Miralles N, Garcia-Fruitós E, Mitraki A, Peternel S, Rinas U, et al. Bacterial inclusion bodies are industrially exploitable amyloids. FEMS Microbiol Rev. 2019;43:53–72.
12. Margreiter G, Messner P, Caldwell KD, Bayer K. Size characterization of inclusion bodies by sedimentation fiel - fl w fractionation. J Biotechnol. 2015;138:67–73.
13. Rinas U, Bailey JE. Protein compositional analysis of inclusion bodies produced in recombinant Escherichia coli. Appl Microbiol Biotechnol. 1992;37:609–14.
14. Carrió MM, Villaverde A. Construction and deconstruction of bacterial inclusion bodies. J Biotechnol. 2002;96:3–12.
15. Fahnert B, Lilie H, Neubauer P. Inclusion bodies: formation and utilisation. Adv Biochem Eng Biotechnol. 2004;89:93–142.
16. Jäger VD, Lamm R, Kloß R, Kaganovitch E, Grünberger A, Pohl M, et al. A synthetic reaction cascade implemented by colocalization of two proteins within catalytically active inclusion bodies. ACS Synth Biol. 2018;7:2282–95.
17. Krauss U, Jäger VD, Diener M, Pohl M, Jaeger KE. Catalytically-active inclusion bodies—carrier-free protein immobilizates for application in biotechnology and biomedicine. J Biotechnol. 2017;258:136–47.

Gil-Garcia *et al. Microb Cell Fact*   (2020) 19:117

Page 15 of 16

18. Pesarrodona M, Jauset T, Díaz-Riascos ZV, Sánchez-Chardi A, Beaulieu ME, Seras-Franzoso J, et al. Targeting antitumoral proteins to breast cancer by local administration of functional inclusion bodies. Adv Sci. 2019;6:1900849.

19. Céspedes MV, Cano-Garrido O, Álamo P, Sala R, Gallardo A, Serna N, et al. Engineering secretory amyloids for remote and highly selective destruction of metastatic foci. Adv Mater. 2019;1907348:1–9.

20. Ventura S, Villaverde A. Protein quality in bacterial inclusion bodies. Trends Biotechnol. 2006;24:179–85.

21. Carrió M, González-Montalbán N, Vera A, Villaverde A, Ventura S. Amyloid-like properties of bacterial inclusion bodies. J Mol Biol. 2005;347:1025–37.

22. Morell M, Bravo R, Espargaró A, Sisquella X, Avilés FX, Fernàndez-Busquets X, et al. Inclusion bodies: specificity in their aggregation process and amyloid-like structure. Biochim Biophys Acta Mol Cell Res. 2008;1783:1815–25.

23. García-Fruitós E, González-Montalbán N, Morell M, Vera A, Ferraz RM, Arís A, et al. Aggregation as bacterial inclusion bodies does not imply inactivation of enzymes and fluorescent proteins. Microb Cell Fact BioMed Central. 2005;4:1–6.

24. García-Fruitós E, Sabate R, De Groot NS, Villaverde A, Ventura S. Biological role of bacterial inclusion bodies: a model for amyloid aggregation. FEBS J. 2011;278:2419–27.

25. Cano-Garrido O, Rodríguez-Carmona E, Díez-Gil C, Vázquez E, Elizondo E, Cubarsi R, et al. Supramolecular organization of protein-releasing functional amyloids solved in bacterial inclusion bodies. Acta Biomater. 2013;9:6134–42.

26. Fowler DM, Koulov AV, Balch WE, Kelly JW. Functional amyloid—from bacteria to humans. Trends Biochem Sci. 2007;32:217–24.

27. Fowler DM, Koulov AV, Alory-Jost C, Marks MS, Balch WE, Kelly JW. Functional amyloid formation within mammalian tissue. PLoS Biol. 2006;4:0100–7.

28. Maji SK, Perrin MH, Sawaya MR, Jessberger S, Vadodaria K, Rissman RA, et al. Functional amyloids as natural storage of peptide hormones in pituitary secretory granules. Science. 2009;325:328–32.

29. Jackson MP, Hewitt EW. Why are functional amyloids non-toxic in humans? Biomolecules. 2017;7:1–13.

30. Berson JF, Theos AC, Harper DC, Tenza D, Raposo G, Marks MS. Pro-protein convertase cleavage liberates a fibrillogenic fragment of a resident glycoprotein to initiate melanosome biogenesis. J Cell Biol. 2003;161:521–33.

31. Guyonnet B, Egge N, Cornwall GA. Functional amyloids in the mouse sperm acrosome. Mol Cell Biol. 2014;34:2624–34.

32. Huang PS, Oberdorfer G, Xu C, Pei XY, Nannenga BL, Rogers JM, et al. High thermodynamic stability of parametrically designed helical bundles. Science. 2014;346:481–5.

33. Fletcher JM, Harniman RL, Barnes FRH, Boyle AL, Collins A, Mantell J, et al. Self-assembling cages from coiled-coil peptide modules. Science. 2013;340:595–9.

34. Ing NL, Spencer RK, Luong SH, Nguyen HD, Hochbaum AI. Electronic conductivity in biomimetic α-helical peptide nanofibers and gels. ACS Nano. 2018;12:2652–61.

35. Gradišar H, Božič S, Doles T, Vengust D, Hafner-Bratkovič I, Mertelj A, et al. Design of a single-chain polypeptide tetrahedron assembled from coiled-coil segments. Nat Chem Biol. 2013;9:362–6.

36. Ljubetič A, Lapenta F, Gradišar H, Drobnak I, Aupič J, Strmšek Ž, et al. Design of coiled-coil protein-origami cages that self-assemble in vitro and in vivo. Nat Biotechnol. 2017;35:1094–101.

37. Jäger VD, Kloss R, Grünberger A, Seide S, Hahn D, Karmainski T, et al. Tailoring the properties of (catalytically)-active inclusion bodies. Microb Cell Fact. BioMed Central. 2019;18:1–20.

38. Diener M, Kopka B, Pohl M, Jaeger KE, Krauss U. Fusion of a coiled-coil domain facilitates the high-level production of catalytically active enzyme inclusion bodies. ChemCatChem. 2016;8:142–52.

39. Lamm R, Jäger VD, Heyman B, Berg C, Cürten C, Krauss U, et al. Detailed small-scale characterization and scale-up of active YFP inclusion body production with Escherichia coli induced by a tetrameric coiled coil domain. Elsevier Ltd: J Biosci Bioeng; 2020.

40. Kloss R, Limberg MH, Mackfeld U, Hahn D, Grünberger A, Jäger VD, et al. Catalytically active inclusion bodies of L-lysine decarboxylase from E. coli for 1,5-diaminopentane production. Sci Rep. 2018;8:1–11.

41. Kloss R, Karmainski T, Jäger VD, Hahn D, Grünberger A, Baumgart M, et al. Tailor-made catalytically active inclusion bodies for different applications in biocatalysis. Catal Sci Technol. 2018;8:5816–26.

42. Ebersbach G, Galli E, Møller-Jensen J, Löwe J, Gerdes K. Novel coiled-coil cell division factor ZapB stimulates Z ring assembly and cell division. Mol Microbiol. 2008;68:720–35.

43. Den Blaauwen T. Prokaryotic cell division: flexible and diverse. Curr Opin Microbiol. 2013;16:738–44.

44. Wang X, Zhou B, Hu W, Zhao Q, Lin Z. Formation of active inclusion bodies induced by hydrophobic self-assembling peptide GFIL8. Microb Cell Fact. 2015;14:1–8.

45. Rueda F, Gasser B, Sánchez-Chardi A, Roldán M, Villegas S, Puxbaum V, et al. Functional inclusion bodies produced in the yeast Pichia pastoris. Microb Cell Fact. 2016;15:166.

46. Pelassa I, Corà D, Cesano F, Monje FJ, Montarolo PG, Fiumara F. Association of polyalanine and polyglutamine coiled coils mediates expansion disease-related protein aggregation and dysfunction. Hum Mol Genet. 2014;23:3402–20.

47. Päiviö A, Nordling E, Kallberg Y, Thyberg J, Johansson J. Stabilization of discordant helices in amyloid fibril-forming proteins. Protein Sci. 2004;13:1251–9.

48. Brandenburg E, Von Berlepsch H, Gerling UIM, Böttcher C, Koksch B. Inhibition of amyloid aggregation by formation of helical assemblies. Chem A Eur J. 2011;17:10651–61.

49. Lupas A, Van Dyke M, Stock J. Predicting coiled coils from protein sequences. Science. 1991;252:1162–4.

50. Gruber M, Söding J, Lupas AN. Comparative analysis of coiled-coil prediction methods. J Struct Biol. 2006;155:140–5.

51. Delorenzi M, Speed T. An HMM model for coiled-coil domains and a comparison with PSSM-based predictions. Bioinformatics. 2002;18:617–25.

52. Ludwiczak J, Winski A, Szczepaniak K, Alva V, Dunin-Horkawicz S. DeepCoil-a fast and accurate prediction of coiled-coil domains in protein sequences. Bioinformatics. 2019;35:2790–5.

53. Villar-Piqué A, Espargaró A, Sabaté R, de Groot NS, Ventura S. Using bacterial inclusion bodies to screen for amyloid aggregation inhibitors. Microb Cell Fact. 2012;11:1–11.

54. De Groot NS, Aviles FX, Vendrell J, Ventura S. Mutagenesis of the central hydrophobic cluster in Aβ42 Alzheimer's peptide: side-chain properties correlate with aggregation propensities. FEBS J. 2006;273:658–68.

55. García-Fruitós E, Arís A, Villaverde A. Localization of functional polypeptides in bacterial inclusion bodies. Appl Environ Microbiol. 2007;73:289–94.

56. Jong WSP, Vikström D, Houben D, Berg van Saparoea HB, Gier JW, Luirink J. Application of an E. coli signal sequence as a versatile inclusion body tag. Microb Cell Fact. 2017;16:1–13.

57. Jong WSP, ten Hagen-Jongman CM, Vikström D, Dontje W, Abdallah AM, de Gier JW, et al. Mutagenesis-based characterization and improvement of a novel inclusion body tag. Front Bioeng Biotechnol. 2020;7:1–13.

58. Conchillo-Solé O, de Groot NS, Avilés FX, Vendrell J, Daura X, Ventura S. AGGRESCAN: a server for the prediction and evaluation of "hot spots" of aggregation in polypeptides. BMC Bioinform. 2007;8:65.

59. Fernandez-Escamilla AM, Rousseau F, Schymkowitz J, Serrano L. Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. Nat Biotechnol. 2004;22:1302–6.

60. Chiti F, Calamai M, Taddei N, Stefani M, Ramponi G, Dobson CM. Studies of the aggregation of mutant proteins in vitro provide insights into the genetics of amyloid diseases. Proc Natl Acad Sci USA. 2002;99:16419–26.

61. Chiti F, Stefani M, Taddei N, Ramponi G, Dobson CM. Rationalization of the effects of mutations on peptide and protein aggregation rates. Nature. 2003;424:805–8.

62. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. J Mol Biol. 1982;157:105–32.

63. Kallberg Y, Gustafsson M, Persson B, Thyberg J, Johansson J. Prediction of amyloid fibril-forming proteins. J Biol Chem. 2001;276:12945–50.

64. Gendoo DM, Harrison PM. Discordant and chameleon sequences: their distribution and implications for amyloidogenicity. Protein Sci. 2011;20:567–79.

65. McGuffin LJ, Bryson K, Jones DT. The PSIPRED protein structure prediction server. Bioinformatics. 2000;16:404–5.

Gil-Garcia *et al. Microb Cell Fact*   (2020) 19:117

Page 16 of 16

66. Garnier J, Osguthorpe DJ, Robson B. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J Mol Biol. 1978;120:97–120.

67. Gil-Garcia M, Banó-Polo M, Varejao N, Jamroz M, Kuriata A, Díaz-Caballero M, et al. Combining structural aggregation propensity and stability predictions to redesign protein solubility. Mol Pharm. 2018;15:3846–59.

68. Zambrano R, Jamroz M, Szczasiuk A, Pujols J, Kmiecik S, Ventura S. AGGRESCAN3D (A3D): server for prediction of aggregation properties of protein structures. Nucleic Acids Res. 2015;43:W306–13.

69. Kuriata A, Iglesias V, Kurcinski M, Ventura S, Kmiecik S. Aggrescan3D standalone package for structure-based prediction of protein aggregation properties. Bioinformatics. 2019;35:3834–5.

70. Kuriata A, Iglesias V, Pujols J, Kurcinski M, Kmiecik S, Ventura S. Aggrescan3D (A3D) 2.0: prediction and engineering of protein solubility. Nucleic Acids Res. 2019;47:W300–7.

71. Tedeschi G, Mangiagalli M, Chmielewska S, Lotti M, Natalello A, Brocca S. Aggregation properties of a disordered protein are tunable by pH and depend on its net charge per residue. Biochim Biophys Acta. 2017;1861:2543–50.

72. Zhou NE, Zhu B-Y, Kay CM, Hodges RS. The two-stranded α-helical coiled-coil is an ideal model for studying protein stability and subunit interactions. Biopolymers. 1992;32:419–26.

73. Zhou NE, Kay CM, Hodges RS. Synthetic model proteins: the relative contribution of leucine residues at the nonequivalent positions of the 3–4 hydrophobic repeat to the stability of the two-stranded α-helical coiled-coil. Biochemistry. 1992;31:5739–46.

74. Reisdorf WC, Krimm S. Infrared amide I′band of the coiled coil. Biochemistry. 1996;35:1383–6.

75. Heimburg T, Schünemann J, Weber K, Geisler N. FTIR-spectroscopy of multistranded coiled coil proteins. Biochemistry. 1999;38:12727–34.

76. Manas ES, Getahun Z, Wright WW, DeGrado WF, Vanderkooi JM. Infrared spectra of amide groups in α-helical proteins: evidence for hydrogen bonding between helices and water. J Am Chem Soc. 2000;122:9883–90.

77. Deliolanis NC, Kasmieh R, Wurdinger T, Tannous BA, Shah K, Ntziachristos V. Performance of the red-shifted fluo escent proteins in deep-tissue molecular imaging applications. J Biomed Opt. 2008;13:044008.

78. Shaner NC, Steinbach PA, Tsien RY. A guide to choosing fluo escent proteins. Nat Methods. 2005;2:905–9.

79. Shaner NC, Patterson GH, Davidson MW. Advances in fluo escent protein technology. J Cell Sci. 2007;120:4247–60.

80. De Groot NS, Ventura S. Effect of temperature on protein quality in bacterial inclusion bodies. FEBS Lett. 2006;580:6471–6.

81. De Groot NS, Sabate R, Ventura S. Amyloids in bacterial inclusion bodies. Trends Biochem Sci. 2009;34:408–16.

82. de Groot NS, Ventura S. Protein activity in bacterial inclusion bodies correlates with predicted aggregation rates. J Biotechnol. 2006;125:110–3.

83. Lupas AN, Gruber M. The structure of α-helical coiled coils. Adv Prot Chem. 2005;70:37–78.

84. Dasari M, Espargaro A, Sabate R, Lopez Del Amo JM, Fink U, Grelle G, et al. Bacterial inclusion bodies of alzheimer's disease β-amyloid peptides can be employed to study native-like aggregation intermediate states. ChemBioChem. 2011;12:407–23.

85. Gifre-Renom L, Seras-Franzoso J, Rafael D, Andrade F, Cano-Garrido O, Martinez-Trucharte F, et al. The biological potential hidden in inclusion bodies. Pharmaceutics. 2020;12:157.

86. Wang R, Li J, Dang D, Hu J, Hu Y, Fan J. Bacterial production of maize and human serine racemases as partially active inclusion bodies for D-serine synthesis. Enzyme Microb Technol. 2020;137:109547.

87. Carratalá JV, Cano-Garrido O, Sánchez J, Membrado C, Pérez E, Conchillo-Solé O, et al. Aggregation-prone peptides modulate activity of bovine interferon gamma released from naturally occurring protein nanoparticles. N Biotechnol. 2020;57:11–9.

88. Céspedes MV, Fernández Y, Unzueta U, Mendoza R, Seras-Franzoso J, Sánchez-Chardi A, et al. Bacterial mimetics of endocrine secretory granules as immobilized in vivo depots for functional protein drugs. Sci Rep. 2016;6:1–10.

89. Rodríguez-Carmona E, Cano-Garrido O, Seras-Franzoso J, Villaverde A, García-Fruitós E. Isolation of cell-free bacterial inclusion bodies. Microb Cell Fact. 2010;9:1–9.

Publisher's Note

CHAPTER 4

Multifunctional Antibody-Conjugated Coiled-Coil Protein Nanoparticles for Selective Cell Targeting

Marcos Gil-Garcia and Salvador Ventura

Full length article

Multifunctional antibody-conjugated coiled-coil protein nanoparticles for selective cell targeting

Marcos Gil-Garcia, Salvador Ventura*

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

A B S T R A C T

Nanostructures decorated with antibodies (Abs) are applied in bioimaging and therapeutics. However, most covalent conjugation strategies affect Abs functionality. In this study, we aimed to create protein-based nanoparticles to which intact Abs can be attached through tight, specific, and noncovalent interactions. Initially considered waste products, bacterial inclusion bodies (IBs) have been used in biotechnology and biomedicine. However, the amyloid-like nature of IBs limits their functionality and raises safety concerns. To bypass these obstacles, we have recently developed highly functional α-helix-rich IBs exploiting the natural self-assembly capacity of coiled-coil domains. We used this approach to create spherical, submicrometric, biocompatible and fluorescent protein nanoparticles capable of capturing Abs with high affinity. We showed that these IBs can be exploited for Ab-directed cell targeting. Simultaneous decoration of the nanoparticles with two different Abs in a controllable ratio enabled the construction of a bispecific antibody mimic that redirected T lymphocytes specifically to cancer cells. Overall, we describe an easy and cost-effective strategy to produce multivalent, traceable protein nanostructures with the potential to be used for biomedical applications.

Statement of significance

Functional inclusion bodies (IBs) are promising platforms for biomedical and biotechnological applications. These nanoparticles are usually sustained by amyloid-like interactions, which imposes some limitations on their use. In this work, we exploit the natural coiled-coil self-assembly properties to create highly functional, nonamyloid, and fluorescent IBs capable of capturing antibodies. These protein-based nanoparticles are successfully used to specifically and simultaneously target two unrelated cell types and bring them close together, becoming a technology with potential application in bioimaging and immunotherapy.

1. Introduction

The use of microorganisms as cell factories to produce recombinant proteins often leads to the formation of protein inclusion bodies (IBs) in the cytoplasm [1,2]. The formation of such protein-based nanoclusters is the consequence of an inside-the-cell assembly reaction consisting in the establishment of noncovalent intermolecular interactions between different copies of the recombinant protein (Fig. S1). For decades, these refractile aggregates have been considered waste products [3]. However, it has gradually been revealed that IBs might contain a significant degree of properly folded and functional proteins [4,5]. These nanometric protein particles are mechanically stable [6,7] and their production and purification are cost-effective and fully scalable, making them attractive and ready-to-use functional nanostructured materials [8–10]. In this regard, IBs are increasingly used in biotechnological and biomedical applications, i. e., as reusable biocatalysts [11–13] or in cancer therapy, since they can penetrate cells and release antitumoral polypeptides [14,15].

---

* Corresponding author.
E-mail addresses: marcos.gil.garcia@uab.cat (M. Gil-Garcia),
salvador.ventura@uab.cat (S. Ventura).

The mechanical stability of IBs relies on the amyloid-like nature of the interactions that stabilize these nanoparticles [16–18]. This amyloidogenic character raises two critical concerns. First, the formation of nonnative intermolecular β-sheet conformations comes at the expense of the native structure, implying that a significant fraction of the protein inside IBs becomes necessarily inactivated [18]. Second, together with the release of adequately folded moieties, the liberation of toxic oligomeric β-sheet species from these nanostructures cannot be disregarded. To offer an alternative to classical amyloid-like IBs, we have recently developed functional IBs exploiting the antiparallel coiled-coil fold of the ZapB protein. This domain promotes intracellular self-assembly into nonamyloid, nontoxic, α-helix-rich IBs. The native-like nature of the interactions that sustain these IBs may preclude the inactivation of ZapB-fused globular domains during the assembly reaction, resulting in highly active nanometric particles [19]. This high functionality seems to be a common property of coiled-coil stabilized IBs [11,13,20–22]. Active ZapB-based IBs can be produced in high yields and easily purified to homogeneity [19], thus they are amenable for use in preparative applications.

The globular nature of proteins confers the main advantage onto protein-based materials, allowing the alteration of material functionality by genetic redesign to fit the intended application [23]. We explored the potential of ZapB-based protein nanoparticles to generate functionalized nanostructures by creating a tripartite fusion protein consisting of ZapB, as the assembling unit, sequentially connected to the green fluorescent protein (GFP) and the Z-domain, as the active moieties (ZapB-GFP-Z). GFP is widely used for *in vivo* imaging [24], whereas the Z-domain [25] is an engineered analog of the B-domain of *Staphylococcus aureus* protein A with high affinity (in the nanomolar range) for IgGs [26–28]. The binding of this allα–helix protein to the Fc region of an IgG is mainly mediated by hydrophobic and noncovalent interactions involving residues located at helices 1 and 2 (Fig. S2) [29,30]. Considering this, the objective was to create traceable fluorescent submicrometric particles that can be easily decorated with an antibody of interest and directed to a specific cellular antigen.

Bispecific antibodies (BsAbs) are mAbs engineered to recognize and bind to two different epitopes simultaneously. BsAbs are increasingly used in cancer therapy since they can recognize targets in cancer cells and immune cells, helping to bring the two cell types together and facilitating tumor cell destruction [31–33]. However, the generalized use of these BsAbs in the clinic is hindered by the high costs and low yields of production and the short half-life of these molecules *in vivo* [34]. Dual-targeting nanoparticles conjugated to two different monoclonal antibodies (mAbs) [35] or binding proteins [36] are being developed to bypass these obstacles and generate new therapeutic applications. We show how ZapB-GFP-Z IBs behave as spherical and fluorescent protein nanoparticles that, when decorated with two different antibodies, promote selective interactions between different cell types, specifically redirecting T lymphocytes to tumoral cells.

2. Materials and methods

2.1. Protein production and purification

The ZapB-GFP-Z gene fragment (the DNA sequence is presented in the Supplementary Information) was cloned into a pET-28a vector between the NcoI and BamHI restriction sites.

For the production of ZapB-GFP and ZapB-GFP-Z IBs, *Escherichia coli* (*E. coli*) BL21 (DE3) (Invitrogen, USA) competent cells were transformed with the corresponding plasmids and grown aerobically in Luria-Bertani broth (LB) medium supplemented with 100 μg/mL ampicillin and 50 μg/mL kanamycin, respectively. After absorbance OD_{600} of 0.6 was reached, protein ex-

pression was induced by treatment with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 6 h at 30 °C. The cells were collected after centrifugation at 5000 g for 20 min and the pelleted cells were resuspended in 400 μL of buffer A (50 mM Tris and 100 mM NaCl, pH 7.4) supplemented with 4 μL of 100 mM PMSF and 6 μL of 10 mg/mL lysozyme. Then, the cells were incubated for 1 h at 37 °C, cooled on ice and sonicated for 3 min at 15% amplitude in 1 s cycles. 4 μL of Nonidet P40 (NP-40) was added to the sonicated cells, and the mixture was incubated for 1 h at 4 °C. Next, 12 μL of 1 M $MgSO_4$ and 10 μL of 1 mg/mL DNase I were added and the mixture was incubated for 45 min at 37 °C. The mixture was centrifuged at 15,000 g for 15 min at 4 °C, and the pellet was washed with buffer A containing 0.5% Triton X-100. All incubations were performed under agitation. Purified IBs were washed three times with buffer A, and the remaining detergent was removed. The purity of the IBs was confirmed by SDS-PAGE, and the protein concentration was estimated by measuring the absorbance of IBs dissolved in guanidine hydrochloride at 280 nm using a Specord 200 Plus spectrophotometer (Analytik Jena, Germany).

For the production of soluble and His-tagged Z-domain, *E. coli* BL21 (DE3) (Invitrogen, USA) competent cells were transformed with the corresponding plasmid and grown aerobically in LB medium supplemented with 50 μg/mL kanamycin. After an absorbance OD_{600} of 0.6 was reached, protein expression was induced by treatment with 1 mM IPTG for 16 h at 20 °C. The cells were collected after centrifugation at 5000 g for 20 min and resuspended in buffer A containing 20 mM imidazole and 1 mM PMSF. The solution was sonicated on ice, and the supernatant was collected after centrifugation at 15,000 g for 30 min, and the protein was purified by a 5 mL HisTrap FF column (GE Healthcare, USA) using an ÄKTA chromatograph (GE Healthcare, USA).

α-Synuclein was produced, purified and aggregated as previously described in Pujols et al. [37].

2.2. Cell fractionation

The distribution of the expressed ZapB-GFP-Z protein in *E. coli* was analyzed by SDS-PAGE. After protein expression, the cells were centrifuged at 5000 g for 20 min and resuspended in buffer A. Then, the cells were lysed by sonication and centrifuged at 15,000 g for 15 min at 4 °C, and the soluble fraction (supernatant) was separated from the insoluble fraction (pellet). The insoluble fraction was resuspended in buffer A to be the same volume as the soluble fraction, and both fractions were heated for 10 min at 98 °C. Finally, 10 μL of the different fractions were loaded onto SDS-PAGE gels.

2.3. Preparation of antibody-loaded IBs

For the preparation of the IgG-IB complexes, 5 μM ZapB-GFP-Z and 5 μM ZapB-GFP IBs were incubated in the presence of 1 μg of IgG for 45 min at room temperature. Due to the nature of the noncovalent but strong interaction between the Z-domain and the IgGs, the preparation of antibody-decorated IBs was performed by simple mixing. Then, the IBs were harvested by centrifugation at 15,000 g for 20 min and resuspended in buffer A. Next, five washing steps were performed to remove unbound antibody molecules and prevent nonspecific binding.

For the competition assay with ZapB-GFP-Z IBs and soluble Z-domain protein, 1 μg of IgG labeled with Alexa Fluor 555 was incubated in the presence of 5 μM ZapB-GFP-Z IBs, either with an excess (50 μM) or without soluble Z-domain. After incubation for 45 min at room temperature, the samples were centrifuged, and the presence of the IgGs labeled with Alexa Fluor 555 in the insoluble fraction (corresponding to the IBs) was analyzed by measuring

their fluorescence in a Jasco FP-8200 fluorescence spectrofluorometer (Jasco Corporation, Japan).

For the determination of the antibody-binding capacity of the ZapB-GFP-Z IBs, 0.5 μg of IgG labeled with Alexa Fluor 555 was incubated in the presence of increasing concentrations of ZapB-GFP-Z IBs (from 0.5 to 20 μM) for 45 min. Then, the GFP and Alexa Fluor 555 fluorescence of the IBs was measured using a Jasco FP-8200 fluorescence spectrofluorometer (Jasco Corporation, Japan).

2.4. Stability in bovine serum

ZapB-GFP-Z IBs were incubated in bovine serum at room temperature for 15 days. Then, samples obtained at different time points were loaded onto SDS-PAGE gels to evaluate the integrity of the ZapB-GFP-Z protein fusion.

2.5. Epifluorescence microscopy

Ten microliters of 5 μM decorated IBs in buffer A were deposited on top of microscopy glass slides and covered with coverslips. Fluorescence was observed using an Eclipse Ts2R-FL inverted microscope (Nikon, Japan) with a C-LED385 filter for eFluor 450 fluorescence (excitation (390/38 nm) and emission (475/90 nm)), C-LED470 filter for GFP fluorescence (excitation (470/40 nm) and emission (534/55 nm)) and C-LED525 filter for Alexa Fluor 555 fluorescence (excitation (525/50 nm) and emission (597/58 nm)).

2.6. Confocal microscopy

Ten microliters of the solution containing 5 μM ZapB-GFP-Z IBs decorated with two different IgGs labeled with eFluor 450 and Alexa Fluor 555 were deposited on top of microscopy glass slides and covered with coverslips. Then, the fluorescence intensity of the different fluorophores (eFluor 450, GFP and Alexa Fluor 555) was observed upon excitation at a specific wavelength for each fluorophore in a Leica SP5 confocal fluorescence microscope (Leica Microsystems, Germany). Undecorated ZapB-GFP-Z IBs at a concentration of 5 μM were used as negative controls, as no fluorescence signal corresponding to eFluor 450 or Alexa Fluor 555 fluorophores was observed.

2.7. Fourier transform infrared spectroscopy (FTIR)

ZapB-GFP-Z IBs at 5 μM were washed with H_2O to remove the presence of salts, placed on ATR crystals and dried under N_2 flow. The experiment was carried out in a Bruker Tensor 27 FTIR (Bruker Optics, USA) supplied with a Specac Golden Gate MKII ATR accessory. The spectrum consisted of 32 acquisitions measured at a resolution of 1 cm^{-1}. The data were acquired and normalized using OPUS MIR Tensor 27 software (Bruker Optics, USA). The IR spectrum was fitted employing the nonlinear peak-fitting equation in PeakFit package v4.12 (Systat Software, USA). The area for each Gaussian curve was calculated in the amide I region from 1700 to 1600 cm^{-1} using the second derivative deconvolution method with PeakFit package v4.12 (Systat Software, USA).

2.8. Circular dichroism (CD) spectroscopy

CD spectroscopy was used to analyze the secondary structure content of the ZapB-GFP-Z IBs. Specifically, 5 μM ZapB-GFP-Z IBs were resuspended in buffer A, and the far-UV CD spectrum was recorded at a 1 nm bandwidth, scan speed of 100 nm/min and response time of 1 s using a Jasco-815 spectropolarimeter (Jasco Corporation, Japan) at 25 °C.

2.9. Scanning electron microscopy (SEM) and dynamic light scattering (DLS)

Scanning electron microscopy (SEM) was used to analyze the morphology of the undecorated and antibody-decorated ZapB-GFP-Z IBs. Specifically, 10 μL of 5 μM IBs with and without antibodies resuspended in H_2O were deposited on silicon wafers (Ted Pella Inc., USA), air-dried and observed using a Merlin scanning electron microscope (Zeiss Merlin, Germany) operating at 2 kV.

Dynamic light scattering (DLS) was used for quantitative size determination of 5 μM ZapB-GFP-Z IBs undecorated or decorated with antibodies. For the determination of the stability of the ZapB-GFP-Z IBs, the size of the nanoparticles was monitored for 15 days. The size of the IBs was determined using a Zetasizer Nano S90 (Malvern Instruments Limited, UK) at 25 °C.

2.10. Congo red (CR) binding assay

A Congo red (CR) assay was performed to determine the presence of amyloid structures. For the CR assay, 5 μM ZapB-GFP-Z IB was resuspended in buffer A and mixed with CR to a final concentration of 10 μM CR. Aggregated α-synuclein was used as a positive control for protein amyloid aggregation. Optical absorption spectra were recorded in the range from 375 to 700 nm in a Specord 200 Plus spectrophotometer (Analytik Jena, Germany). The spectrum of the protein alone was acquired to subtract protein scattering.

2.11. Fluorescence spectra

The emission spectrum of Alexa Fluor 555 IgG was obtained by recording the emitted fluorescence between 560 and 700 nm using an excitation wavelength of 555 nm.

The emission spectrum of eFluor 450 IgG was obtained by recording the emitted fluorescence between 440 and 660 nm using an excitation wavelength of 410 nm.

The emission spectrum of the ZapB-GFP-Z IBs in buffer A was obtained by recording the emitted fluorescence between 500 and 600 nm. The excitation spectrum was obtained by exciting the samples at a range of 400–500 nm.

Three spectra were accumulated at 25 °C using a Jasco FP-8200 fluorescence spectrofluorometer (Jasco Corporation, Japan) with slit widths of 5 nm, a 0.5 nm interval, and a 1000 nm/min scan rate for each sample.

For the thermal unfolding of unloaded and IgG-loaded ZapB-GFP-Z IBs, the GFP fluorescence signal was recorded in a range of 25–95 °C with an increasing heat rate of 1 °C/min and slit widths of 5 nm and 600 rpm. ZapB-GFP-Z IBs were excited at 495 nm and emission was recorded at 514 nm using a Jasco FP-8200 fluorescence spectrofluorometer (Jasco Corporation, Japan).

2.12. Toxicity assay

HeLa cells and MRC-5 cells were seeded into 96-well plates and cultured for 24 h in Dulbecco's Modified Eagle Medium (DMEM) and Minimum Essential Medium α (MEM α), respectively, supplemented with 10% Fetal Bovine Serum (FBS). The ZapB-GFP-Z IBs were resuspended in PBS at pH 7.4 and added at a concentration ranging from 2 to 12 μM. PBS was used alone as a control.

For the evaluation of the potential antibody-associated ZapB-GFP-Z IBs toxicity, HeLa cells were seeded into 96-well plates and cultured in DMEM supplemented with 10% FBS for 24 h. ZapB-GFP-Z IBs loaded with anti-EGFR IgGs were resuspended in PBS at pH 7.4 and added at a concentration ranging from 2 to 12 μM. PBS alone and anti-EGFR IgGs were used as controls.

In the case of the potential toxicity induced by ZapB-GFP-Z IBs loaded with anti-CD3 IgGs, T cells (Jurkat cell line) were seeded

into 96-well plates cultured in Roswell Park Memorial Institute (RPMI) 1640 + GlutaMAX medium supplemented with 10% FBS for 24 h. ZapB-GFP-Z IBs decorated with anti-CD3 IgGs were resuspended in PBS at pH 7.4 and added at a concentration ranging from 2 to 12 µM. PBS alone, anti-CD3 IgGs and undecorated ZapB-GFP-Z IBs at 10 µM were used as controls.

Treated and control cells were incubated at 37 °C in 5% CO_2 for 72 h. Then, 10 µL of PrestoBlue cell viability reagent (Thermo Fisher Scientific) was added and incubated for 30 min. Cell viability was determined by quantifying the fluorescence with a Victor3 fluorescent plate reader (Perkin Elmer, USA).

2.13. Statistical methods

All experimental values are presented as the means ± standard error of the mean (SEM). Measurements are taken in duplicate. GraphPad Prism software (GraphPad Software, USA) was used to calculate the mean and SEM values.

2.14. Specific cell targeting by decorated IBs

HeLa cells were cultured on a 35 mm cell culture dish (SARST-EDT AG & Co, Germany) to a confluence of 70–80%. The initial DMEM was replaced with fresh medium containing 10 µM ZapB-GFP-Z-anti-EGFR IgG IB labeled with Alexa Fluor 555 and prepared as previously described in the "Preparation of antibody-loaded IBs" section. Then, HeLa cells were incubated with ZapB-GFP-Z-anti-EGFR IgG IBs for 1 h at 37 °C and 5% CO_2, and ZapB-GFP-Z-IgG anti-rabbit IBs were used as controls. The incubated cells were washed five times with fresh medium to remove unbound molecules.

T cells were maintained in RPMI 1640 + GlutaMAX medium supplemented with 10% FBS. T cells were harvested and resuspended in fresh medium containing 5 µM ZapB-GFP-Z-anti-CD3 IgG IB labeled with eFluor 450. The cells were incubated with ZapB-GFP-Z-anti-CD3 IgG IBs for 1 h at 37 °C and 5% CO_2, precipitated and washed five times with fresh medium. ZapB-GFP-Z-IgG anti-rabbit and anti-EGFR IBs were used as controls.

For the binding of unrelated cell types, HeLa cells were cultured on a 35 mm dish culture (SARSTEDT AG & Co, Germany) to a confluence of 70–80%. Then, the initial medium was replaced with T cells and dual antibody-decorated 10 µM ZapB-GFP-Z IBs (anti-EGFR labeled with Alexa Fluor 555 and anti-CD3 labeled with eFluor 450). The cells with antibody-decorated IBs were incubated for 1 h at 37 °C and 5% CO_2 and washed five times with fresh medium. IBs decorated with only anti-EGFR IgGs were used as controls.

Finally, the cells incubated with decorated IBs were visualized using an Eclipse Ts2R-FL inverted microscope (Nikon, Japan), and images were acquired using the appropriate filters and brightfield.

2.15. Flow cytometry assay

ZapB-GFP-Z IBs were decorated with anti-EGFR IgG labeled with Alexa Fluor 555 as previously described. Then, HeLa cells were incubated with the decorated IBs for 1 h, pelleted and washed five times to discard unbound molecules. HeLa cells incubated with ZapB-GFP IBs (previously incubated with anti-EGFR IgGs), ZapB-GFP-Z IBs decorated with anti-rabbit IgG, ZapB-GFP-Z IBs alone, and buffer alone were used as controls. For the T cell experiments, the procedure was the same as previously described for HeLa cells, decorating ZapB-GFP-Z IBs with anti-CD3 IgG. In this case, T cells incubated with ZapB-GFP IBs (previously incubated with anti-CD3 IgGs), ZapB-GFP-Z IBs decorated with anti-rabbit IgG, ZapB-GFP-Z IBs alone, and buffer alone were used as controls.

A total of 10,000 cells incubated with IBs were analyzed using a FACSCanto flow cytometer (BD Biosciences, USA) equipped with a FITC laser to detect the green fluorescence of GFP. Fluorescence intensities of cell-bound IBs were analyzed and quantified using FlowJo (BD Biosciences, USA).

For the evaluation of the localization of HeLa and T cells in the presence of dual-decorated (anti-EGFR and anti-CD3 IgGs) ZapB-GFP-Z IBs, HeLa cells were cultured on a 35 mm cell culture dish (SARSTEDT AG & Co, Germany) to a confluence of 70–80%. Then, the initial medium was replaced with T cells and dual antibody-decorated 10 µM ZapB-GFP-Z IBs (anti-EGFR and anti-CD3 IgG). Cells with decorated IBs were incubated for 1 h at 37 °C and 5% CO_2 and washed five times with fresh medium. ZapB-GFP-Z IBs decorated with only anti-EGFR IgGs were used as controls. Then, trypsin was added to the solution and neutralized with FBS-supplemented medium. Then, the cells were centrifuged, resuspended in PBS at pH 7.4 and analyzed using a FACSCanto flow cytometer (BD Biosciences, USA). FSC and SSC signals were measured for the detection of both cell types, and untreated HeLa and T cells were used alone as controls. Finally, the percentage and quantification of the cells were determined using FlowJo (BD Biosciences, USA).

3. Results

3.1. ZapB-GFP-Z forms nonamyloid α-helix-rich IBs

To generate nonamyloid fluorescent IBs with antibody capturing activity, ZapB was N-terminally fused to two consecutive globular domains: GFP and the Z-domain. This created a tripartite fusion protein named ZapB-GFP-Z. The Z-domain [25] is an engineered analog of the B domain of *Staphylococcus aureus* protein A, consists of 58 residues (6.6 kDa), and folds into three α-helices. Therefore, the fusion protein consisted of sequentially ordered coiled-coil, all-β, and all α- domains (Fig. S3).

ZapB-GFP-Z was expressed in *E. coli* and located in the insoluble cell fraction (Fig. S4), from which IBs composed of assembled copies of the fusion protein were purified to homogeneity (Fig. S5) at a yield of 98.5 ± 5.5 mg/L. The secondary structure of the IBs was analyzed using Fourier Transform Infrared Spectroscopy (FTIR). We recorded the infrared spectrum of the ZapB-GFP-Z IBs in the amide I region (1700–1600 cm^{-1}), corresponding to the absorption of the main chain carbonyl group. As shown in Fig. 1A, the primary contributor to the ZapB-GFP-Z IBs spectrum was a signal located at 1654 cm^{-1} (44% of the area), assigned to the α-helix conformation. This outcome is reasonable since the ZapB and Z-domains both consist of α-helices. A second band at 1631 cm^{-1} accounts for 35% of the spectrum area and likely represents the low-frequency coiled-coil signal [38] and the β-sheet signal of the GFP β-barrel. The third band at 1676 cm^{-1} (21% of the area) possibly is attributable to the mixed contribution of structurally different domains at high frequencies. Although the far-UV Circular Dichroism (CD) spectrum of the ZapB-GFP-Z IBs is influenced by the β-barrel structure of GFP, the two main α-helical signals are still evident, with a minimum at 224 nm and an inflection at 210 nm, in good agreement with the FTIR data (Fig. S6) and with the previous secondary structure analysis of ZapB IBs and ZapB-GFP IBs [19]. These results corroborate the native-like secondary structure of the ZapB-GFP-Z polypeptides embedded in these IBs.

Scanning Electron Microscopy (SEM) revealed that the ZapB-GFP-Z IBs were of submicrometric size and displayed a spherical shape (Fig. 1B). Dynamic light scattering (DLS) data revealed a moderately polydisperse size distribution with an average diameter of 560 ± 157 nm (Fig. S7) that remained stable in size for at least 15 days (Fig. S8). This submicrometric size is the result of the high order assembly of multiple copies of the ZapB-GFP-Z fusion protein inside the bacteria, already before their purification. To assess whether ZapB-GFP-Z IBs were stable under close to physiolog-
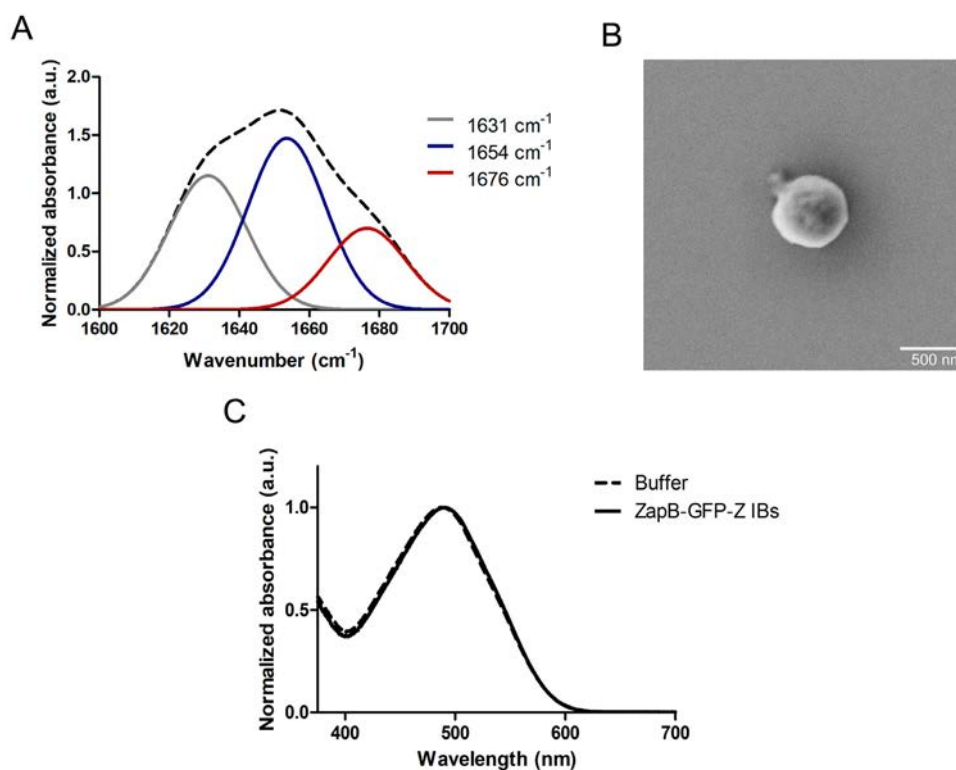
Fig. 1. Conformational properties of the ZapB-GFP-Z IBs. (A) FTIR absorbance spectrum of the ZapB-GFP-Z IBs in the amide I region (dashed line). The different component bands (solid lines) were obtained by deconvolution of the spectrum using PeakFit software. (B) SEM image of a ZapB-GFP-Z IB. (C) CR absorbance spectra of the ZapB-GFP-Z IBs (solid line) and buffer alone (dashed line).

ical conditions, they were incubated with bovine serum for up to 15 days. The SDS-PAGE analysis revealed that the fusion protein in these IBs remained stable for at least 10 days at room temperature (Fig. S9).

An amyloid-like nature for ZapB-GFP-Z IBs was excluded using one of the most common dyes to detect amyloid-like structures, the Congo red (CR) [39]. The CR absorption spectrum is sensitive to the presence of amyloid structures, displaying a redshift in their presence, caused by the cooperative binding of CR molecules along the fibril axis [40]. As shown in Fig. 1C, the spectra of CR dye in the presence and absence of ZapB-GFP-Z IBs were very similar. In contrast, when CR was incubated with aggregated α-synuclein, the displacement of the spectrum was evident (Fig. S10).

Finally, the thermal stability of the fusion protein within the ZapB-GFP-Z IBs was analyzed by monitoring the changes in the GFP fluorescence intensity upon heating. As shown in Fig. S11, the melting curve is indicative of cooperative GFP unfolding with a T_m of 74.5 ± 3.0 °C.

Overall, the biophysical analysis indicates that ZapB-GFP-Z IBs are α-helix-rich and stable nanometric spherical particles, likely sustained by coiled-coil interactions, and devoid of amyloid-like intermolecular β-sheet structures.

3.2. ZapB-GFP-Z IBs are bifunctional

The conformational traits of ZapB-GFP-Z IBs suggest that the different domains in the protein fusion might maintain their native structure in the assembled state. Therefore, they were expected to emit green fluorescence and bind antibodies simultaneously. To confirm that the spectral properties of the GFP embedded in the ZapB-GFP-Z IBs were preserved, the excitation and emission spectra were recorded. As shown in Fig. S12, the GFP located in the IBs displays the characteristic excitation and emission profiles, in

good agreement with our previous studies [19], with excitation and emission maxima at 495 nm and 514 nm, respectively. To confirm the expected bifunctional activity of these IBs, we incubated ZapB-GFP-Z IBs with an IgG antibody labeled with Alexa Fluor 555 at room temperature for 45 min. Then, the IBs were precipitated and washed five times to minimize nonspecific binding, and the fluorescence of these nanoparticles was analyzed by fluorescence microscopy. The ZapB-GFP-Z IBs emitted the characteristic GFP green fluorescence and the labeled antibody red fluorescence, with a perfect overlap between both signals (Fig. 2). When ZapB-GFP IBs without the Z-domain were used as controls, green fluorescence was evident, but we could not detect red fluorescence coming from the labeled IgG (Fig. S13), indicating that the Z-domain was essential for the binding of IgG to the IBs. The ability of the ZapB-GFP-Z IBs to bind antibodies was tested in the presence and in the absence of 10-fold molar excess of soluble Z-domain. As shown in Fig. 3A, the binding of these IBs to antibodies labeled with Alexa Fluor 555 was strongly reduced in the presence of the competing soluble Z-domain, as it was evident by the much lower fluorescence intensity. This indicates that, as intended, the binding of the antibodies to the IBs is mediated by the Z-domain.

To determine the IgG loading capacity of these IBs, we performed a titration assay by measuring the fluorescence of the labeled IgG bound to the IBs. As shown in Fig. 3B, the GFP fluorescence of these IBs increases linearly with the concentration. In contrast, the binding of the Alexa Fluor 555-labeled IgG reached a plateau at 2 μM ZapB-GFP-Z IB, revealing a conventional saturation binding curve, with an apparent binding capacity of 130 ng of antibody per μg of IB and 867 ng of antibody per μg of Z-domain in the IB (Fig. 3C).

We characterized the size and morphology of these nanoparticles when loaded with antibodies using DLS and SEM. The pres-
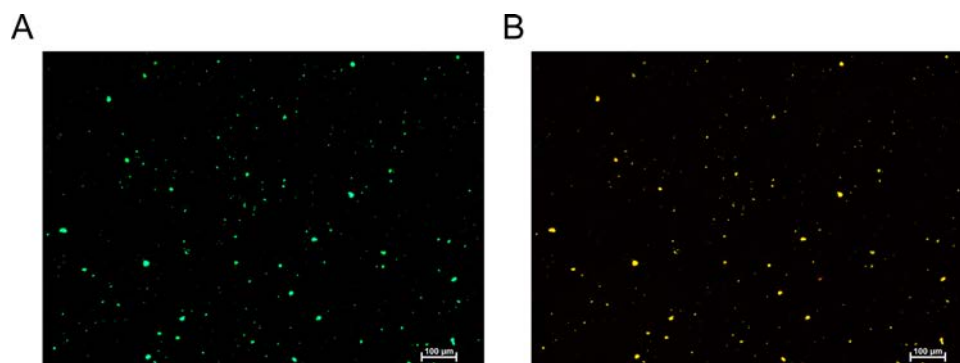
Fig. 2. Bifunctional activity of the ZapB-GFP-Z IBs. Representative fluorescence microscopy images of (A) GFP fluorescence and (B) IgG labeled with Alexa Fluor 555 fluorescence. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

ence of the IgGs increased the average diameter of the IBs to 724.8 ± 292.7 nm, while they kept the spherical shape (Fig. S14). The stability of the IBs constituent fusion proteins was not significantly altered by the presence of the IgGs, as it was deduced from the cooperative melting curve of GFP, with a Tm of 75.0 ± 2.1 °C, fairly similar to that of undecorated IBs (Fig. S11).

3.3. ZapB-GFP-Z IBs are biocompatible

Once the functionality of ZapB-GFP-Z IBs was confirmed, we sought to determine whether they induce toxicity in human cells since it would preclude any further cellular assay. HeLa cells were incubated with different concentrations of ZapB-GFP-Z IBs (from 0 to 12 μM), and we monitored their viability using a PrestoBlue® fluorescent assay. As it can be observed in Fig. 4A, ZapB-GFP-Z IBs did not show associated toxicity at any of the assayed concentrations. The IBs toxicity was further tested in the MRC-5 nontumor human lung fibroblast line. As shown in Fig. 4B, ZapB-GFP-Z IBs were innocuous to MRC-5 cells at all the assayed concentrations, corroborating that these functional protein-based nanoparticles have optimal biocompatibility.

3.4. Antibody-decorated ZapB-GFP-Z IBs specifically target human cancer cells

We aimed to assess whether ZapB-GFP-Z IBs can target specific antigens in living cells once they have been loaded with antibodies through their Z-domain. First, we decorated ZapB-GFP-Z IBs with an anti-EGFR antibody labeled with Alexa Fluor 555 as described above. Anti-EGFR antibodies recognize epidermal growth factor receptor (EGFR) overexpressed on the membrane of epithelial cancer cells such as HeLa cells [41,42]. We incubated the anti-EGFR antibody-loaded ZapB-GFP-Z IBs with HeLa cells. The presence of IBs attached to the cell membranes was corroborated using fluorescence and brightfield microscopy (Fig. 5). We observed a good overlay between green and red fluorescence signals, emitted from the GFP and the antibody, respectively, indicating that the nanoparticles maintained their bifunctionality in the assay. In contrast, when HeLa cells were incubated in the presence of ZapB-GFP-Z IBs loaded with an anti-rabbit antibody, IBs binding to cells was not detected (Fig. S15), indicating that the recognition of HeLa cells by IBs is specific and depends on the presence of anti-EGFR antibodies anchored to their Z-domains.

The binding of IBs to cells was also investigated using flow cytometry (FC). We incubated ZapB-GFP-Z IBs alone and loaded with either anti-EGFR or anti-rabbit antibodies with HeLa cells, followed by a washing step. The cells were immediately analyzed by FC, taking advantage of GFP fluorescence using a FITC emission detector. The analysis revealed that approximately 80% of the HeLa

cells were labeled with anti-EGFR decorated IBs (Fig. S16), whereas no significant labeling was observed in the other cases. Similarly, when ZapB-GFP IBs were incubated with the anti-EGFR antibody, and these IBs were added to HeLa cells, the cells remained unlabeled. Furthermore, when ZapB-GFP-Z IBs loaded with anti-EGFR antibody were incubated with cells deficient in EGFR expression, such as T lymphocytes, the IBs did not recognize them (Fig. S17).

Finally, we wanted to discard any antigen-specific toxicity of the antibody-decorated IBs. ZapB-GFP-Z IBs loaded with anti-EGFR antibody were incubated with HeLa cells. As shown in Fig. S18, these functionalized IBs did not induce any detectable antigen-associated toxicity in the HeLa cells, which displayed viability values similar to the cells treated with buffer or anti-EGFR soluble antibody.

Overall, we can conclude that ZapB-GFP-Z IBs can be used to target specific human cell epitopes and that this activity depends both on the presence of the Z-domain and the identity of the conjugated antibody, without any antigen-associated toxicity. The intrinsic green fluorescence of these nanoparticles allows to localize them easily without need for post-assembly IBs labeling.

3.5. Dual antibody binding by the ZapB-GFP-Z IBs

In principle, these bifunctional IBs can be endorsed with multivalency simply by loading different IgG antibodies onto the Z-domain. To confirm this possibility, ZapB-GFP-Z IBs were simultaneously incubated with two different antibodies labeled either with Alexa Fluor 555 (red fluorescence) or eFluor 450 (blue fluorescence). As shown in Fig. 6, ZapB-GFP-Z nanoparticles emitted three different fluorescence signals, green from GFP, and red and blue fluorescence from the antibodies, with a perfect overlap, indicating that both antibodies were bound to the IBs, as intended. This dual antibody-loading was confirmed using confocal microscopy. A good overlap of the three fluorescence signals in the dual-decorated IBs was observed, without detecting blue or red fluorescence signals in the undecorated ZapB-GFP-Z IBs (Fig. S19). Moreover, controlling the proportion of the antibodies in the initial mixture allowed us to obtain IBs decorated with a defined ratio of these molecules (Fig. S20).

Overall, the data indicate that ZapB-GFP-Z IBs can be functionalized with the desired proportion of two, and likely more, different antibodies.

3.6. Dual antibody-conjugated ZapB-GFP-Z IBs direct T lymphocytes to HeLa cells

Once we confirmed the ability of ZapB-GFP-Z IBs to capture different antibodies simultaneously, we wondered whether they can be used as scaffolds to bring two different cell types together. To this end, we first analyzed the ability of the ZapB-GFP-Z IBs to tar-
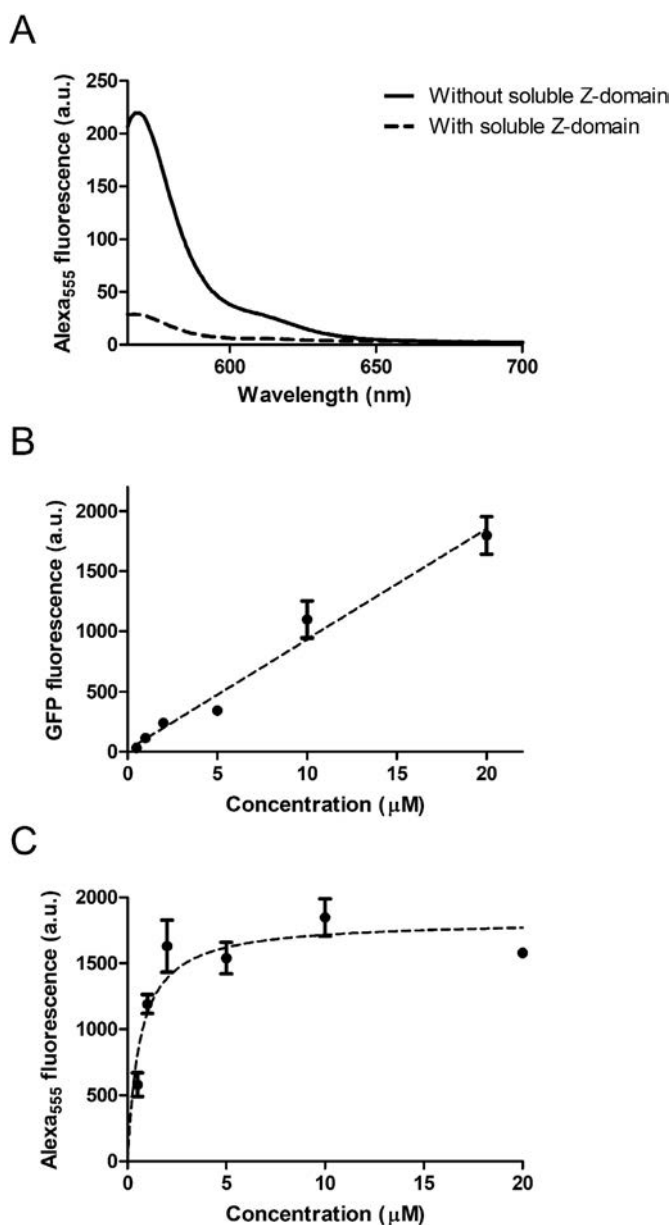
Fig. 3. Antibody-binding capacity of the ZapB–GFP–Z IBs. (A) Alexa Fluor 555 fluorescence intensity of ZapB–GFP–Z IBs incubated with labeled IgG in the presence and absence of excess of competing soluble Z-domain. (B) Fluorescence intensity of GFP of IgG-loaded ZapB–GFP–Z IBs as a function of the IBs concentration. (C) Alexa Fluor 555 fluorescence intensity of IgG-bound to ZapB–GFP–Z IBs as a function of the IBs concentration.

get T lymphocytes when they were decorated with anti-CD3 IgG labeled with eFluor 450. This anti-CD3 antibody recognizes the TCR/CD3 complex of T cells [43]. ZapB-GFP-Z IBs loaded with anti-CD3 IgG were incubated with T cells for one hour and washed to prevent nonspecific binding. As shown in Fig. 7, the presence of conjugated IBs attached to T cells was evident in the fluorescence and brightfield microscopy images. In addition, there was a perfect overlap between the green and blue fluorescence emitted from the GFP and anti-CD3 antibody, respectively. When T cells were incubated with ZapB-GFP-Z IBs loaded with anti-rabbit IgG, these nanoparticles did not recognize the T lymphocytes (Fig. S21).

The binding of IBs to T cells was also investigated using FC, as previously described with HeLa cells. As shown in Fig. S22, the analysis indicated that approximately 60% of T cells were labeled when incubated with ZapB-GFP-Z IBs conjugated to the anti-CD3 antibody. In contrast, T cells incubated with IBs loaded with non-specific antibodies, unloaded IBs or IBs incubated with the anti-CD3 antibody but devoid of the Z-domain remained unlabeled and were nonfluorescent.

To confirm that ZapB-GFP-Z IBs decorated with anti-CD3 antibody did not induce antigen-specific toxicity, T cells were incubated with different concentrations of anti-CD3-loaded IBs for 72 h. As is shown in Fig. S23, these functionalized IBs did not induce any significant toxicity at the tested concentrations.

The data indicate that, as shown for HeLa cells, the binding to T lymphocytes is controlled by the specificity of the bound antibody. Next, we loaded ZapB-GFP-Z IBs simultaneously with two antibodies, namely, fluorescently labeled anti-EGFR and anti-CD3. These dual-antibody decorated IBs were incubated with HeLa and T cells for one hour, and the culture was washed to remove unbound particles. Brightfield microscopy analysis showed abundant circular T cells and polygonal HeLa cells connected by fluorescent IBs (Figs. 8 and S24). In contrast, when HeLa cells and T cells were incubated in the presence of ZapB-GFP-Z IBs loaded only with anti-EGFR IgG, the IBs targeted HeLa cells but did not capture T lymphocytes (Fig. S25).

The IBs mediated colocalization of the two cell types was further studied using FC. As shown in Fig. S26, the dual decoration of ZapB-GFP-Z IBs with anti-EGFR and anti-CD3 antibodies promoted the association of T cells and HeLa cells. However, in the absence of anti-CD3 antibody, the ZapB-GFP-Z IBs failed to promote this cell association. A quantitative analysis indicated that dual anti-EGFR and anti-CD3 antibody decoration of ZapB-GFP-Z IBs increased the number of T cells associated to HeLa cells by at least 14-fold, when compared with IBs decorated only with anti-EGFR antibody (Table S1).

Overall, the data indicate that dual antibody-loaded ZapB-GFP-Z IBs can simultaneously bind at least two different antigens and approach unrelated cell types.
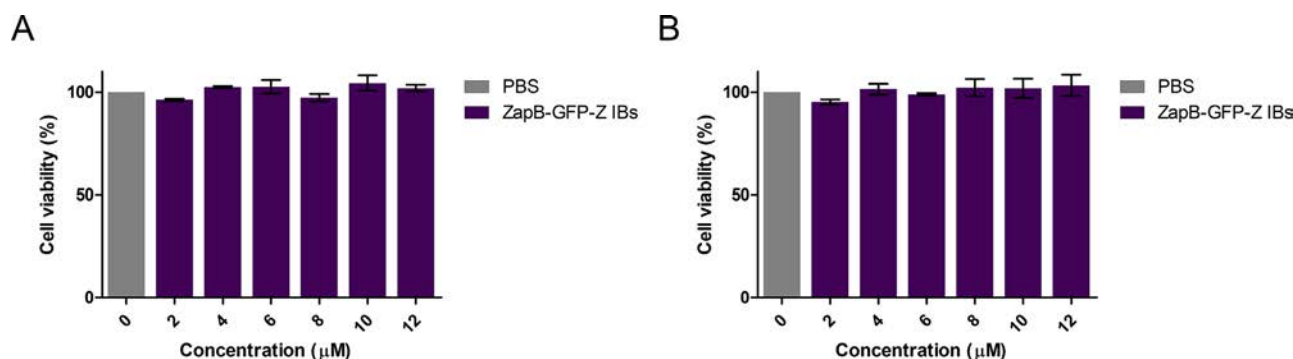
Fig. 4. Cytotoxicity induced by the ZapB-GFP-Z IBs. HeLa (A) and MRC-5 (B) cell viability was tested at different concentrations (from 2 to 12 µM) of ZapB-GFP-Z IBs. IBs were added to the cells and incubated for 72 h. PBS buffer was used as the control.
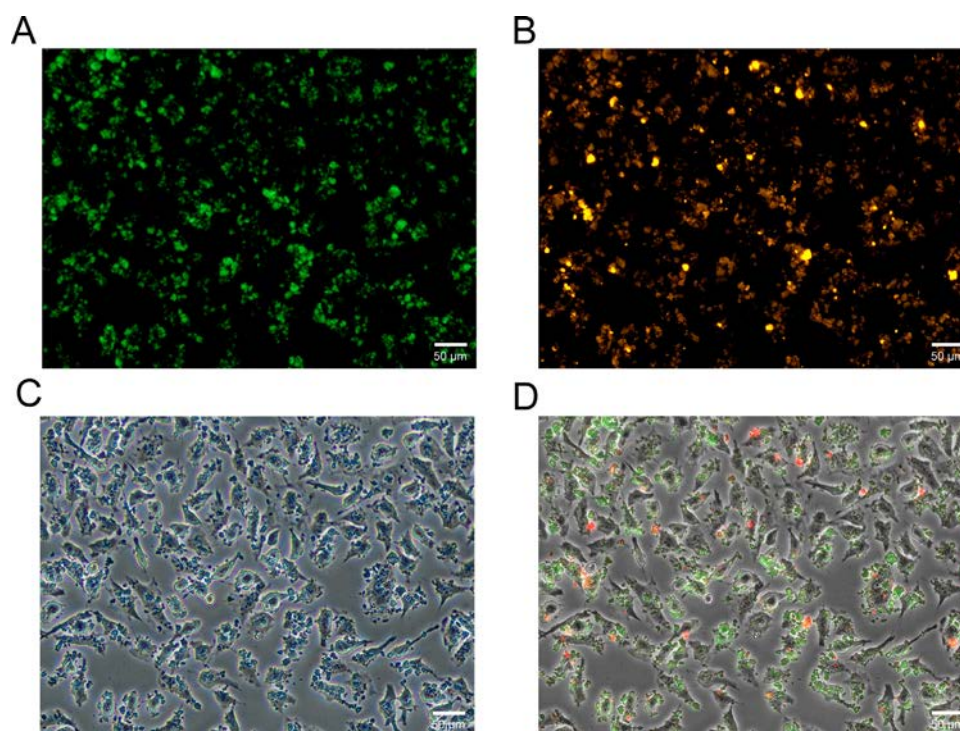
Fig. 5. Specific binding of functionalized ZapB-GFP-Z IBs to human cells. Representative microscopy images of green fluorescence emitted from the ZapB-GFP-Z IBs (A), red fluorescence emitted from anti-EGFR IgG labeled with Alexa Fluor 555 (B), HeLa cells with brightfield microscopy (C), and the three images merged (D). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4. Discussion

Functional IBs are becoming promising sources of active protein for biotechnological and biomedical purposes [44]. These protein nanoparticles are easy to produce and purify at preparative yields and have been applied in tissue engineering, protein delivery, as vaccines, and for catalysis [45–48].

We have recently succeeded in exploiting the ability of a natural coiled-coil domain to self-assemble inside bacterial cells to generate highly active IBs [19]. These α-helix-rich protein nanoparticles overcome the drawbacks of classical β-sheet-rich, amyloid-like IBs. Here, we address whether this inside-the-cell coiled-coil-driven assembly strategy can be used to create multifunctional IBs containing two different folded and active globular domains.

The tripartite ZapB-GFP-Z fusion protein was successfully expressed in *E. coli* cells as IBs. These nanoparticles have been purified to homogeneity, and they display a spherical submicrometric size that remains stable in solution for more than a week, suggesting suitability for their use in nanotechnological or biomedical applications [49,50]. In this study, we showed that these IBs are sustained by native-like α-helical interactions and that the activity of the all-β GFP and all α- Z-domain moieties is preserved, as evidenced by the perfect overlap between the green fluorescence of GFP and the red or blue fluorescence emitted from the labeled IgG antibody attached to the Z-domain. This outcome suggests a generic ability of proteins to keep their active structures within ZapB-based IBs, as indicated by the cooperative melting of the GFP moiety. Importantly, ZapB-GFP-Z IBs keep a submicrometric size when bound to antibodies, without their shape or the constituent proteins stability being altered.

Biomaterials for *in vivo* applications need to be nontoxic to human cells [19]. As intended, generating IBs without amyloid characteristics resulted in innocuous and biocompatible IBs potentially suitable for biomedical purposes, as it has been corroborated with different cell lines. Another requirement for future *in vivo* uses is

negligible antigen-specific toxicity when they are loaded with antibodies, a feature that IgG-decorated ZapB-GFP-Z IBs fulfill.

Selective targeting of defined cell types has emerged as a safe alternative to conventional therapies, minimizing toxic side effects. The ability of BsAbs to simultaneously recognize two different epitopes is increasingly exploited in immuno-oncology, gradually displacing monovalent therapies, since this capacity allows redirection of immune cells to selected cancer cell types [51].

Dual-targeting nanoparticles that combine two different mAbs are receiving increasing attention as alternatives to BsAbs for cancer immunotherapy [35]. The FDA has approved many monovalent mAbs for targeted cancer treatment, and these mAbs are manufactured at a much lower cost than BsAbs. In addition, mAb immobilization in a nanoscopic assembly results in significant stabilization relative to short-lived soluble BsAbs. Current methods for developing multivalent antibody nanostructures rely on covalent strategies that require consecutive chemical conjugations steps [52], which can potentially modify the antibody structure and function. In contrast, in ZapB-GFP-Z protein nanoparticles, antibody decoration relies on the strong, specific, and noncovalent interaction between the Z-domain and the Fc-region of IgG, as shown in the competition binding assay, which is not expected to alter the antibody affinity/specificity for its target.

Once we demonstrated the ability of antibody decorated ZapB-GFP-Z IBs to recognize different cell types specifically, we rationalized that our multifunctional nanoparticles can be a convenient way to promote dual epitope targeting when decorated with two different specific antibodies and, indeed, dual-loaded nanoparticles effectively redirect T cells to cancer cells.

The intrinsic fluorescence of these nanostructures makes them traceable *in vivo*, without the need for fluorescent dyes, and they can be simultaneously loaded with two, and likely more, antibodies in a user-defined stoichiometry. These features, together with the high yield of production confer them advantages, relative to BsAbs, from a biotechnological point of view.
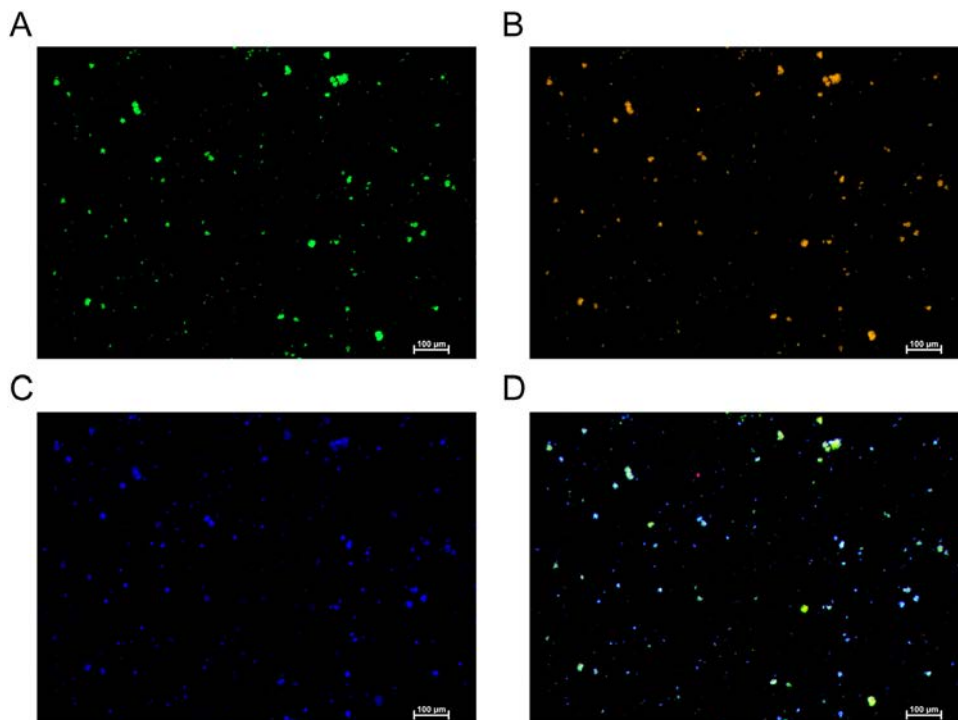
Fig. 6. Multifunctional ZapB-GFP-Z IBs decorated with two different antibodies. Representative microscopy images of GFP fluorescence (A), red fluorescence emitted from IgG labeled with Alexa Fluor 555 (B), blue fluorescence emitted from IgG labeled with eFluor 450 (C), and the three images merged (D). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
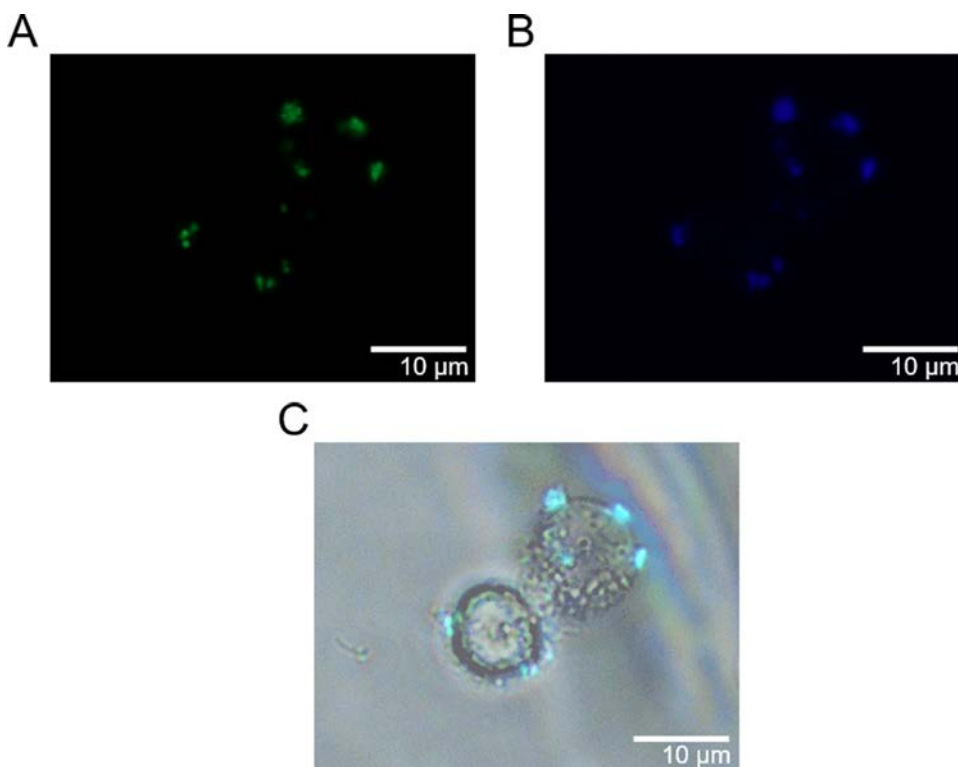
Fig. 7. Specific binding of bifunctional ZapB-GFP-Z IBs to T lymphocytes. Representative microscopy images of GFP fluorescence (A), blue fluorescence emitted from anti-CD3 IgG labeled with eFluor 450 (B), and merging of these images and brightfield image (C). (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)
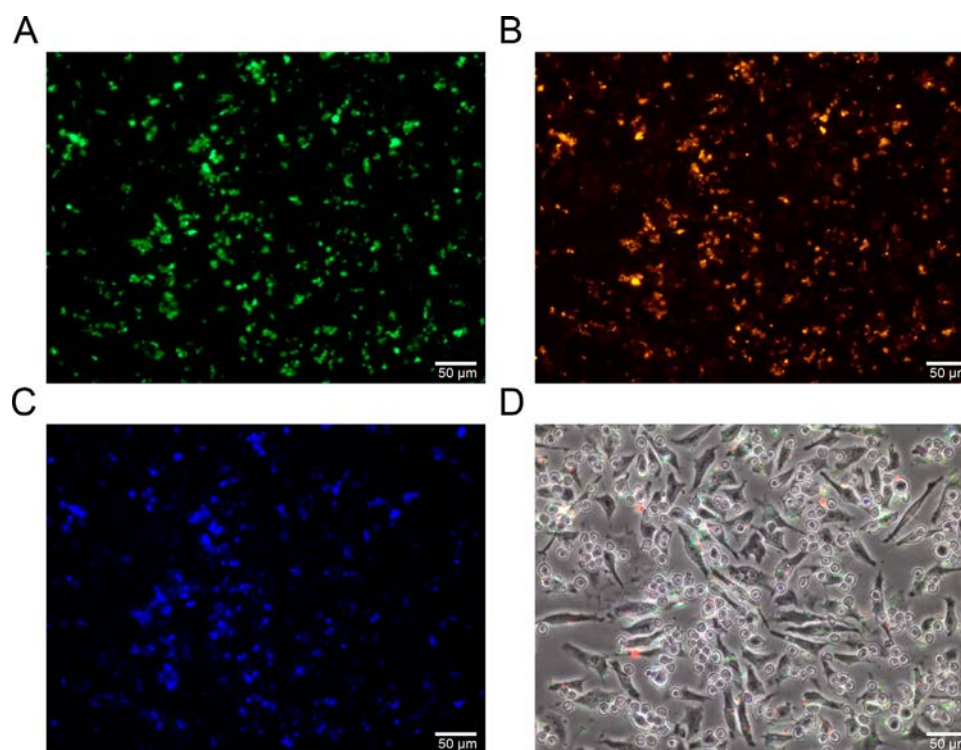
Fig. 8. Dual IgG-decorated ZapB-GFP-Z IBs direct T cells expressing CD3 to HeLa cells expressing EGFR. Representative microscopy images of GFP fluorescence (A), red fluorescence emitted from anti-EGFR IgG labeled with Alexa Fluor 555 (B), blue fluorescence emitted from anti-CD3 IgG labeled with eFluor 450 (C), and merging of these three images and brightfield (D). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Overall, the present proof-of-concept study describes a novel kind of protein-based nanoparticles with potential applications in biomedicine, although the *in vivo* usefulness of this approach needs to be further validated.

5. Conclusion

In this work, we created fluorescent nonamyloid IBs able to capture antibodies with high affinity. These submicrometric, spherical, and α-helix-rich protein nanoparticles are innocuous for human cells. They can potentially recognize any cell type of interest when loaded with the proper antibody. Furthermore, these functional IBs can be easily functionalized by adding different antibodies simultaneously, promoting the association of unrelated cell types, acting as mimics of BsAbs. Their production is cost-effective and fully scalable, and therefore, they are convenient sources of *in vivo* traceable protein nanoparticles for potential biomedical applications.

Funding

Declaration of Competing Interest

The authors declare that they have no competing interests.

Acknowledgments

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.actbio.2021.06.040.

References

[1] A. Schrödel, A. De Marco, Characterization of the aggregates formed during recombinant protein expression in bacteria, BMC Biochem. 6 (2005) 1–11, doi:10.1186/1471-2091-6-10.

[2] N. Ferrer-Miralles, J. Domingo-Espín, J. Corchero, E. Vázquez, A. Villaverde, Microbial factories for recombinant pharmaceuticals, Microb. Cell Fact.. 8 (2009) 1–8. 10.1186/1475-2859-8-17.

[3] F. Baneyx, M. Mujacic, Recombinant protein folding and misfolding in Escherichia coli, Nat. Biotechnol. 22 (2004) 1399–1407, doi:10.1038/nbt1029.

[4] U. Rinas, E. Garcia-Fruitós, J.L. Corchero, E. Vázquez, J. Seras-Franzoso, A. Villaverde, Bacterial inclusion bodies: discovering their better half, Trends Biochem. Sci. 42 (2017) 726–737, doi:10.1016/j.tibs.2017.01.005.

[5] A. De Marco, N. Ferrer-Miralles, E. Garcia-Fruitós, A. Mitraki, S. Peternel, U. Rinas, M.A. Trujillo-Roldán, N.A. Valdez-Cruz, E. Vázquez, A. Villaverde, Bacterial inclusion bodies are industrially exploitable amyloids, FEMS Microbiol. Rev. 43 (2019) 53–72, doi:10.1093/femsre/fuy038.

[6] E. García-Fruitós, E. Rodríguez-Carmona, C. Díez-Gil, R.M. Ferraz, E. Vázquez, J.L. Corchero, M. Cano-Sarabia, I. Ratera, N. Ventosa, J. Veciana, A. Villaverde, Surface cell growth engineering assisted by a novel bacterial nanomaterial, Adv. Mater. 21 (2009) 4249–4253, doi:10.1002/adma.200900283.

[7] G. Margreiter, P. Messner, K.D. Caldwell, K. Bayer, Size characterization of inclusion bodies by sedimentation field-flow fractionation, J. Biotechnol. 138 (2008) 67–73, doi:10.1016/j.jbiotec.2008.07.1995.

[8] V.D. Jäger, R. Lamm, K. Küsters, G. Ölçücü, M. Oldiges, K.E. Jaeger, J. Büchs, U. Krauss, Catalytically-active inclusion bodies for biotechnology-General concepts, optimization, and application, Appl. Microbiol. Biotechnol. 104 (2020) 7313–7329, doi:10.1007/s00253-020-10760-3.

[9] A. Villaverde, J.L. Corchero, J. Seras-Franzoso, E. Garcia-Fruitós, Functional protein aggregates: just the tip of the iceberg, Nanomedicine 10 (2015) 2881–2891 (London, England), doi:10.2217/nnm.15.125.

[10] P. Singhvi, A. Saneja, S. Srichandan, A.K. Panda, Bacterial inclusion bodies: a treasure trove of bioactive proteins, Trends Biotechnol. 38 (2020) 474–486, doi:10.1016/j.tibtech.2019.12.011.

[11] R. Kloss, M.H. Limberg, U. Mackfeld, D. Hahn, A. Grünberger, V.D. Jäger, U. Krauss, M. Oldiges, M. Pohl, Catalytically active inclusion bodies of *L*-lysine decarboxylase from *E. coli* for 1,5-diaminopentane production, Sci. Rep. 8 (2018) 1–11, doi:10.1038/s41598-018-24070-2.

[12] U. Krauss, V.D. Jäger, M. Diener, M. Pohl, K.E. Jaeger, Catalytically-active inclusion bodies-carrier-free protein immobilizates for application in biotechnology and biomedicine, J. Biotechnol. 258 (2017) 136–147, doi:10.1016/j.jbiotec.2017.04.033.

[13] V.D. Jäger, R. Lamm, R. Kloß, E. Kaganovitch, A. Grünberger, M. Pohl, J. Büchs, K.E. Jaeger, U. Krauss, A synthetic reaction cascade implemented by colocalization of two proteins within catalytically active inclusion bodies, ACS Synth. Biol. 7 (2018) 2282–2295, doi:10.1021/acssynbio.8b00274.

[14] M. Pesarrodona, T. Jauset, Z.V. Díaz-Riascos, A. Sánchez-Chardi, M.E. Beaulieu, J. Seras-Franzoso, L. Sánchez-García, R. Baltà-Foix, S. Mancilla, Y. Fernández, U. Rinas, S. Schwartz, L. Soucek, A. Villaverde, I. Abasolo, E. Vázquez, Targeting antitumoral proteins to breast cancer by local administration of functional inclusion bodies, Adv. Sci. 6 (2019), doi:10.1002/advs.201900849.

[15] M.V. Céspedes, O. Cano-Garrido, P. Álamo, R. Sala, A. Gallardo, N. Serna, A. Falgàs, E. Voltà-Durán, I. Casanova, A. Sánchez-Chardi, H. López-Laguna, L. Sánchez-García, J.M. Sánchez, U. Unzueta, E. Vázquez, R. Mangues, A. Villaverde, Engineering secretory amyloids for remote and highly selective destruction of metastatic foci, Adv. Mater. 1907348 (2019) 1–9, doi:10.1002/adma.201907348.

[16] S. Ventura, A. Villaverde, Protein quality in bacterial inclusion bodies, Trends Biotechnol. 24 (2006) 179–185, doi:10.1016/j.tibtech.2006.02.007.

[17] M. Carrió, N. González-Montalbán, A. Vera, A. Villaverde, S. Ventura, Amyloid-like properties of bacterial inclusion bodies, J. Mol. Biol. 347 (2005) 1025–1037, doi:10.1016/j.jmb.2005.02.030.

[18] M. Morell, R. Bravo, A. Espargaró, X. Sisquella, F.X. Avilés, X. Fernàndez-Busquets, S. Ventura, Inclusion bodies: specificity in their aggregation process and amyloid-like structure, Biochim. Biophys. Acta-Mol. Cell Res. 1783 (2008) 1815–1825, doi:10.1016/j.bbamcr.2008.06.007.

[19] M. Gil-Garcia, S. Navarro, S. Ventura, Coiled-coil inspired functional inclusion bodies, Microb. Cell Fact. 19 (2020) 1–16, doi:10.1186/s12934-020-01375-4.

[20] V.D. Jäger, R. Kloss, A. Grünberger, S. Seide, D. Hahn, T. Karmainski, M. Piqueray, J. Embruch, S. Longerich, U. Mackfeld, K.E. Jaeger, W. Wiechert, M. Pohl, U. Krauss, Tailoring the properties of (catalytically)-active inclusion bodies, Microb. Cell Fact. 18 (2019) 1–20, doi:10.1186/s12934-019-1081-5.

[21] M. Diener, B. Kopka, M. Pohl, K.E. Jaeger, U. Krauss, Fusion of a coiled-coil domain facilitates the high-level production of catalytically active enzyme inclusion bodies, ChemCatChem 8 (2016) 142–152, doi:10.1002/cctc.201501001.

[22] R. Lamm, V.D. Jäger, B. Heyman, C. Berg, C. Cürten, U. Krauss, K.E. Jaeger, J. Büchs, Detailed small-scale characterization and scale-up of active YFP inclusion body production with Escherichia coli induced by a tetrameric coiled coil domain, J. Biosci. Bioeng. xxx (2020), doi:10.1016/j.jbiosc.2020.02.003.

[23] W. Wang, S. Ventura, Prion domains as a driving force for the assembly of functional nanomaterials, Prion 14 (2020) 170–179, doi:10.1080/19336896.2020.1785659.

[24] R.M. Hoffman, Application of GFP imaging in cancer, Lab. Invest. 95 (2015) 432–452, doi:10.1038/labinvest.2014.154.

[25] M. Tashiro, R. Tejero, D.E. Zimmerman, B. Celda, B. Nilsson, G.T. Montelione, High-resolution solution NMR structure of the *Z* domain of staphylococcal protein A, J. Mol. Biol. 272 (1997) 573–590, doi:10.1006/jmbi.1997.1265.

[26] C. Chen, Q.L. Huang, S.H. Jiang, X. Pan, Z.C. Hua, Immobilized protein ZZ, an affinity tool for immunoglobulin isolation and immunological experimentation, Biotechnol. Appl. Biochem. 45 (2006) 87, doi:10.1042/ba20060055.

[27] W. Choe, T.A. Durgannavar, S.J. Chung, Fc-Binding Ligands of immunoglobulin G: an overview of high affinity proteins and peptides, Materials 9 (2016) (Basel, Switzerland), doi:10.3390/ma9120994.

[28] A. Konrad, A.E. Karlström, S. Hober, Covalent immunoglobulin labeling through a photoactivable synthetic *Z* domain, Bioconjug. Chem. 22 (2011) 2395–2403, doi:10.1021/bc200052h.

[29] M. Ultsch, A. Braisted, H.R. Maun, C. Eigenbrot, 3-2-1: Structural insights from stepwise shrinkage of a three-helix Fc-binding domain to a single helix, Protein Eng. Des.Sel. PEDS 30 (2017) 619–625, doi:10.1093/protein/gzx029.

[30] A.C. Braisted, J.A. Wells, Minimizing a binding domain from protein *A*, PNAS 93 (1996) 5688–5692, doi:10.1073/pnas.93.12.5688.

[31] D. Skokos, J.C. Waite, L. Haber, A. Crawford, A. Hermann, E. Ullman, R. Slim, S. Godin, D. Ajithdoss, X. Ye, B. Wang, Q. Wu, I. Ramos, A. Pawashe, L. Canova, K. Vazzana, P. Ram, E. Herlihy, H. Ahmed, E. Oswald, J. Golubov, P. Poon, L. Havel, D. Chiu, M. Lazo, K. Provoncha, K. Yu, J. Kim, J.J. Warsaw, N.S. Oristian, C.J. Siao, D. Dudgeon, T. Huang, T. Potocky, J. Martin, D. MacDonald, A. Oyejide, A. Rafique, W. Poueymirou, J.R. Kirshner, E. Smith, W. Olson, J. Lin, G. Thurston, M.A. Sleeman, A.J. Murphy, G.D. Yancopoulos, A class of costimulatory CD28-bispecific antibodies that enhance the antitumor activity of CD3-bispecific antibodies, Sci. Trans. Med. 12 (2020), doi:10.1126/scitranslmed.aaw7888.

[32] H. Kantarjian, A. Stein, N. Gökbuget, A.K. Fielding, A.C. Schuh, J.M. Ribera, A. Wei, H. Dombret, R. Foà, R. Bassan, Ö. Arslan, M.A. Sanz, J. Bergeron, F. Demirkan, E. Lech-Maranda, A. Rambaldi, X. Thomas, H.A. Horst, M. Brüggemann, W. Klapper, B.L. Wood, A. Fleishman, D. Nagorsen, C. Holland, Z. Zimmerman, M.S. Topp, Blinatumomab versus chemotherapy for advanced acute Lymphoblastic Leukemia, N. Engl. J. Med. 376 (2017) 836–847, doi:10.1056/NEJMoa1609783.

[33] C. Rader, Bispecific antibodies in cancer immunotherapy, Curr. Opin. Biotechnol. 65 (2020) 9–16, doi:10.1016/j.copbio.2019.11.020.

[34] H. Liu, A. Saxena, S.S. Sidhu, D. Wu, Fc engineering for developing therapeutic bispecific antibodies and novel scaffolds, Front. Immunol. 8 (2017) 1–15, doi:10.3389/fimmu.2017.00038.

[35] A.K. Kosmides, J.W. Sidhom, A. Fraser, C.A. Bessell, J.P. Schneck, Dual targeting nanoparticle stimulates the immune system to inhibit tumor growth, ACS Nano 11 (2017) 5417–5429, doi:10.1021/acsnano.6b08152.

[36] M.C. Lo Giudice, F. Meder, E. Polo, S.S. Thomas, K. Alnahdi, S. Lara, K.A. Dawson, Constructing bifunctional nanoparticles for dual targeting: improved grafting and surface recognition assessment of multiple ligand nanoparticles, Nanoscale 8 (2016) 16969–16975, doi:10.1039/c6nr05478a.

[37] J. Pujols, S. Peña-Díaz, M. Conde-Giménez, F. Pinheiro, S. Navarro, J. Sancho, S. Ventura, High-throughput screening methodology to identify alpha-synuclein aggregation inhibitors, Int. J. Mol. Sci. 18 (2017), doi:10.3390/ijms18030478.

[38] T. Heimburg, J. Schünemann, K. Weber, N. Geisler, FTIR-spectroscopy of multistranded coiled coil proteins, Biochemistry 38 (1999) 12727–12734, doi:10.1021/bi983079h.

[39] R. Sabaté, S. Ventura, Cross-β-sheet supersecondary structure in amyloid folds: techniques for detection and characterization, Methods Mol. Biol. 932 (2013) 237–257, doi:10.1007/978-1-62703-065-6_15.

[40] A. Espargaró, S. Llabrés, S.J. Saupe, C. Curutchet, F.J. Luque, R. Sabaté, On the binding of congo red to amyloid fibrils, Angew. Chem. Int. Ed. Engl. 59 (2020) 8104–8107, doi:10.1002/anie.201916630.

[41] D. Li, F. Chen, J. Ding, N. Lin, Z. Li, X. Wang, Knockdown of HIP1 expression promotes ligand-induced endocytosis of EGFR in HeLa cells, Oncol. Rep. 38 (2017) 3387–3391, doi:10.3892/or.2017.6025.

[42] N. Normanno, A. De Luca, C. Bianco, L. Strizzi, M. Mancino, M.R. Maiello, A. Carotenuto, G. De Feo, F. Caponigro, D.S. Salomon, Epidermal growth factor receptor (EGFR) signaling in cancer, Gene 366 (2006) 2–16, doi:10.1016/j.gene.2005.10.018.

[43] M.S. Kuhns, H.B. Badgandi, Piecing together the family portrait of TCR-CD3 complexes, Immunol. Rev. 250 (2012) 120–143, doi:10.1111/imr.12000.

[44] E. García-Fruitós, E. Vázquez, C. Díez-Gil, J.L. Corchero, J. Seras-Franzoso, I. Ratera, J. Veciana, A. Villaverde, Bacterial inclusion bodies: making gold from waste, Trends Biotechnol. 30 (2012) 65–70, doi:10.1016/j.tibtech.2011.09.003.

[45] S.T.T. Schetters, W.S.P. Jong, L.J.W. Kruijssen, H.B. van den, B. van Saparoea, S. Engels, W.W.J. Unger, D. Houben, J.M.M. den Haan, J. Luirink, Y. van Kooyk, Bacterial inclusion bodies function as vehicles for dendritic cell-mediated *T* cell responses, Cell. Mol. Immunol. 17 (2020) 415–417, doi:10.1038/s41423-019-0298-x.

[46] M.V. Céspedes, Y. Fernández, U. Unzueta, R. Mendoza, J. Seras-Franzoso, A. Sánchez-Chardi, P. Álamo, V. Toledo-Rubio, N. Ferrer-Miralles, E. Vázquez, S. Schwartz, I. Abasolo, J.L. Corchero, R. Mangues, A. Villaverde, Bacterial mimetics of endocrine secretory granules as immobilized *in vivo* depots for functional protein drugs, Sci. Rep. 6 (2016) 1–10, doi:10.1038/srep35765.

[47] R. Kloss, T. Karmainski, V.D. Jäger, D. Hahn, A. Grünberger, M. Baumgart, U. Krauss, K.E. Jaeger, W. Wiechert, M. Pohl, Tailor-made catalytically active inclusion bodies for different applications in biocatalysis, Catal. Sci. Technol. 8 (2018) 5816–5826, doi:10.1039/c8cy01891j.

[48] J. Seras-Franzoso, C. Dez-Gil, E. Vazquez, E. Garca-Fruits, R. Cubarsi, I. Ratera, J. Veciana, A. Villaverde, Bioadhesiveness and efficient mechanotransduction stimuli synergistically provided by bacterial inclusion bodies as scaffolds for tissue engineering, Nanomedicine 7 (2012) 79–93, doi:10.2217/nnm.11.83.

[49] Š. Peternel, R. Komel, Isolation of biologically active nanomaterial (inclusion bodies) from bacterial cells, Microb. Cell Fact. 9 (2010) 1–16, doi:10.1186/1475-2859-9-66.

[50] L. Gifre-Renom, J. Seras-Franzoso, D. Rafael, F. Andrade, O. Cano-Garrido, F. Martinez-Trucharte, E. Ugarte-Berzal, E. Martens, L. Boon, A. Villaverde, G. Opdenakker, S. Schwartz, A. Arís, E. Garcia-Fruitós, The biological potential hidden in inclusion bodies, Pharmaceutics 12 (2020), doi:10.3390/pharmaceutics12020157.

[51] A.F. Labrijn, M.L. Janmaat, J.M. Reichert, P.W.H.I. Parren, Bispecific antibodies: a mechanistic review of the pipeline, Nat. Rev. Drug Discov. 18 (2019) 585–608, doi:10.1038/s41573-019-0028-1.

[52] M. Arruebo, M. Valladares, Á. González-Fernández, Antibody-conjugated nanoparticles for biomedical applications, J. Nanomater. 2009 (2009), doi:10.1155/2009/439389.

CONCLUDING REMARKS

Chapter 1: Coiled-Coil Based Inclusion Bodies and Their Potential Applications

- IBs are a source of active protein in a ready-to-use and assembled form, holding great potential in biomedicine and biotechnology.
- The formation of IBs assisted by aggregation-prone domains often results in the intracellular accumulation of active nanoparticles sustained by amyloid interactions.
- Coiled-coil domains are emerging as an alternative for creating highly functional and safe IBs that overcome the drawbacks related to amyloid conformations.
- Functional IBs have been employed for a plethora of biomedical and biotechnological applications thanks to their excellent mechanical properties, biocompatibility, and versatility, acting, for example, as drug delivery agents and reusable enzymatic microreactors.
- Their easy production, together with their high specific activity, suggests that coiled-coil-based IBs might be a strategy that, together with the artificial IBs technology, can lead to a next generation of functional IBs with enhanced properties.

Chapter 2: Combining Structural Aggregation Propensity and Stability Predictions to Redesign Protein Solubility

- A3D successfully assisted the redesign of the fr-GFP solubility by inserting three basic residues at the protein surface, resulting in an improved version with enhanced solubility and folding properties.
- The introduction of solubilizing but destabilizing mutations negatively impacts the effective protein solubility due to destabilization and misfolding of the native conformation.
- Forecasting the impact of solubilizing mutations on the thermodynamic stability of proteins allows improving their solubility by simple protein redesign procedures.
- Incorporating stability calculations using FoldX into the A3D pipeline could prevent users from selecting mutations that jeopardize successful protein redesign.

- A minimalist redesign of protein solubility does not alter the 3D structure and function of the target protein.

- The reported approach can be transversely applied to redesign the solubility of therapeutic proteins such as human antibodies, providing solubilized variants that display the same stability as the *wild-type* counterpart.

- The obtained data suggest that there is room to improve protein solubility without impacting protein structure and stability by inserting a small number of mutations at the protein surface.

Chapter 3: Coiled-Coil Inspired Functional Inclusion Bodies

- An exhaustive *in silico* analysis of the structural properties of different proteins confirmed the suitability of a natural coiled-coil domain to work as an IB-tag for creating non-amyloid-based IBs.

- This coiled-coil domain is a versatile tool for building up different supramolecular structures that are sustained by native α-helical interactions.

- The described approach allows us to design highly functional and biocompatible IBs maintained by coiled-coil interactions cost-effectively.

- The higher specific activity and better biocompatibility reported for coiled-coil-based IBs in comparison with their amyloid counterparts can be explained by the nature of the intermolecular contacts sustaining each of these assemblies.

- The present study paves the way for creating functional IBs using coiled-coil domains instead of the classical amyloid tags, expanding the toolbox to generate safe nanoparticles with tailored functionalities.

Chapter 4: Multifunctional Antibody-Conjugated Coiled-Coil Protein Nanoparticles for Selective Cell Targeting

- The design of a fusion protein composed of a coiled-coil domain, as the assembling entity, and the GFP and the Z-domain, as active moieties, results in the assembly of fluorescent and antibody-capturing nanoparticles.

- These multifunctional IBs, sustained by α-helical interactions, display a spherical shape and a submicrometric size, which is maintained for a long time, demonstrating that coiled-coil-based protein fusions can self-assemble into homogeneous and highly stable nanoparticles.

- These nanoparticles are biocompatible in two unrelated human cell lines, confirming their safety for *in vivo* biomedical applications.

- Single antibody-decorated IBs target specific human cells expressing the intended receptor without any associated toxicity, allowing to direct these nanoparticles to the desired objective easily.

- A dual antibody-functionalization of these IBs triggers the approach of two unrelated cell types, mimicking the mechanism of action of bispecific antibodies.

- The nanoparticles described in this chapter constitute an appealing and economical platform to target different antigens simultaneously, emerging as a feasible strategy with promising therapeutic and bioimaging applications.

ANNEX: SUPPLEMENTARY INFORMATION

Chapter 2

Combining Structural Aggregation Propensity and Stability Predictions to Redesign Protein Solubility

Marcos Gil-Garcia, Manuel Bañó-Polo, Nathalia Varejão, Michal Jamroz, Aleksander Kuriata, Marta Díaz-Caballero, Jara Lascorz, Bertrand Morel, Susanna Navarro, David Reverter, Sebastian Kmiecik and Salvador Ventura

1   **SUPPORTING INFORMATION FOR**

2   **Combining structural aggregation propensity and stability**

3   **predictions to re-design protein solubility**

4   Marcos Gil-Garcia[1], Manuel Bañó-Polo[1], Nathalia Varejão[1], Michal Jamroz[2], Aleksander Kuriata[2],

5   Marta Diaz-Caballero[1], Jara Lascorz[1], Bertrand Morel[3], Susanna Navarro[1], David Reverter[1],

6   Sebastian Kmiecik[2] and Salvador Ventura[1*]

7   [1]Institut de Biotecnologia i de Biomedicina and Departament de Bioquímica i Biologia Molecular;

8   Universitat Autònoma de Barcelona; Bellaterra (Barcelona) 08193, Spain.

9   [2]Biological and Chemical Research Centre, Faculty of Chemistry; University of Warsaw; Warsaw,

10   Poland.

11   [3]Departamento de Química Física e Instituto de Biotecnología, Facultad de Ciencias, Universidad

12   de Granada, 18071 Granada, Spain.

13

14

15

16

17   To whom correspondence should be addressed

18

19   Tel: 93 58689 56; Fax: 93 5812011, E- mail: salvador.ventura@uab.cat

20

21   **Table S1. Data collection and refinement statistics.** Values in parentheses are for the last shell.

	GFP/KKK
Data collection	
Space group	$P2_1\ 2_1\ 2_1$
Unit cell parameters (Å) a, b, c	31.81, 60.86, 106.41
Resolution range (Å)	60.86-1.13
Rmerge	0.09(0.80)
(I/s(I))	8(1.5)
Completeness (%)	99.5(92.6)
Multiplicity	4.2(3.9)
CC(1/2)	0.99(0.56)
Structure refinement	
Resolution range (Å)	53.20(1.13)
No. Reflections	77574
Rwork/Rfree (%)	18.5/20.4
No. of atoms	
Protein	1848
Water molecules	334
Overall B factors (Å^2)	13.38
R.m.s. deviations	
Bonds (Å)	0.013
Angles (°)	1.67
PDB entry	6FWW

22

23

24   **Table S2. Distances between residues 11, 39 and 221 in fr-GFP and GFP/KKK.** Distances measured using PDB:

25   2B3Q and PDB: 6FWW for fr-GFP and GFP/KKK, respectively.

Lys11 N$^{\zeta}$	Lys221 N$^{\zeta}$	14.79 Å
Val11 C$^{\gamma 1}$	Leu221 C$^{\delta 1}$	12.35* Å
Lys11 N$^{\zeta}$	Lys39 N$^{\zeta}$	4.88 Å
Val11 C$^{\gamma 2}$	Tyr39 C$^{\delta 1}$	7.19* Å
Lys39 N$^{\zeta}$	Lys221 N$^{\zeta}$	14.93 Å
Tyr39 C$^{\delta 2}$	Leu221 C$^{\delta 1}$	10.19* Å

26   *Values measured in the fr-GFP (PDB: 2B3Q)

27

28 **Table S3. A3D analysis of DP47 and its aspartic acid mutants.** More negative average and total A3D scores indicate

29 higher solubility. Negative ΔΔG values correspond to over-stabilizing mutations.

DP47 PROTEIN	AVERAGE SCORE	TOTAL SCORE	ΔΔG (kcal/mol)
DP47	-0.48	-54.248	-
DP47/D	-0.5974	-67.5027	0.3038
DP47/DD	-0.704	-79.5483	1.5
DP47/DDD	-0.7984	-90.2156	1.9113

30

31

32

33

34

35
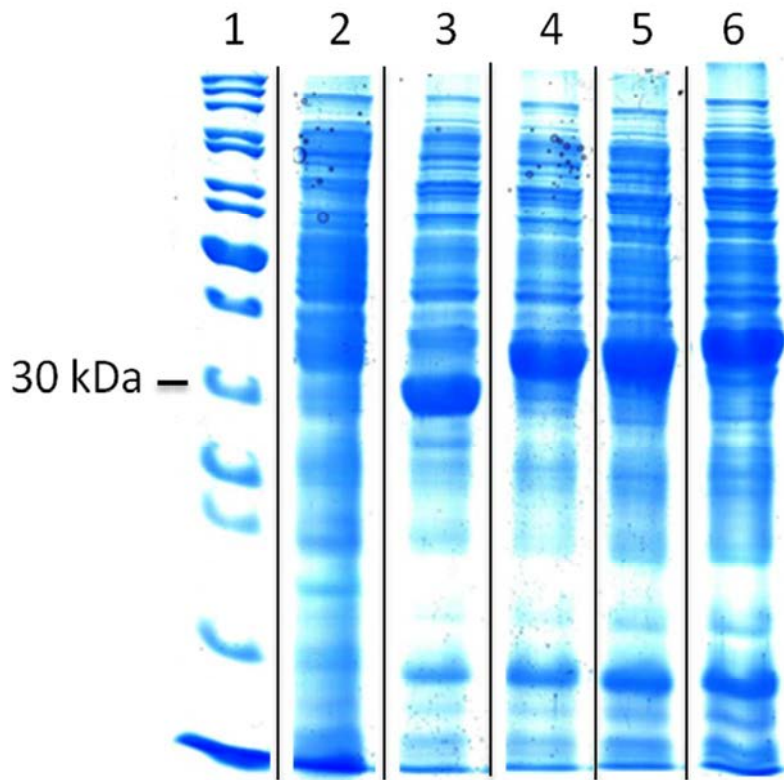
Figure S1. Expression gel of Aβ42-GFP and aspartic acid variants. Lane 1 corresponds to protein ladder, lane 2 to non-induced culture and lanes 3,4,5,6 correspond to Aβ42-GFP, Aβ42-GFP/D, Aβ42-GFP/DD and Aβ42-GFP/DDD respectively. Different lanes were extracted from the same SDS-PAGE gel.

39

40
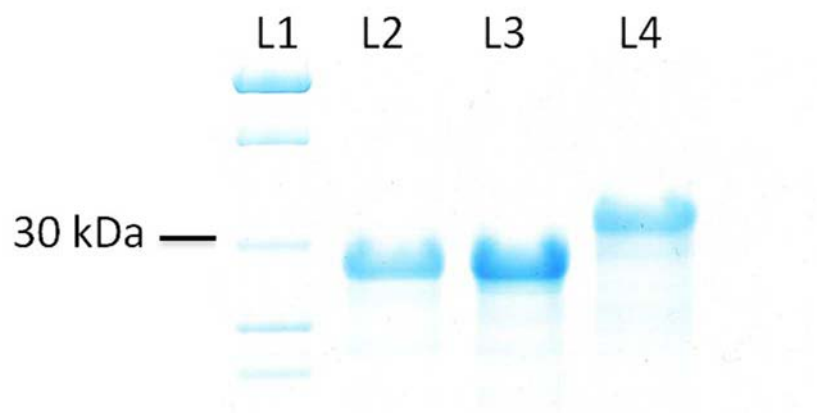
41

42

43 **Figure S2. SDS-PAGE of fr-GFP variants.** Lane 1 corresponds to protein MW ladder, lane 2 to fr-GFP protein, lane 3

44 to GFP/KKK and lane 4 to GFP/DDD.

45
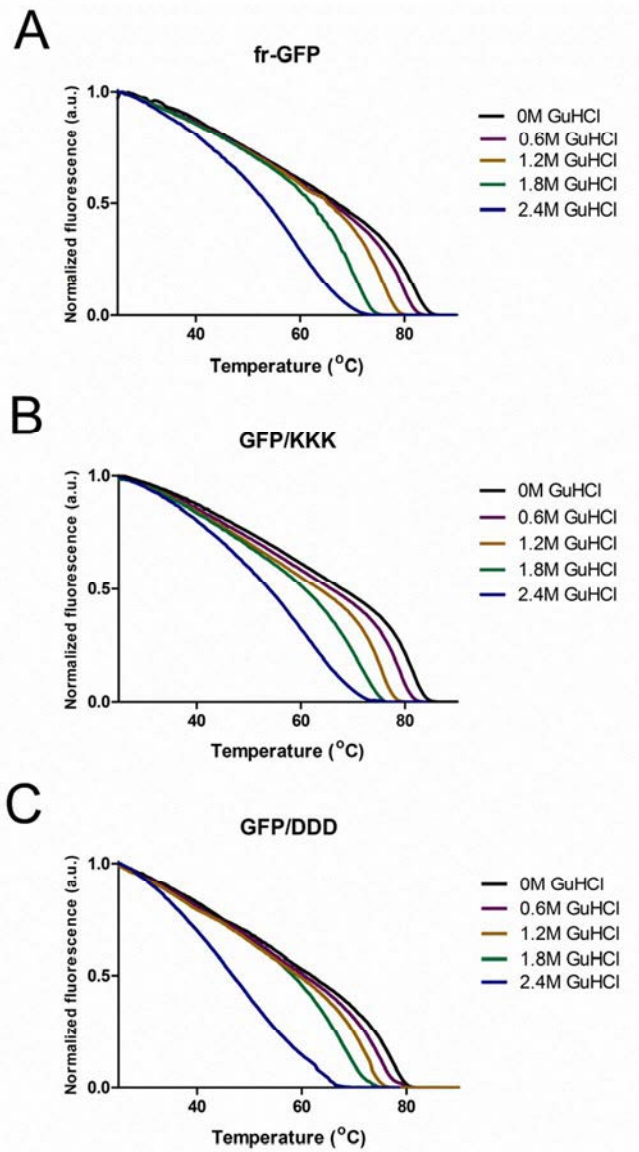
Figure S3. Thermal denaturation of different fr-GFP variants in the presence of increasing concentrations of GuHCl. A), **B)** and **C)** correspond to thermal denaturation curves of fr-GFP, GFP/KKK and GFP/DDD, respectively.

48
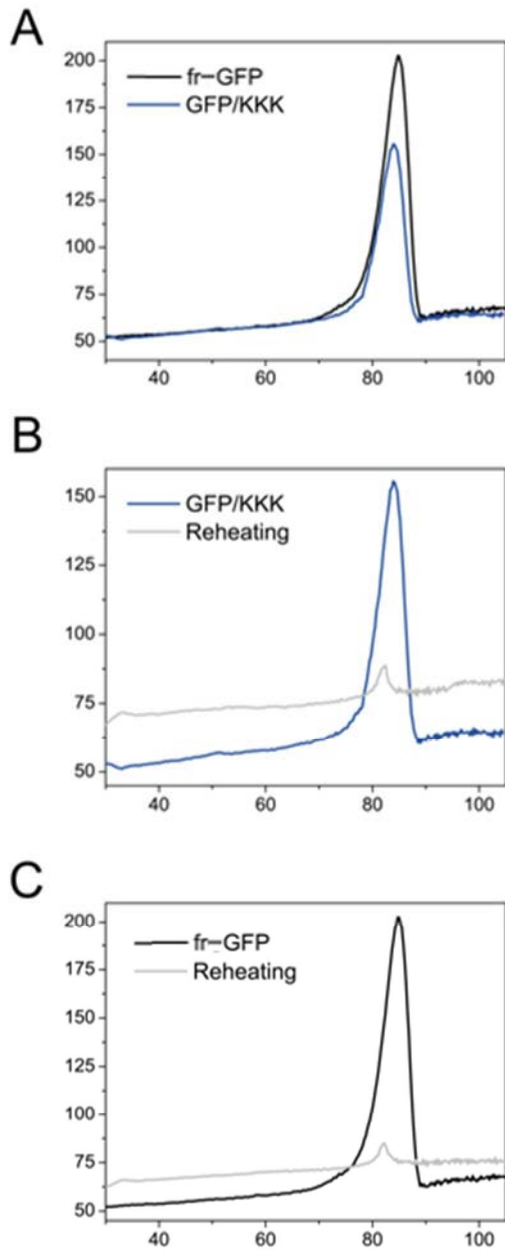
Figure S4. Thermal denaturation of fr-GFP and GFP/KKK analyzed by Differential Scanning Calorimetry (DSC).

A) Thermal denaturation of both proteins showing the transition at 84 ^{o}C. **B)** and **C)** Thermal denaturation of both proteins followed by DSC (black line). The grey line corresponds to the thermogram obtained in a second consecutive heating of the sample.
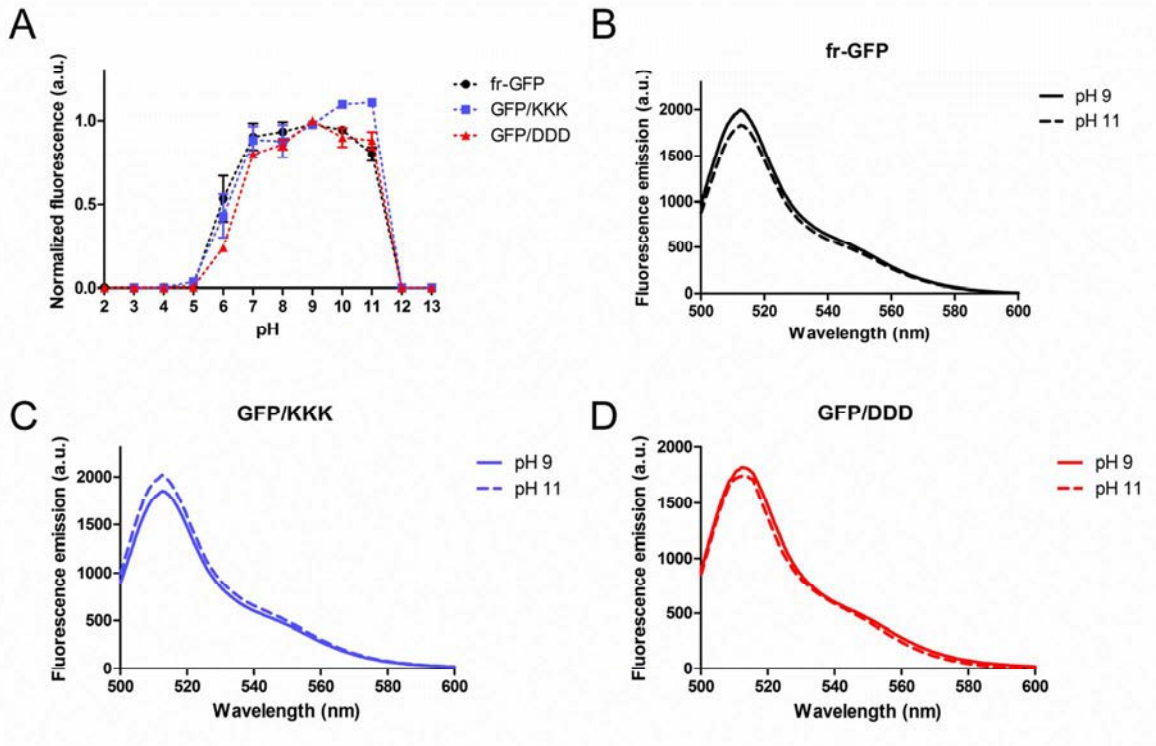
Figure S5. Dependence of GFP fluorescence on the pH. A) Normalized fluorescence of fr-GFP variants in a range of pH from 2 to 13. The value of all proteins at pH 9.0 was normalized to 1, the rest of values are relative to it. **B)**, **C)** and **D)** Fluorescence emission spectrum at pH 9 and pH 11 of fr-GFP, GFP/KKK and GFP/DDD, respectively.

57
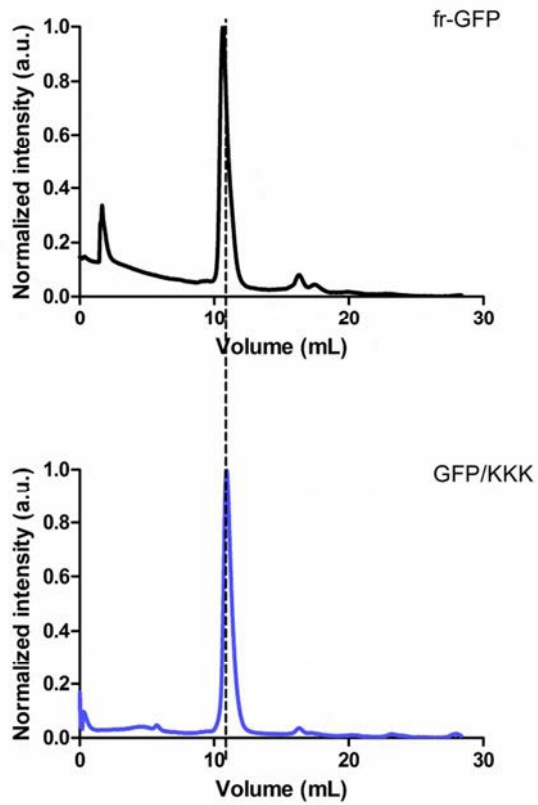
Figure S6. Elution profile of fr-GFP and GFP/KKK in a Size-Exclusion Chromatography. Both proteins elute in the same volume, corresponding with the monomeric form of the protein.

60
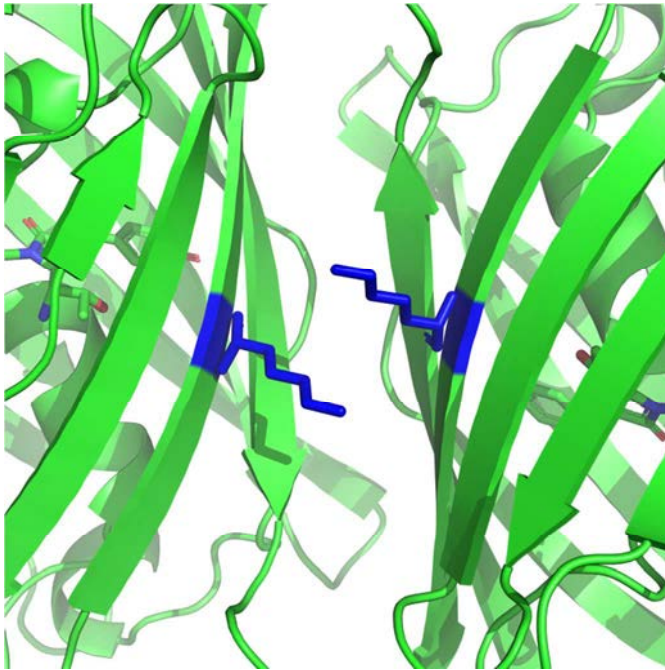
Figure S7. 3D Structure of GFP/KKK dimer model based on PDB: 2B3Q. Lysine 221 in each of the GFP molecules is

indicated in blue.

fr-GFP   **GFP/KKK**

| 0 | 5 | 12.5 | 25 | 37.5 | 50 | 5 | 12.5 | 25 | 37.5 | 50 | Ptn:DNA |
| 0 | 1 | 2.5 | 5 | 7.5 | 10 | 1 | 2.5 | 5 | 7.5 | 10 | Ptn (µM) |

GFP

Free DNA
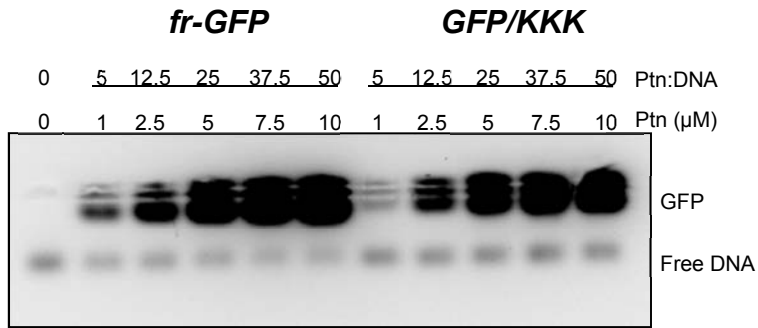
(FITC)-labeled 50b ssDNA

64

65 **Figure S8. Electrophoretic mobility shift assays (EMSA)**. DNA binding properties of fr-GFP and GFP/KKK mutant,

66 were analyzed by EMSA saturation experiments. Protein mixtures were incubated for 30 min at 30°C before loading the

67 TAE-agarose gel. Numbers above gel indicate the molar ratio of protein over (FITC)-labeled 50b ssDNA in each lane.

68

69

Figure S9. "Project details" screen of A3D indicating the predicted effect of selected point mutations in protein stability according to FoldX. "Project details" screens of fr-GFP (A), GFP/KKK (B) and GFP/DDD (C) prediction projects.

13

73
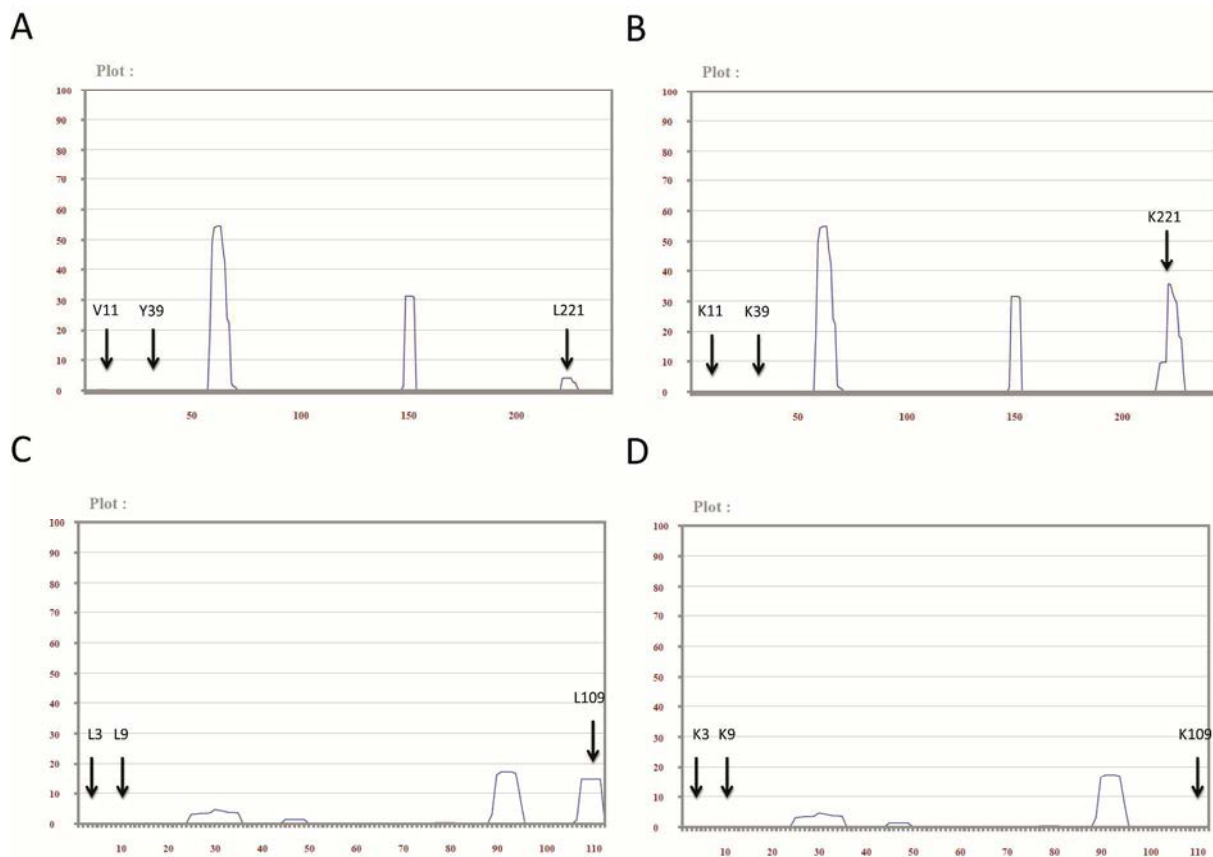
75 **Figure S10. Aggregation profile of A) fr-GFP, B) GFP/KKK, C) DP47** and **D) DP47/KKK according to TANGO.**

76 Mutated positions are indicated.

77

78

79

80

81

82

83

ANNEX: SUPPLEMENTARY INFORMATION

Chapter 3

Coiled-Coil Inspired Functional Inclusion Bodies

Marcos Gil-Garcia, Susanna Navarro and Salvador Ventura

Microbial Cell Factories

2020

Additional File 1:

Coiled-coil inspired functional inclusion bodies

Marcos Gil-Garcia[1], Susanna Navarro[1] and Salvador Ventura[1]*

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

*Correspondence: salvador.ventura@uab.cat
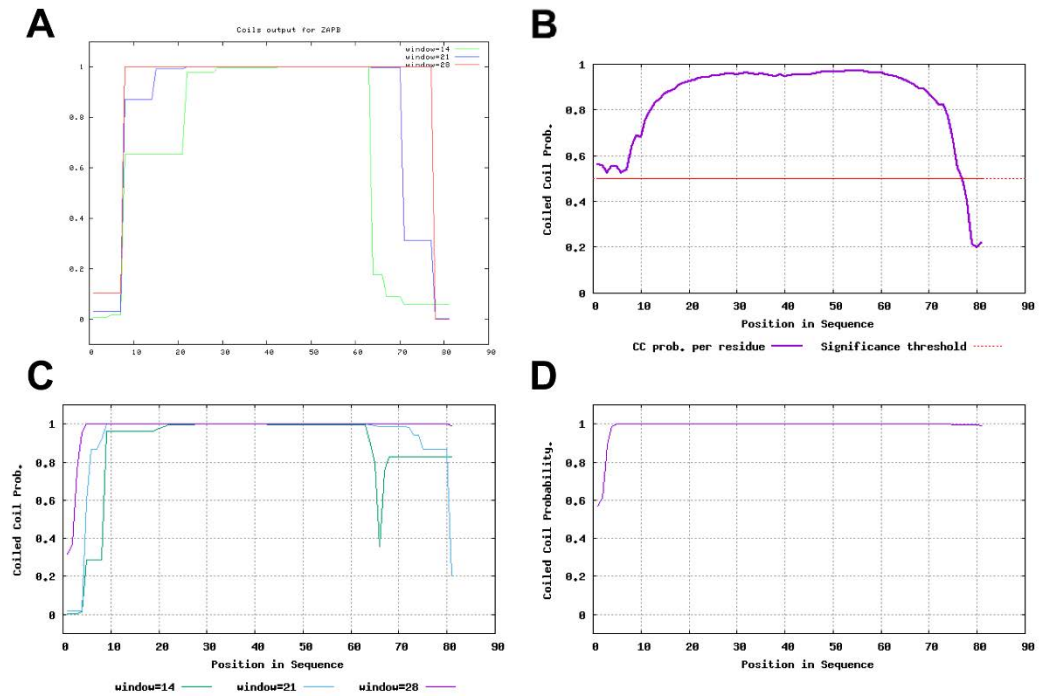
Figure S1. Coiled-coil predictions for ZapB. Coiled-coil predictions for ZapB obtained from four different servers: A) COILS, B) DeepCoil, C) PCoils and D) MARCOIL.
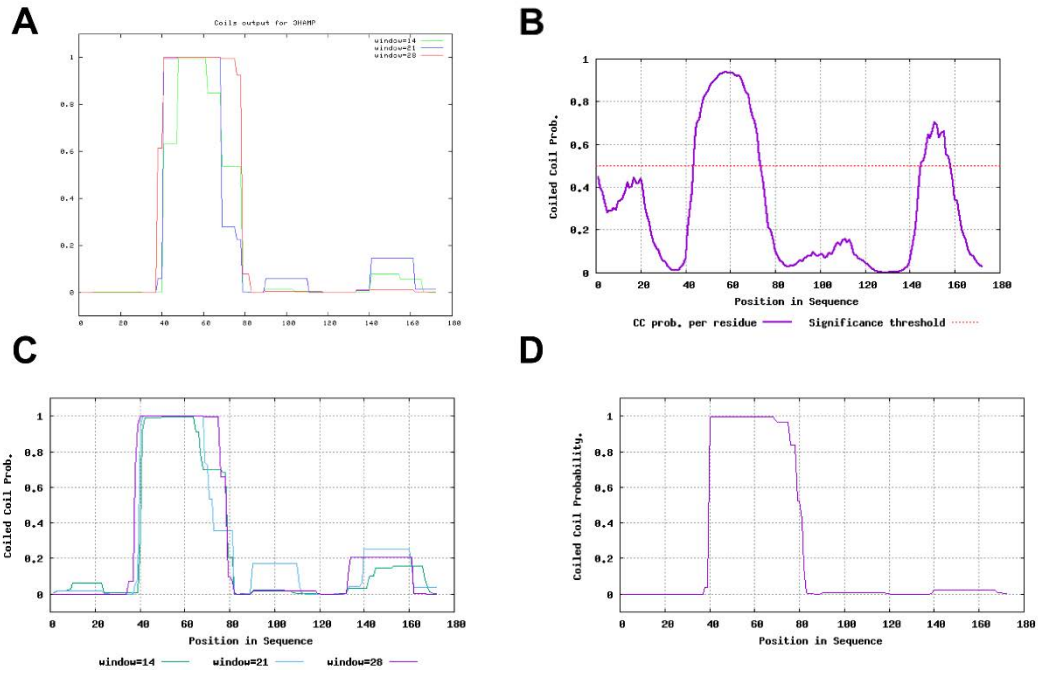
Figure S2. Coiled-coil predictions for 3HAMP. Coiled-coil predictions for 3HAMP obtained from four different servers: A) COILS, B) DeepCoil, C) PCoils and D) MARCOIL.
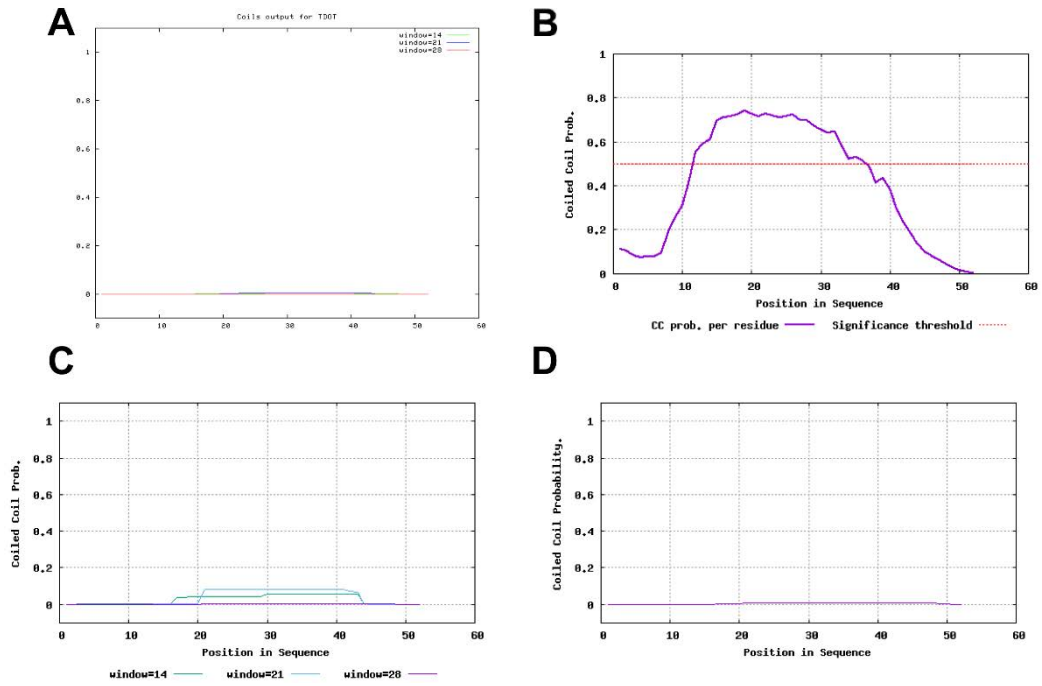
Figure S3. Coiled-coil predictions for TDoT. Coiled-coil predictions for TDoT obtained from four different servers: A) COILS, B) DeepCoil, C) PCoils and D) MARCOIL.
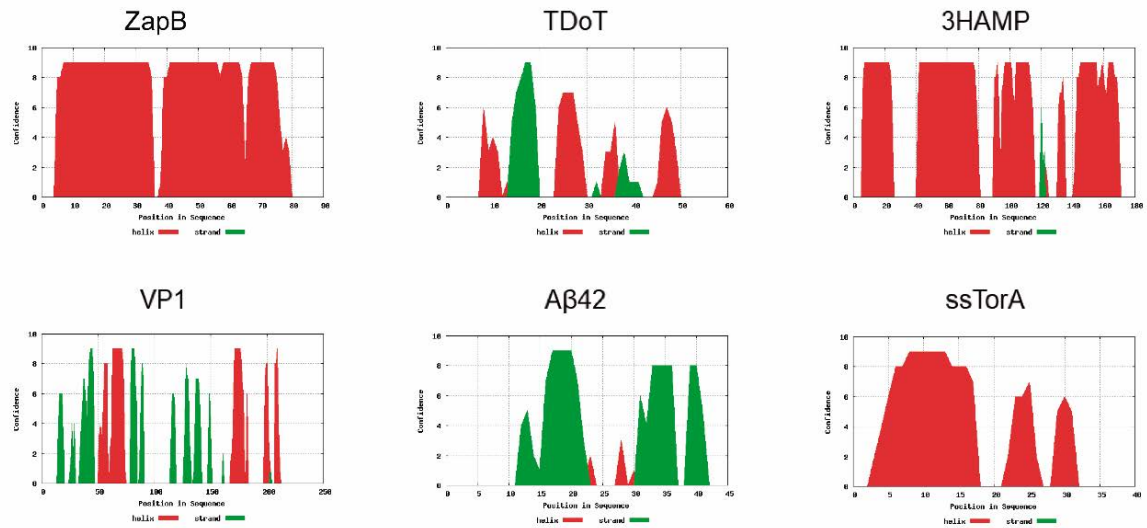
Figure S4. Secondary structure prediction by PSIPRED server. Secondary structure prediction for the

six IB-tags by PSIPRED server.
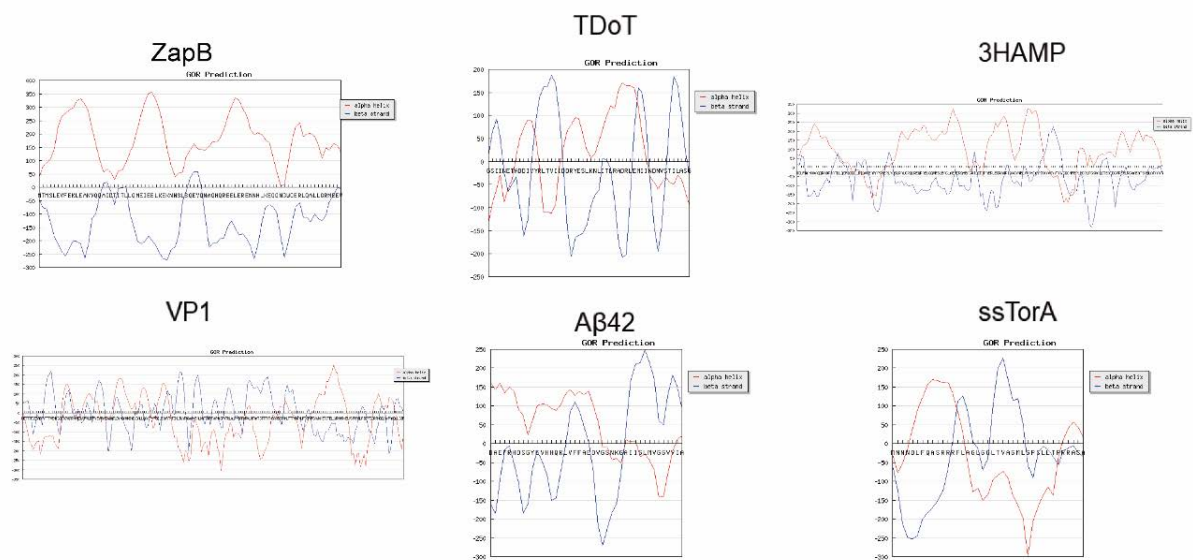
Figure S5. Secondary structure prediction by GOR server. Secondary structure prediction for the six

IB-tags by GOR server.

A3D Score: -2.53	A3D Score: -0.76	A3D Score: -1.26
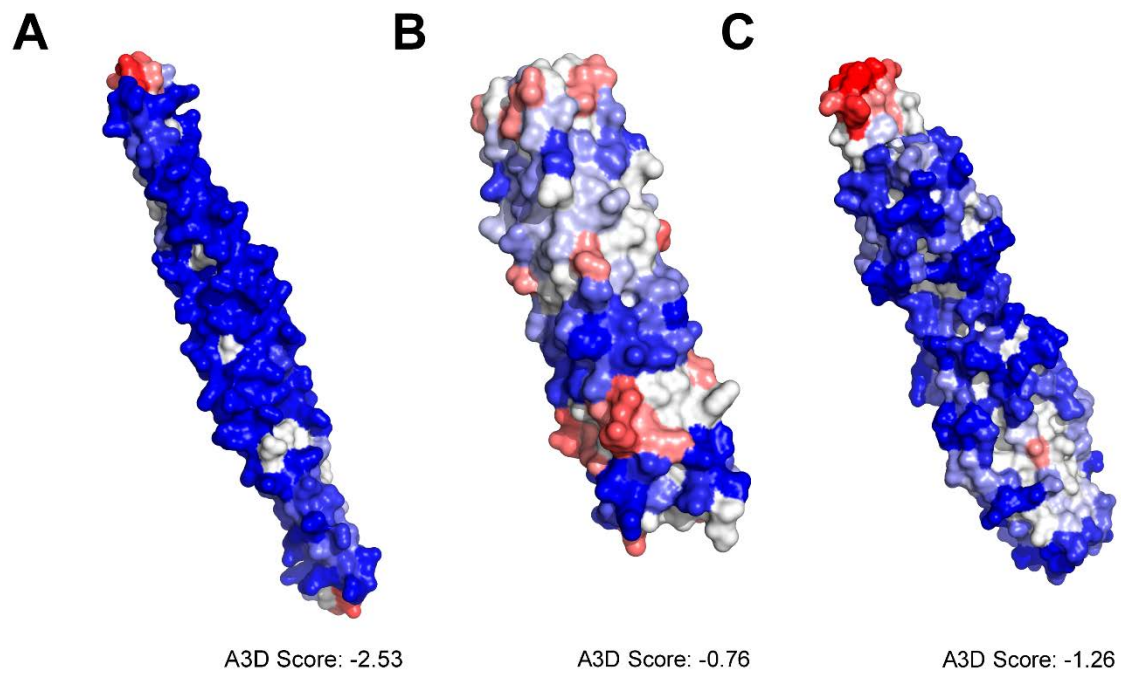
Figure S6. AGGRESCAN3D structural aggregation propensity predictions for ZapB (A), TDoT (B) and 3HAMP (C). The protein surface is colored according to the A3D score in a gradient from blue (high-predicted solubility) to white (negligible impact on protein aggregation) to red (high-predicted aggregation propensity). The A3D Score value for each structure is indicated under the structures. More negative A3D Score values indicate higher solubility.
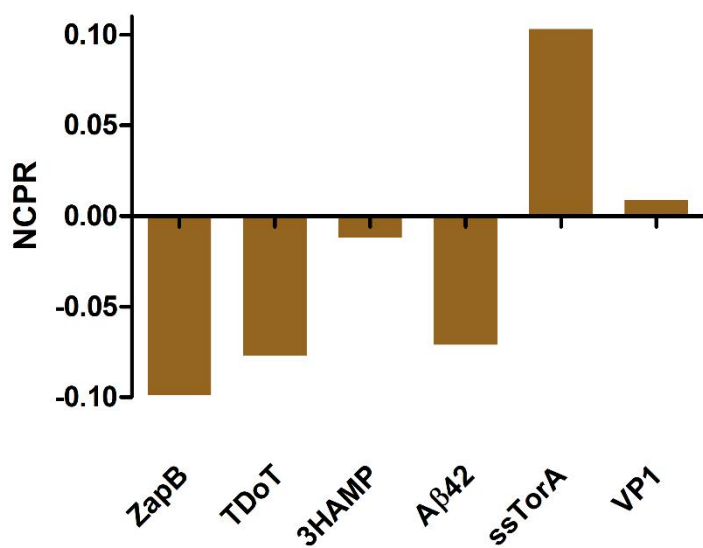
Figure S7. Net charge per residue (NCPR) of the different tags.
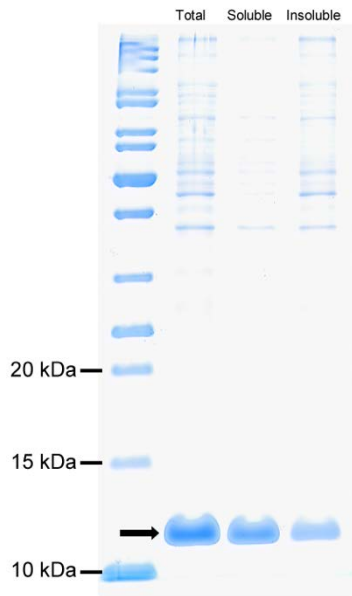
Figure S8. SDS-PAGE of the cellular distribution of ZapB.
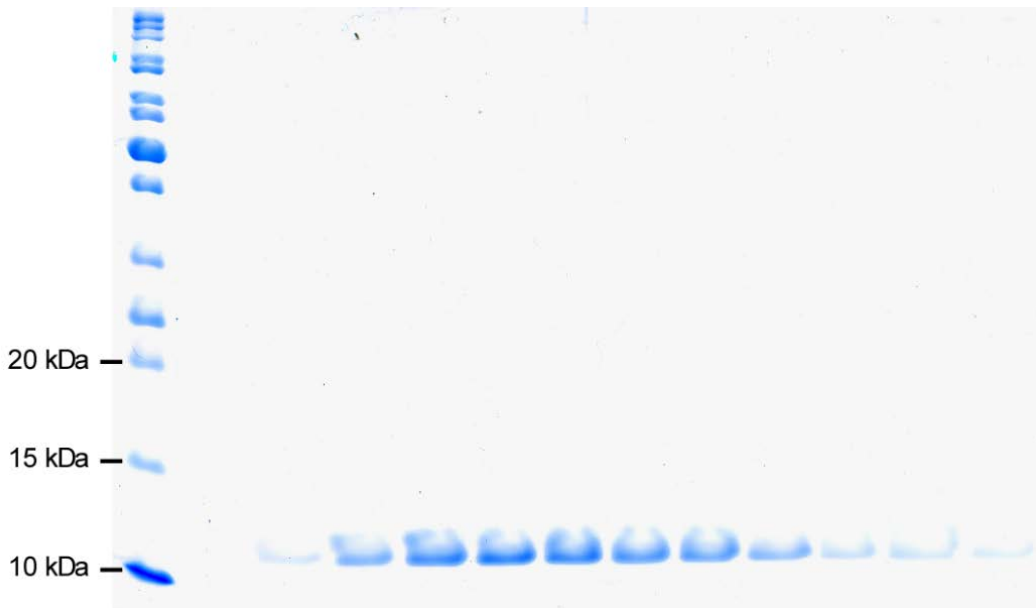
Figure S9. SDS-PAGE of ZapB purification by IMAC. Different elution fractions of ZapB IMAC purification.
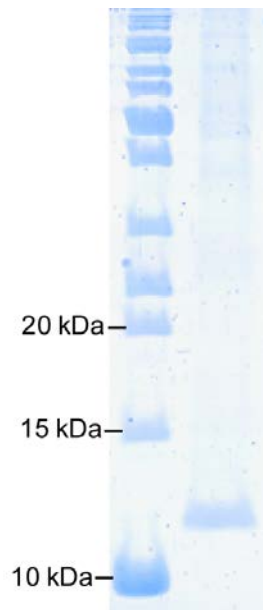
Figure S10. SDS-PAGE of purified ZapB IBs.
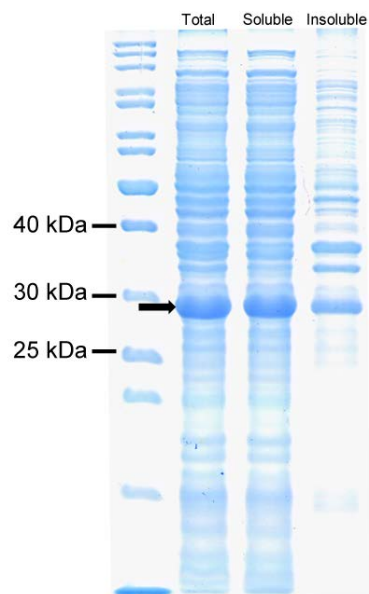
Figure S11. SDS-PAGE of the cellular distribution of GFP.
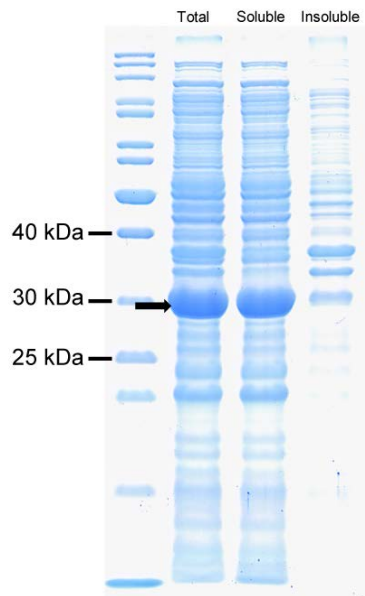
Figure S12. SDS-PAGE of the cellular distribution of mCherry.

A3D Score: -0.94
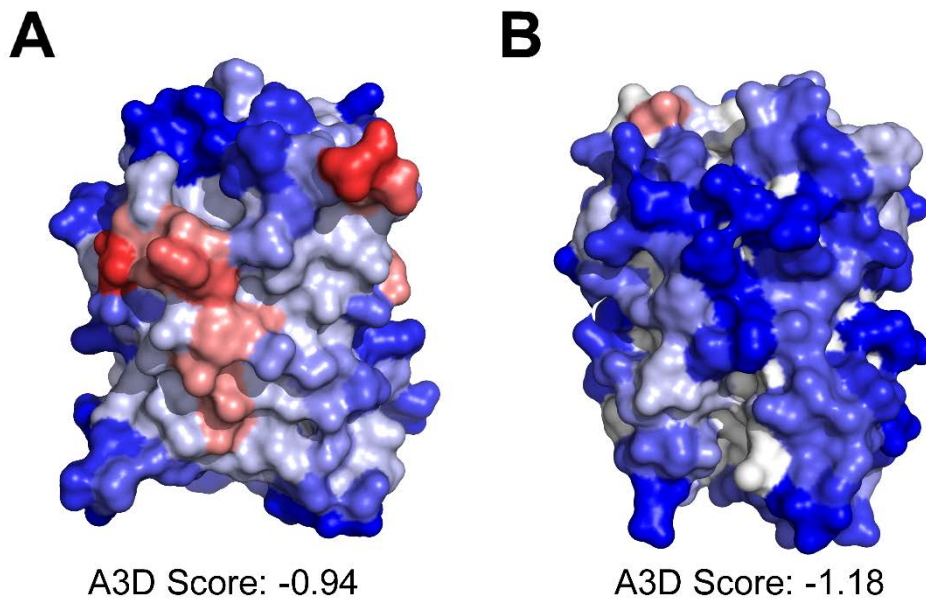
A3D Score: -1.18

Figure S13. AGGRESCAN3D structural aggregation propensity predictions for GFP (A) and mCherry (B). The protein surface is colored according to the A3D score in a gradient from blue (high-predicted solubility) to white (negligible impact on protein aggregation) to red (high-predicted aggregation propensity). The A3D Score value for each structure is indicated under the structures. More negative A3D Score values indicate higher solubility.
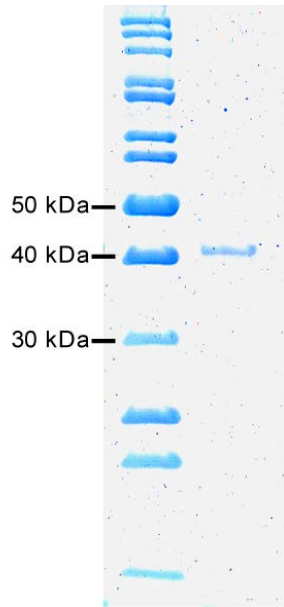
Figure S14. SDS-PAGE of purified ZapB-GFP IBs.
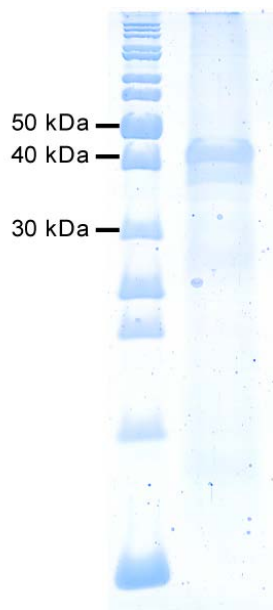
Figure S15. SDS-PAGE of purified ZapB-mCherry IBs.
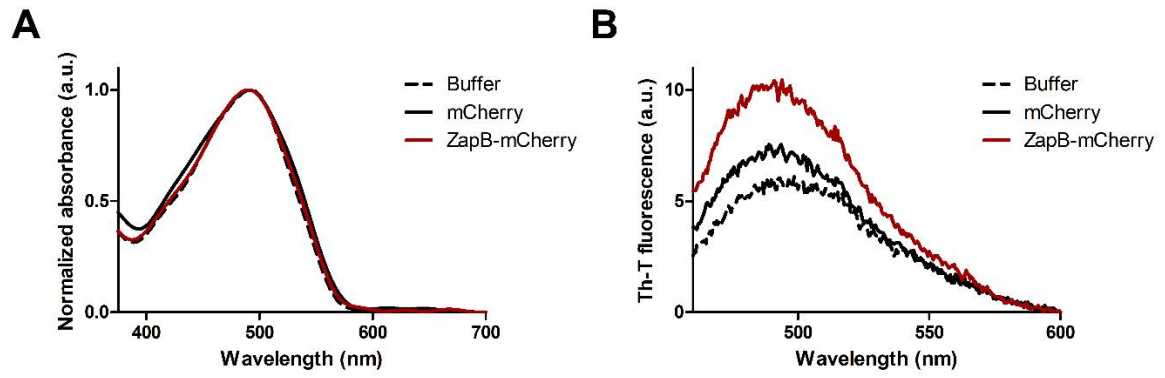
Figure S16. Characterization of the non-amyloid nature of ZapB-mCherry IBs. A) Congo-Red absorbance spectra. B) Th-T fluorescence emission spectra.
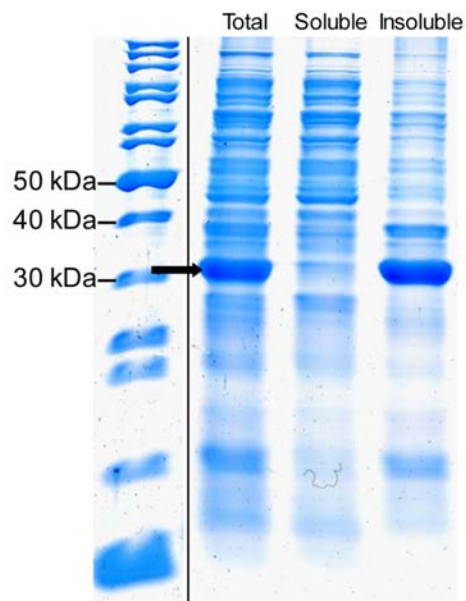
Figure S17. SDS-PAGE of the cellular distribution of Aβ42-GFP.
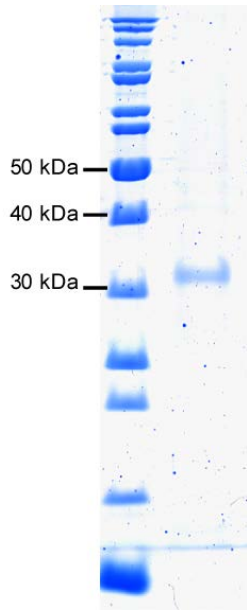
Figure S18. SDS-PAGE of purified Aβ42-GFP IBs.
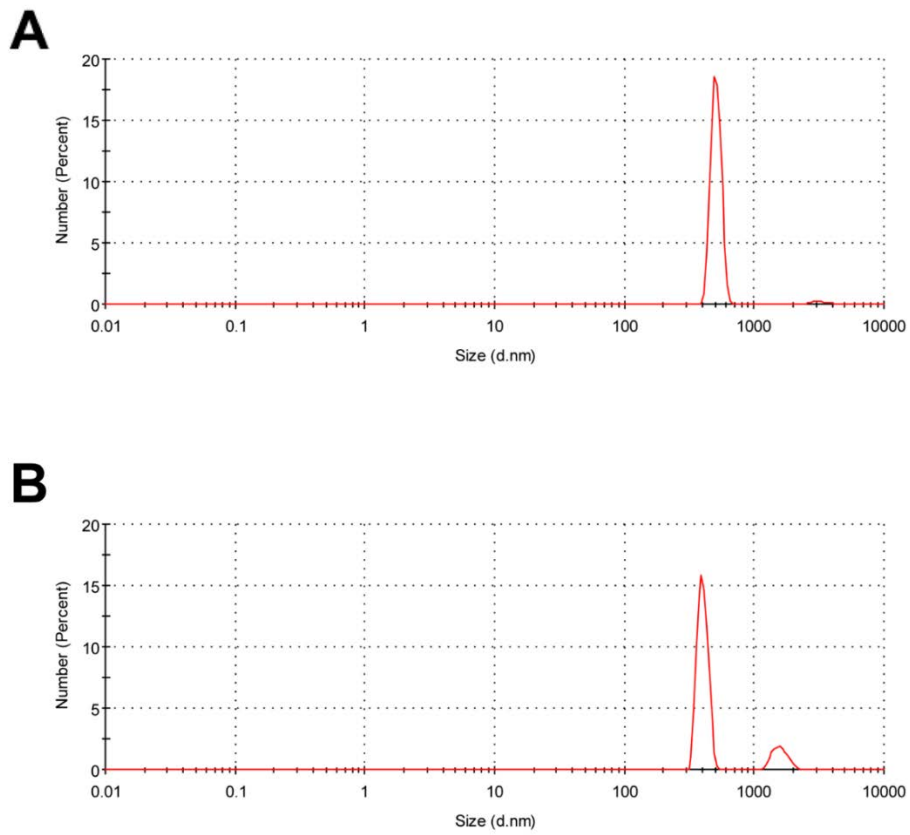
Figure S19. DLS spectra of ZapB-GFP (A) and Aβ42-GFP (B) IBs.
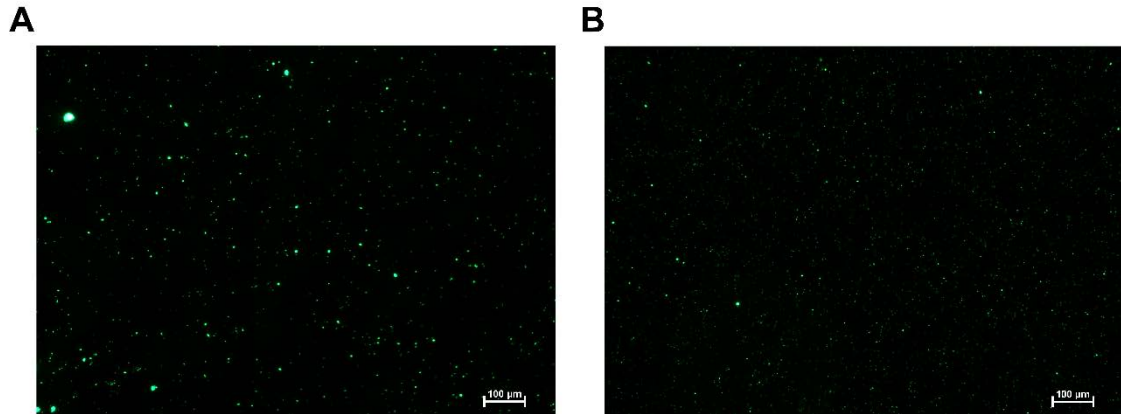
Figure S20. Epifluorescence microscopy images of ZapB-GFP (A) and Aβ42-GFP (B) IBs. A) GFP fluorescence of ZapB-GFP IBs. B) GFP fluorescence of Aβ42-GFP IBs.

DNA and amino acid sequences of ZapB protein

ZapB DNA sequence:

ATGACAATGTCATTAGAAGTGTTTGAGAAACTGGAAGCAAAAGTACAGCAGGCGATTGATACCATCAC

TCTGTTGCAGATGGAAATCGAAGAGCTGAAAGAAAAAAACAACTCACTGTCGCAGGAAGTTCAAAATG

CCCAGCATCAGCGCGAAGAGCTGGAGCGTGAGAACAACCATCTGAAAGAACAGCAGAACGGCTGGCA

GGAACGTCTGCAGGCCCTGCTGGGTCGCATGGAAGAGGTC

ZapB amino acid sequence:

MTMSLEVFEKLEAKVQQAIDTITLLQMEIEELKEKNNSLSQEVQNAQHQREELERENNHLKEQQNGWQER

LQALLGRMEEV

ANNEX: SUPPLEMENTARY INFORMATION

Chapter 4

Multifunctional Antibody-Conjugated Coiled-Coil Protein Nanoparticles for Selective Cell Targeting

Marcos Gil-Garcia and Salvador Ventura

Supplementary material for:

Multifunctional antibody-conjugated coiled-coil protein nanoparticles for selective cell targeting

Marcos Gil-Garcia[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

*Corresponding author: Salvador Ventura

Email addresses: salvador.ventura@uab.cat and marcos.gil.garcia@uab.cat

Telephone number: +34 935868956

Fax number: +34 935812011
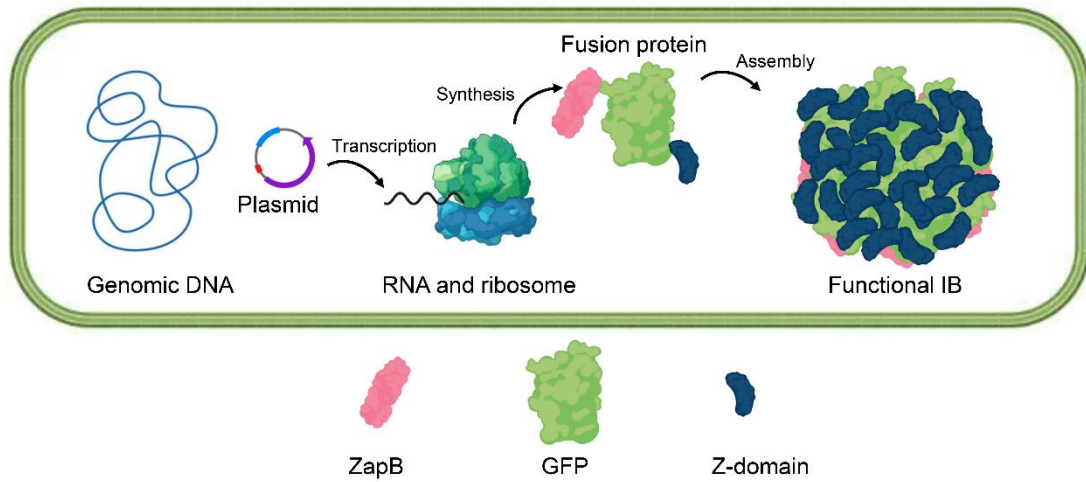
Fig. S1. Schematic representation of inclusion body (IB) formation inside microbial cells. The IBs formation starts after the induction of recombinant protein expression. Once the fusion protein is synthesized, it establishes intermolecular interactions with similar polypeptides, triggering the formation of protein-based nanoclusters, commonly known as IBs, usually located at the poles of the cell. These IBs formed by an accumulation of functional fusion protein can be easily purified as ready-to-use submicrometric nanoparticles. Created with Biorender.com.
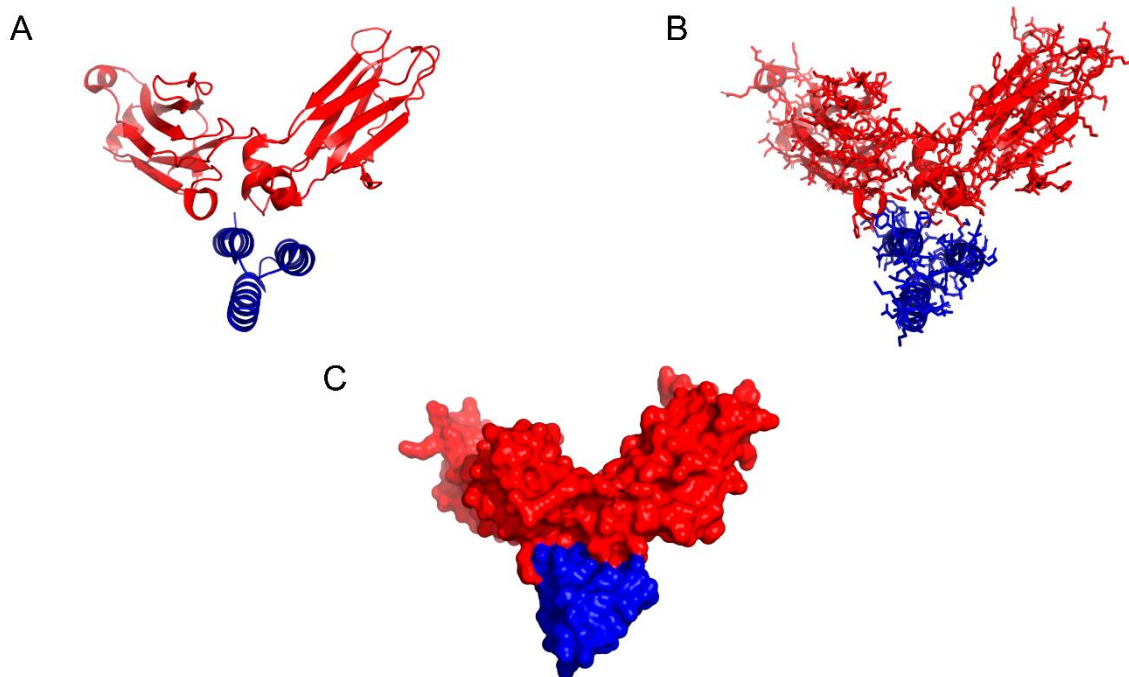
Fig. S2. Structural representation of the interaction between the Z-domain (blue) and the Fc region of an IgG (red) according to the PDB: 5U4Y. Protein structures are shown in cartoon representation (A), in cartoon representation showing the sidechains (B) and in surface representation (C).

A
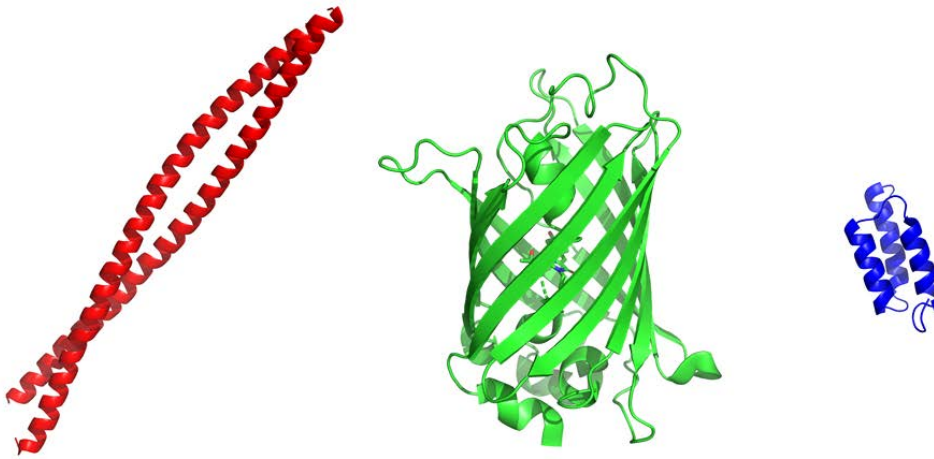
B

MGTMSLEVFEKLEAKVQQAIDTITLLQMEIEELKEKNNSLSQEVQNAQHQREELERENNHL
KEQQNGWQERLQALLGRMEEVSIPGAVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGE
GDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQE
RTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQ
KNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHM
VLLEFVTAAGITLGMDELYKSGGGSGGSVDNKFNKEQQNAFYEILHLPNLNEEQRNAFIQS
LKDDPSQSANLLAEAKKLNDAQAPKHHHHHH

Fig. S3. Schematic representation and sequence of ZapB, GFP and Z-domain. A) Cartoon
representation of ZapB (red, PDB: 2JEE), GFP (green, PDB: 2Y0G) and Z-domain (blue, PDB: 2SPZ). B)
Sequence of ZapB-GFP-Z fusion protein. ZapB, GFP, Z-domain, linkers and 6xHis-tag are shown in red,
green, blue and gray, respectively.
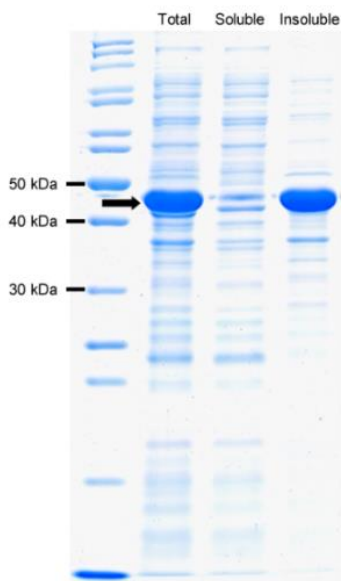
Fig. S4. SDS-PAGE of the cellular distribution of ZapB-GFP-Z. SDS-PAGE of intact cells, soluble, and
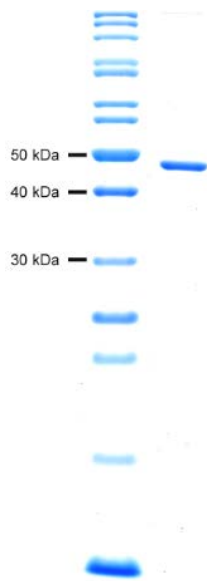insoluble cellular fractions. The arrow indicates ZapB-GFP-Z protein.

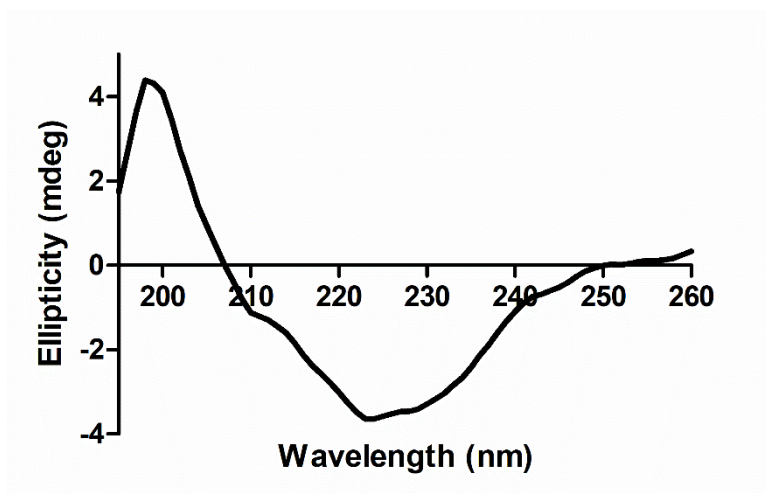**Fig. S5. SDS-PAGE of purified ZapB-GFP-Z IBs.**

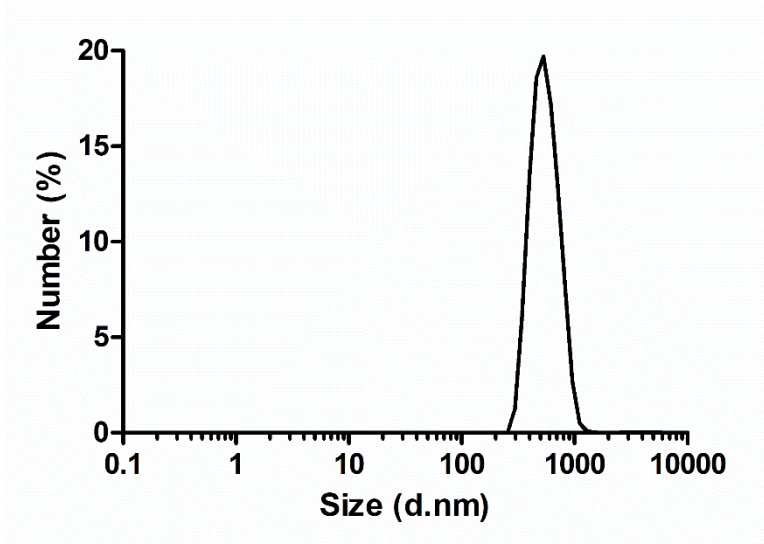**Fig. S6. Far-UV circular dichroism spectrum of ZapB-GFP-Z IBs.**

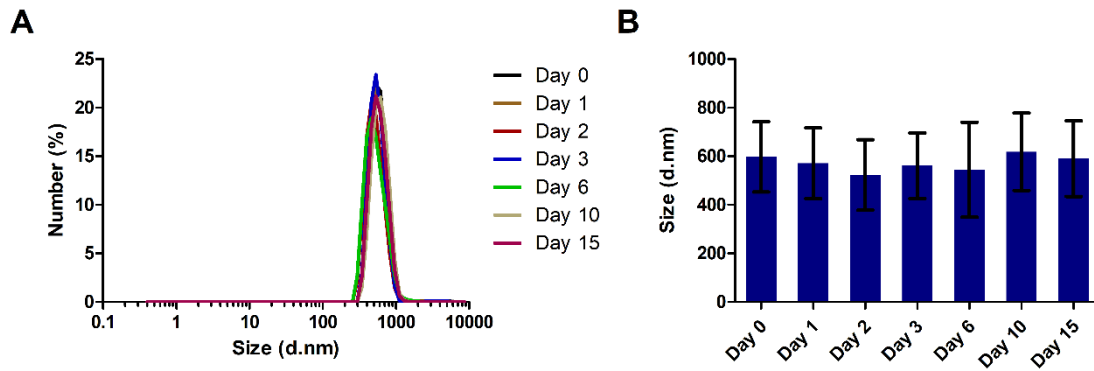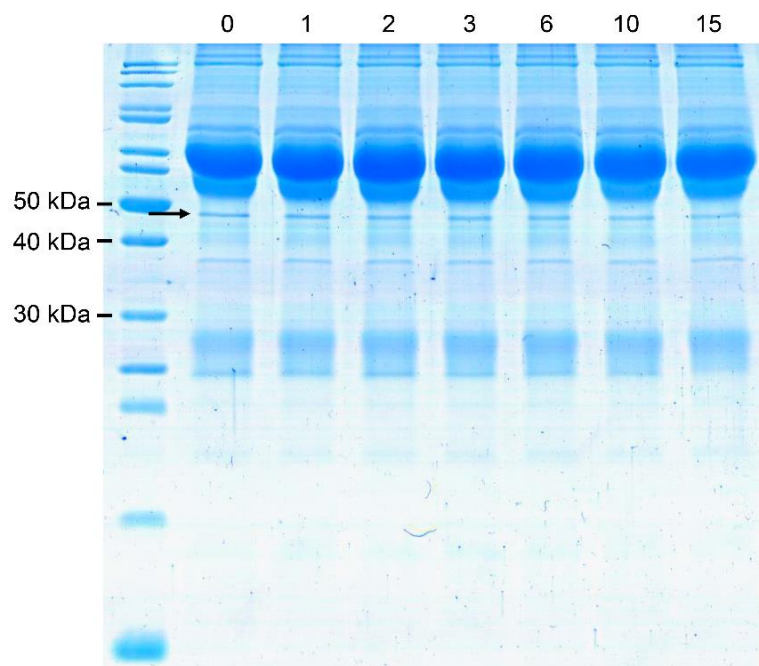**Fig. S7. DLS analysis of ZapB-GFP-Z IBs.**

Fig. S8. Stability of ZapB-GFP-Z IBs as measured by DLS. A) Size distribution graphics of ZapB-GFP-Z

IBs at specific time points. B) Size distribution values of ZapB-GFP-Z IBs during two weeks.

Fig. S9. SDS-PAGE analysis of the stability of ZapB-GFP-Z IBs in bovine serum. The different numbers at the top of each lane correspond to the time of incubation (in days). The arrow indicates the ZapB-GFP-Z protein band.
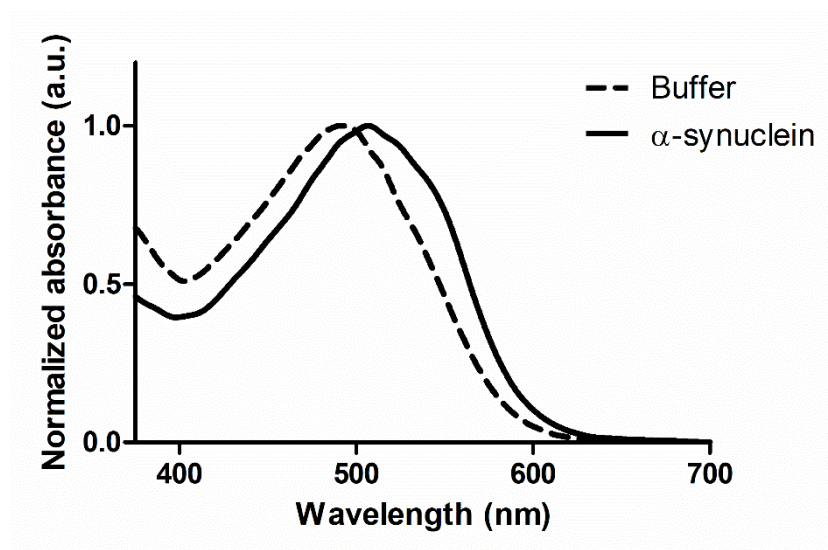
Fig. S10. CR absorbance spectra of amyloid fibrils of α-synuclein (solid line) and buffer alone (dashed line).
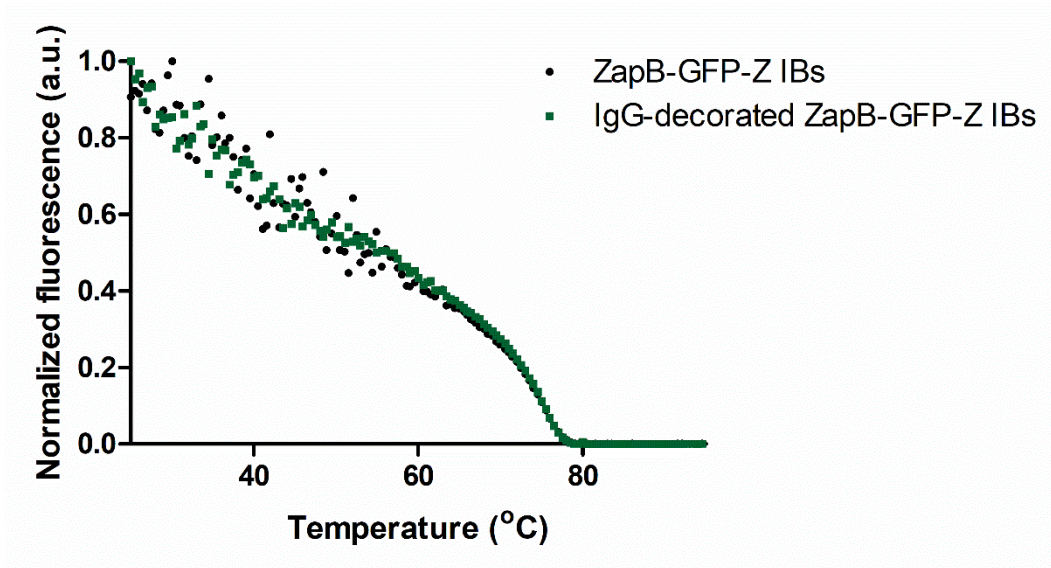
Fig. S11. Thermal stability of ZapB-GFP-Z IBs. Thermal unfolding curves monitored by changes in GFP fluorescence intensity from 25 to 95 °C of undecorated and IgG-decorated ZapB-GFP-Z IBs.
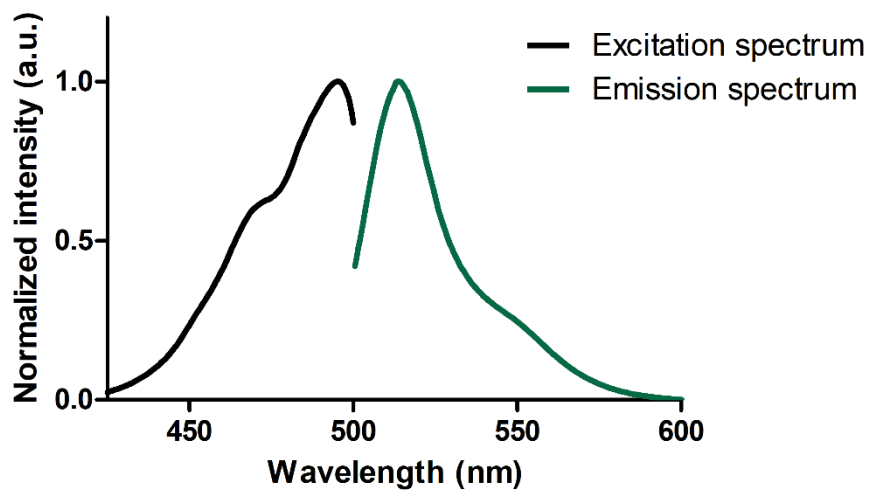

Fig. S12. Spectral properties of ZapB-GFP-Z IBs. Excitation and emission spectra of ZapB-GFP-Z IBs.

GFP fluorescence

Alexa Fluor 555 fluorescence
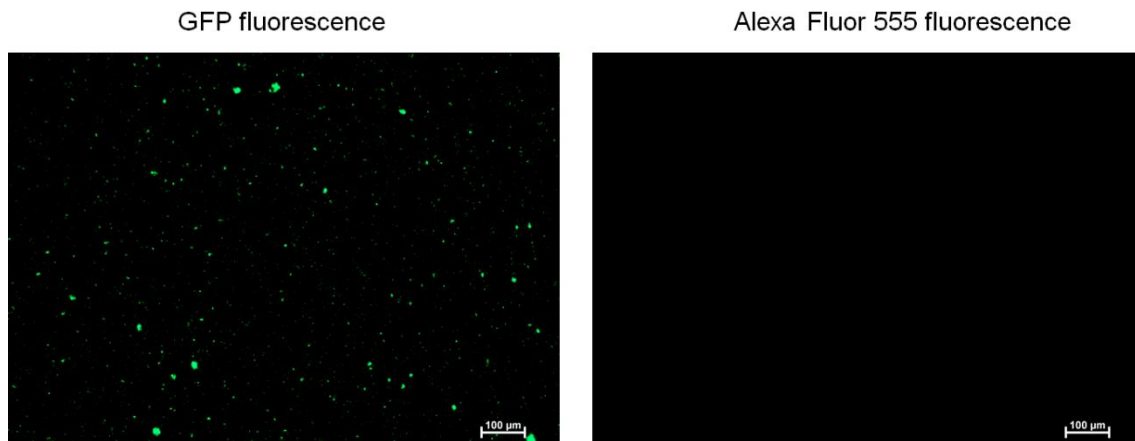
Fig. S13. Fluorescence microscopy of ZapB-GFP IBs incubated in the presence of IgG labeled with Alexa Fluor 555. Representative fluorescence microscopy images of (left) green fluorescence from ZapB-GFP IBs and (right) red fluorescence from IgG labeled with Alexa Fluor 555.

A

B
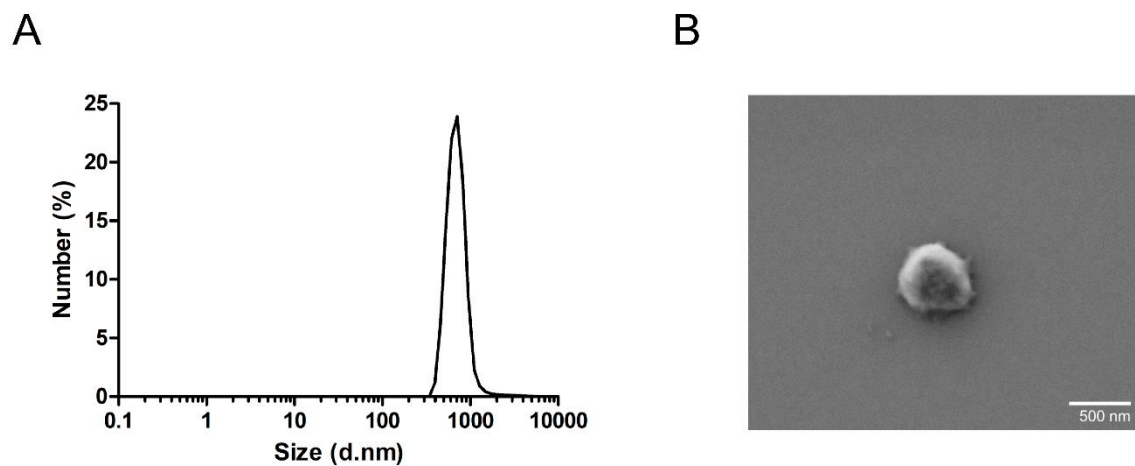
Fig. S14. Size distribution and shape of antibody-decorated ZapB-GFP-Z IBs. A) Size distribution graphic of antibody-decorated ZapB-GFP-Z IBs as measured by DLS. B) SEM image of an antibody-decorated ZapB-GFP-Z IB.

| Brightfield | GFP fluorescence |
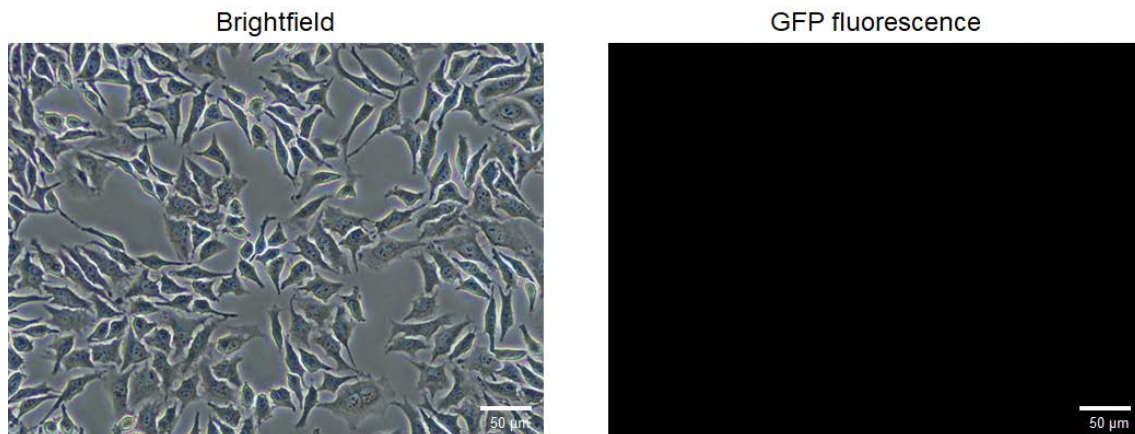
Fig. S15. Fluorescence microscopy of HeLa cells incubated with ZapB-GFP-Z IBs decorated with an anti-rabbit IgG. Representative fluorescence microscopy images of (left) brightfield and (right) green fluorescence from ZapB-GFP-Z IBs.
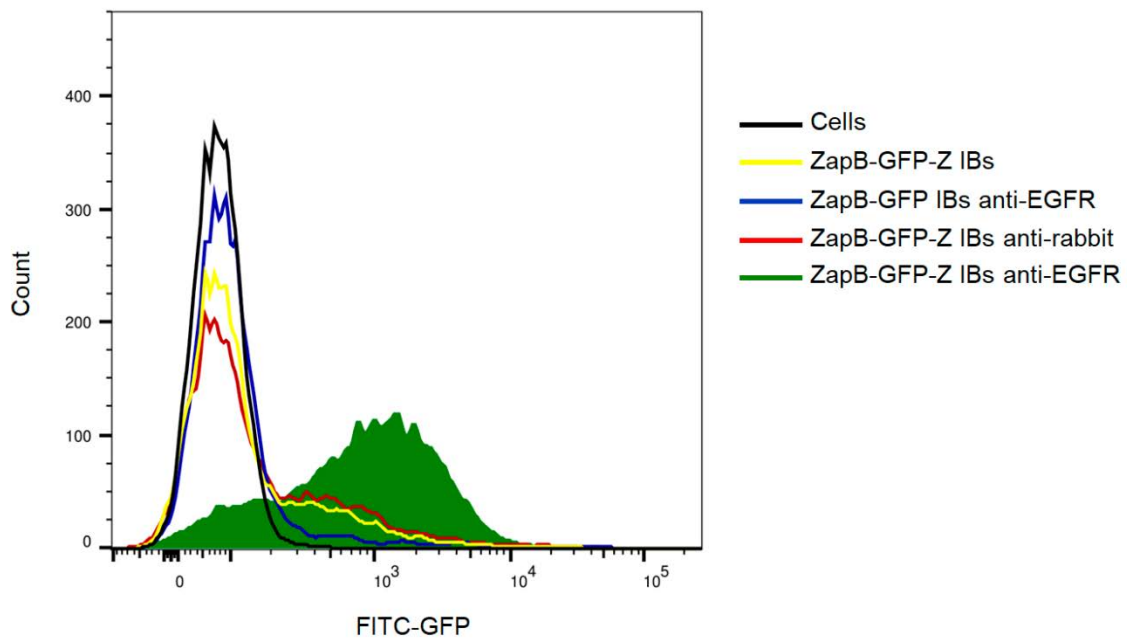
Fig. S16. Quantitative analysis by flow cytometry of the specific binding of decorated IBs to HeLa cells. HeLa cells were incubated with anti-EGFR IgG-decorated ZapB-GFP-Z IBs (green line) and different controls such as ZapB-GFP-Z IBs (yellow line), ZapB-GFP IBs previously incubated with an anti-EGFR IgG (blue line) and ZapB-GFP-Z IBs decorated with an anti-rabbit IgG (red line). The black line corresponds to non-incubated HeLa cells. The green fluorescence of GFP was used to analyze the binding capability of the different IBs to HeLa cells.
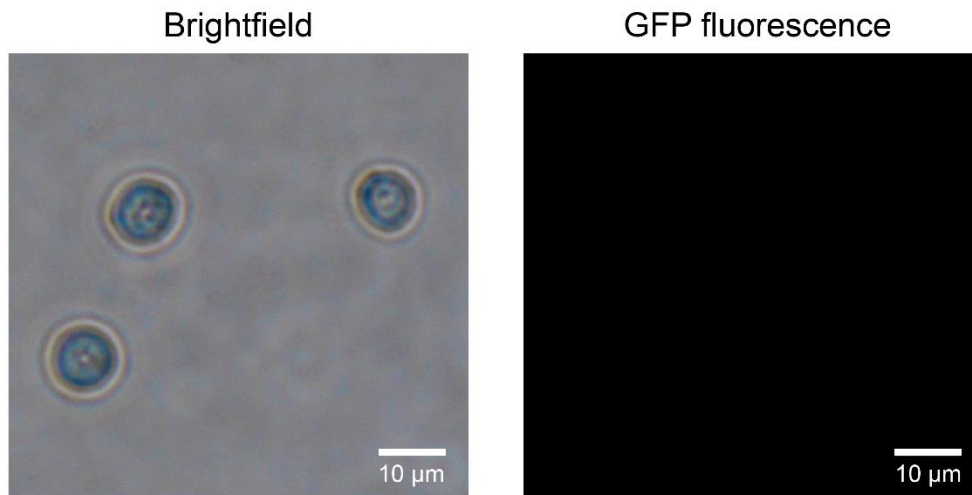
Fig. S17. Fluorescence microscopy analysis of T cells incubated with ZapB-GFP-Z IBs decorated with an anti-EGFR IgG. Representative fluorescence microscopy images of (left) brightfield and (right) green fluorescence from ZapB-GFP-Z IBs.
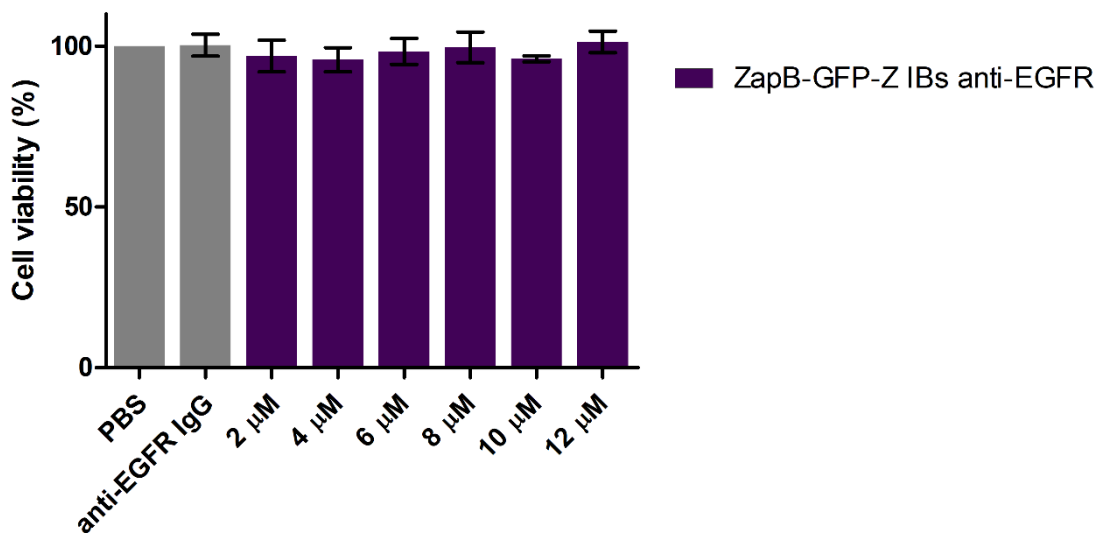

Fig. S18. Cytotoxicity induced by the ZapB-GFP-Z IBs decorated with anti-EGFR IgGs. HeLa cells viability was tested after adding different concentrations (from 2 to 12 μM) of anti-EGFR IgG-decorated ZapB-GFP-Z IBs. IBs were added to the cells and incubated for 72 h. PBS buffer and anti-EGFR IgG were used as controls.
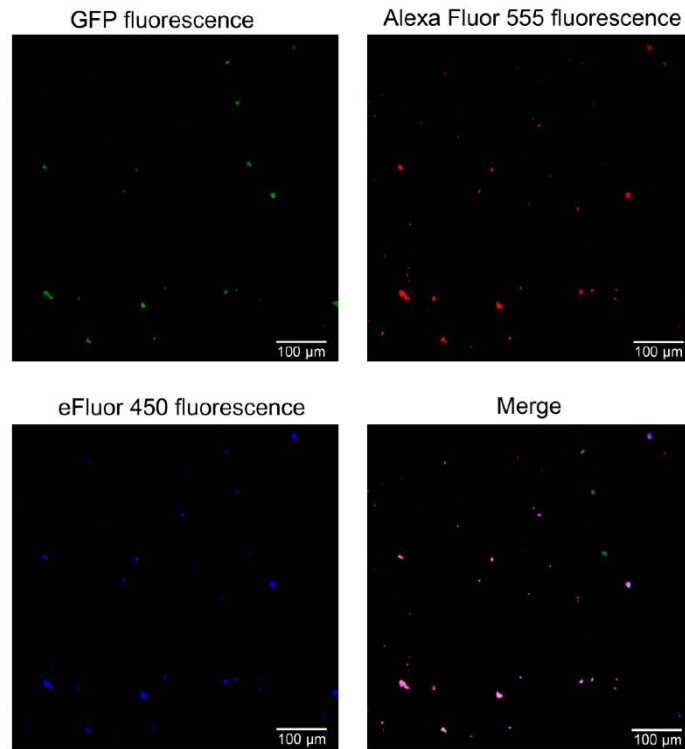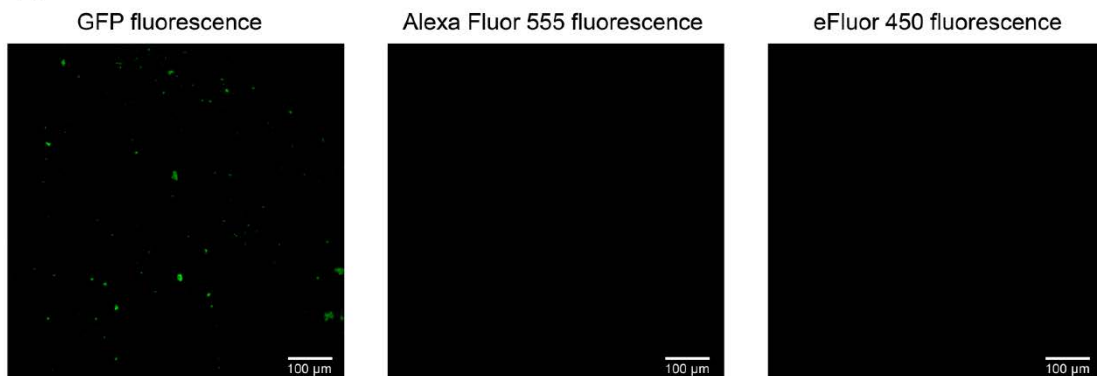
Fig. S19. Dual-antibody decoration of ZapB-GFP-Z IBs analyzed by confocal microscopy. A) Representative confocal microscopy images of antibody-decorated IBs. GFP fluorescence (top, left), red fluorescence emitted from IgG labeled with Alexa Fluor 555 (top, right), blue fluorescence emitted from IgG labeled with eFluor 450 (bottom, left), and merging of these images (bottom, right). B) Representative confocal microscopy images of undecorated IBs. GFP fluorescence channel (left), red fluorescence channel (middle), blue fluorescence channel (right).
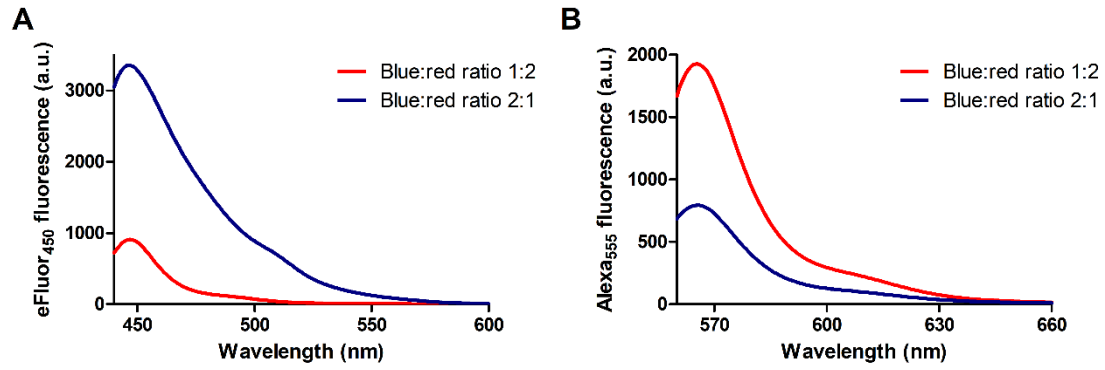
Fig. S20. Fluorescence spectra of ZapB-GFP-Z IBs incubated with two labeled antibodies at different ratios. eFluor 450 (A) and Alexa Fluor 555 (B) fluorescence spectra of ZapB-GFP-Z IBs incubated with a mixture of blue and red labeled antibodies at different ratios. The color of the curve corresponds to that of the antibody incubated at 2-fold concentration.
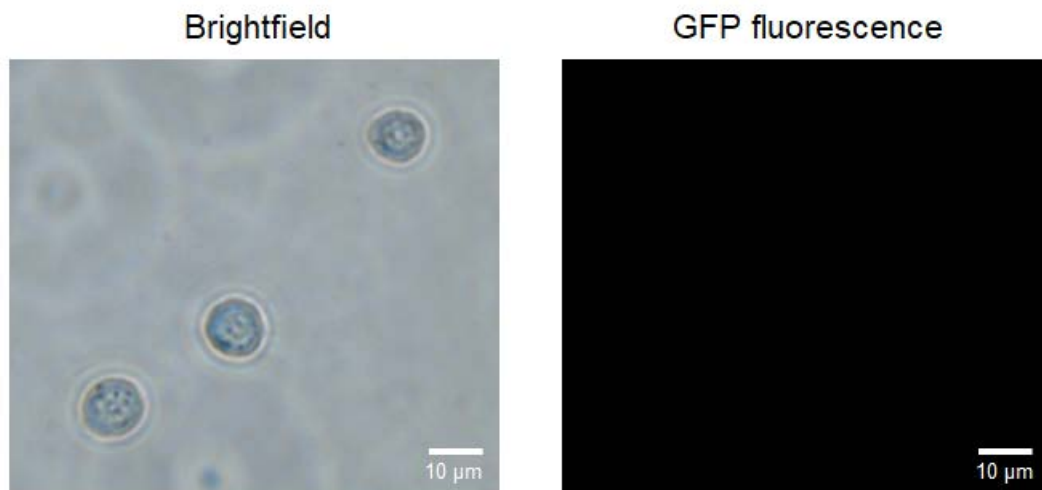

Fig. S21. Fluorescence microscopy of T cells incubated with ZapB-GFP-Z IBs decorated with an anti-rabbit IgG. Representative fluorescence microscopy images of (left) brightfield and (right) green fluorescence from ZapB-GFP-Z IBs.
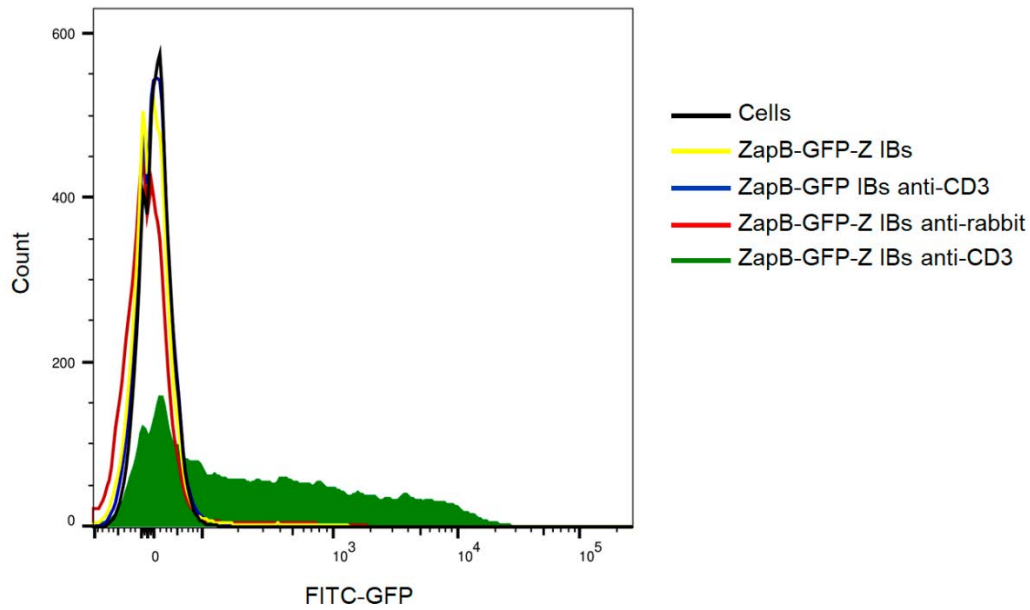
Fig. S22. Quantitative analysis by flow cytometry of the specific binding of decorated IBs to T cells. T cells were incubated with anti-CD3 IgG-decorated ZapB-GFP-Z IBs (green line) and different controls such as ZapB-GFP-Z IBs (yellow line), ZapB-GFP IBs previously incubated with an anti-CD3 IgG (blue line) and ZapB-GFP-Z IBs decorated with an anti-rabbit IgG (red line). The black line corresponds to non-incubated T cells. The green fluorescence of GFP was used to analyze the binding capability of the different IBs to T cells.
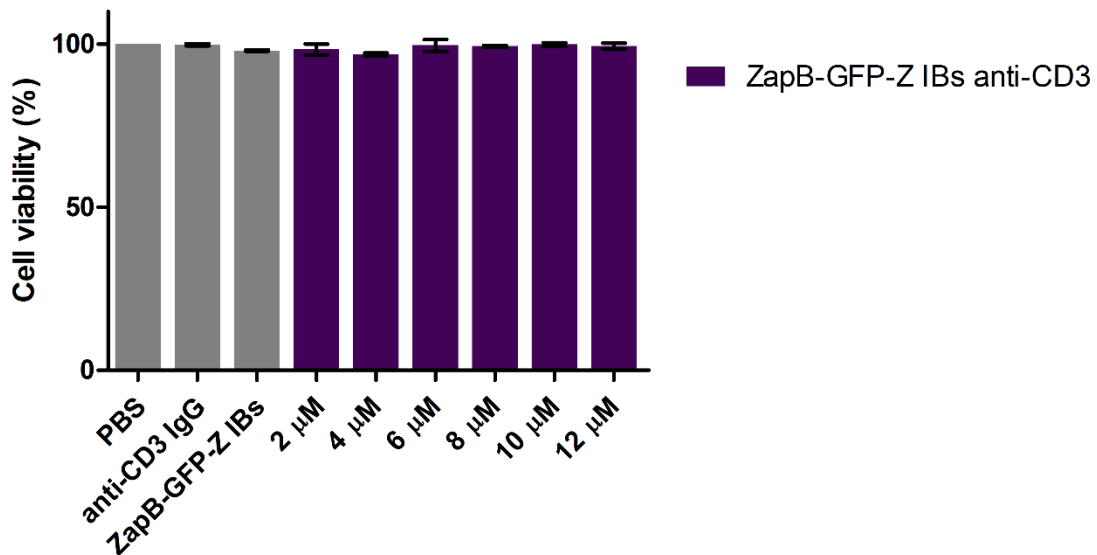
Fig. S23. Cytotoxicity induced by the ZapB-GFP-Z IBs decorated with anti-CD3 IgGs. T cells viability was tested after adding different concentrations (from 2 to 12 μM) of anti-CD3 IgG-decorated ZapB-GFP-Z IBs. IBs were added to the cells and incubated for 72 h. PBS buffer, anti-CD3 IgG and undecorated ZapB-GFP-Z IBs were used as controls.
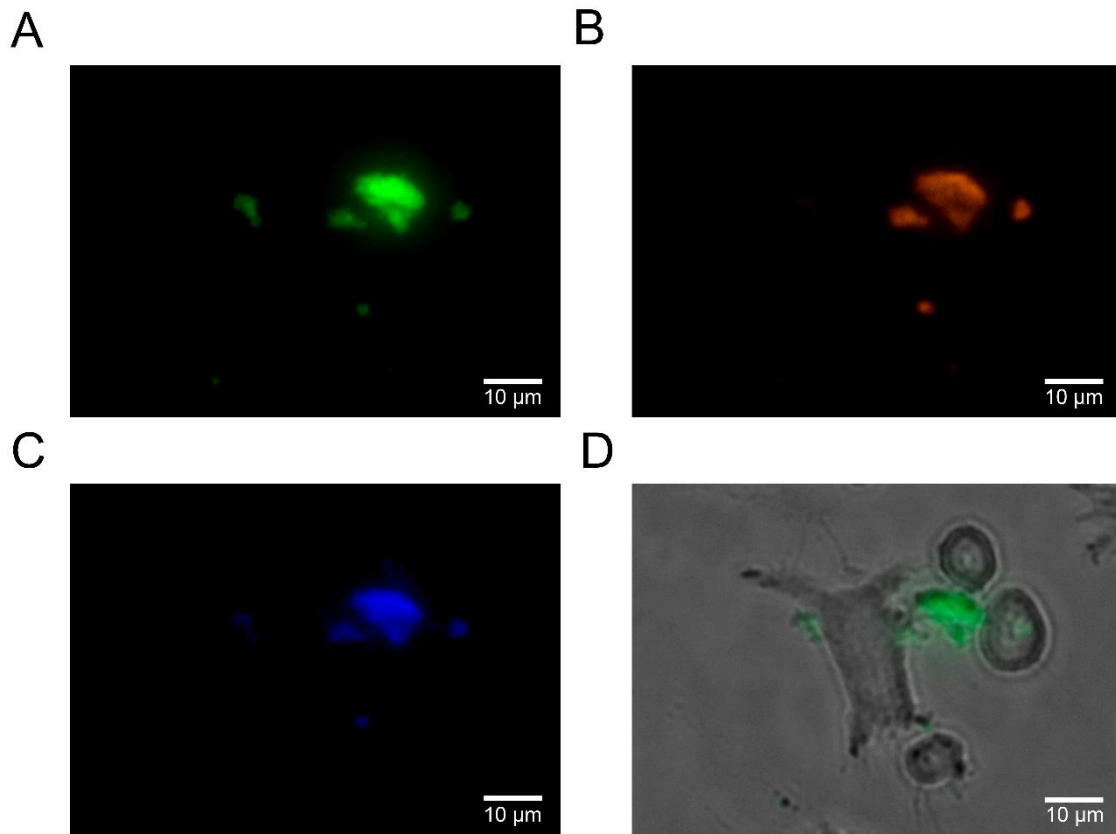
Fig. S24. ZapB-GFP-Z IBs loaded with anti-EGFR and anti-CD3 IgGs direct T cells to HeLa cells. Representative microscopy images of GFP fluorescence (A), red fluorescence emitted from anti-EGFR IgG labeled with Alexa Fluor 555 (B), blue fluorescence emitted from anti-CD3 IgG labeled with eFluor 450 (C), and merging of GFP fluorescence and brightfield (D).
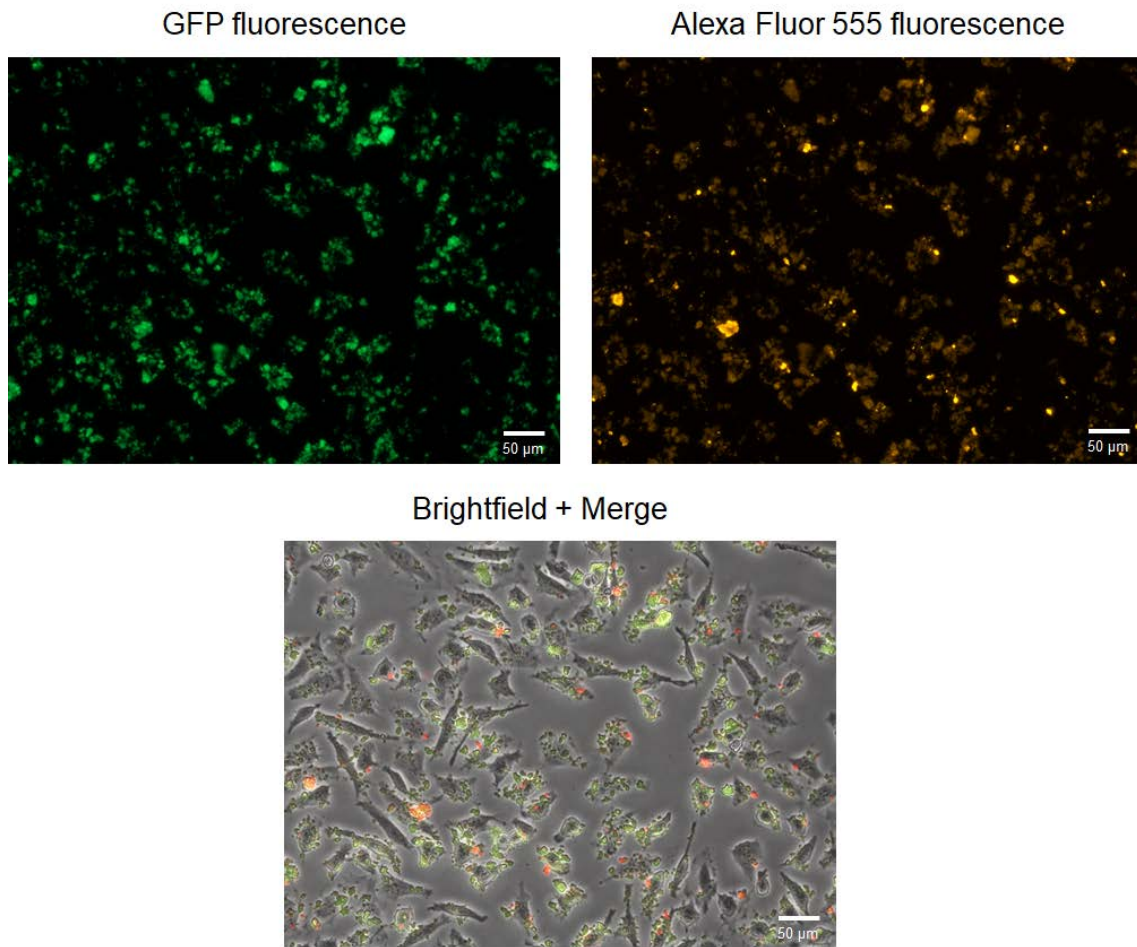
Fig. S25. Fluorescence microscopy of HeLa cells incubated with ZapB-GFP-Z IBs decorated with anti-EGFR IgG in the presence of T cells. Representative fluorescence microscopy images of (left) green fluorescence from ZapB-GFP-Z IBs, (right) red fluorescence from anti-EGFR IgG labeled with Alexa Fluor 555 and (bottom) brightfield and merge.
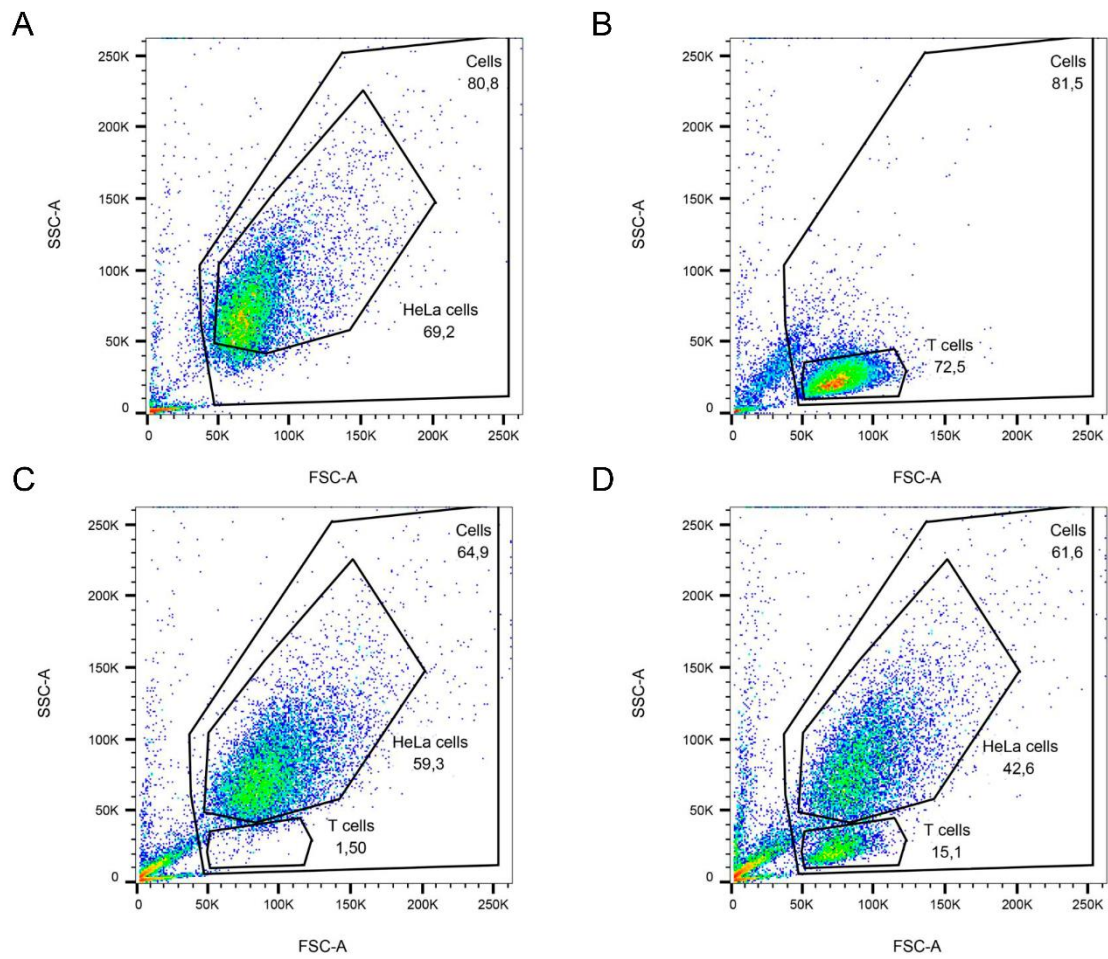
Fig. S26. Quantitative analysis by flow cytometry of the specific binding of T cells to HeLa cells mediated by dual-antibody decorated ZapB-GFP-Z IBs. Flow cytometry dot plots of HeLa cells alone (A), T cells alone (B), HeLa and T cells incubated with ZapB-GFP-Z IBs loaded only with anti-EGFR IgGs (C) and HeLa and T cells incubated with ZapB-GFP-Z IBs loaded with anti-EGFR and anti-CD3 IgGs (D).

Table S1. Quantification of the binding of T cells to HeLa cells mediated by antibody-decorated ZapB-GFP-Z IBs.

	HeLa cells (%)	T cells (%)	T cells/HeLa cells Ratio
Untreated HeLa cells	69.2	2.3*	-
Untreated T cells	3.74#	72.5	-
HeLa and T cells + anti-EGFR IgG-loaded IBs	59.3	1.5	0.025
HeLa and T cells + anti-EGFR and anti-CD3 IgG-loaded IBs	42.6	15.1	0.35

* This value corresponds to events of HeLa cells compatible in size and complexity with T cells.

This value corresponds to events of T cells compatible in size and complexity with HeLa cells.

Annex 1. DNA sequence of ZapB-GFP-Z protein. DNA sequences corresponding to ZapB, GFP and Z-domain proteins are colored in red, green and blue, respectively. DNA sequences corresponding to linkers and 6xHis-Tag are colored in black.

ACAATGTCATTAGAAGTGTTTGAGAAACTGGAAGCAAAAGTACAGCAGGCGATTGATACCATCACTCT
GTTGCAGATGGAAATCGAAGAGCTGAAAGAAAAAAACAACTCACTGTCGCAGGAAGTTCAAAATGCCC
AGCATCAGCGCGAAGAGCTGGAGCGTGAGAACAACCATCTGAAAGAACAGCAGAACGGCTGGCAGG
AACGTCTGCAGGCCCTGCTGGGTCGCATGGAAGAGGTCAGCATCCCCGGCGCCGTGAGCAAGGGC
GAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAG
TTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGC
ACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGC
TTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACG
TCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCG
AGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCC
TGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAA
CGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCA
CTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCAC
CCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGAC
CGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGAGCGGCGGCGGCAGCGGCGGCAGCG
TGGATAACAAGTTCAACAAAGAGCAGCAAAACGCGTTTTACGAAATCCTGCACCTGCCGAACCTGAAC
GAGGAACAGCGTAACGCGTTCATTCAAAGCCTGAAGGACGATCCGAGCCAAAGCGCGAACCTGCTG
GCGGAAGCGAAAAAACTGAACGATGCGCAAGCGCCGAAACATCACCATCACCATCAC

REFERENCES

Almeida, Z. L., and Brito, R. M. M. (2020). Structure and Aggregation Mechanisms in Amyloids. *Molecules (Basel, Switzerland)* 25. doi:10.3390/molecules25051195.

Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science (New York, N.Y.)* 181, 223–230. doi:10.1126/science.181.4096.223.

Anfinsen, C. B., Haber, E., Sela, M., and White, F. H. J. (1961). 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* 47, 1309–1314. doi:10.1073/pnas.47.9.1309.

Bahramali, G., Goliaei, B., Minuchehr, Z., and Marashi, S.-A. (2017). A network biology approach to understanding the importance of chameleon proteins in human physiology and pathology. *Amino acids* 49, 303–315. doi:10.1007/s00726-016-2361-6.

Bai, Y., Luo, Q., and Liu, J. (2016). Protein self-assembly via supramolecular strategies. *Chemical Society reviews* 45, 2756–2767. doi:10.1039/c6cs00004e.

Baker, D. (2000). A surprising simplicity to protein folding. *Nature* 405, 39–42. doi:10.1038/35011000.

Batlle, C., Iglesias, V., Navarro, S., and Ventura, S. (2017). Prion-like proteins and their computational identification in proteomes. *Expert review of proteomics* 14, 335–350. doi:10.1080/14789450.2017.1304214.

Beerten, J., Schymkowitz, J., and Rousseau, F. (2012). Aggregation prone regions and gatekeeping residues in protein sequences. *Current topics in medicinal chemistry* 12, 2470–2478. doi:10.2174/1568026611212220003.

Ben-Sasson, A. J., Watson, J. L., Sheffler, W., Johnson, M. C., Bittleston, A., Somasundaram, L., et al. (2021). Design of biologically active binary protein 2D materials. *Nature* 589, 468–473. doi:10.1038/s41586-020-03120-8.

Bhaskar, S., and Lim, S. (2017). Engineering protein nanocages as carriers for biomedical applications. *NPG Asia materials* 9, e371. doi:10.1038/am.2016.128.

Bielajew, B. J., Hu, J. C., and Athanasiou, K. A. (2020). Collagen: quantification, biomechanics, and role of minor subtypes in cartilage. *Nature reviews. Materials*

5, 730–747. doi:10.1038/s41578-020-0213-1.

Black, S. D., and Mould, D. R. (1991). Development of hydrophobicity parameters to analyze proteins which bear post- or cotranslational modifications. *Analytical biochemistry* 193, 72–82. doi:10.1016/0003-2697(91)90045-u.

Boeynaems, S., Alberti, S., Fawzi, N. L., Mittag, T., Polymenidou, M., Rousseau, F., et al. (2018). Protein Phase Separation: A New Phase in Cell Biology. *Trends in cell biology* 28, 420–435. doi:10.1016/j.tcb.2018.02.004.

Boke, E., and Mitchison, T. J. (2017). The balbiani body and the concept of physiological amyloids. *Cell cycle (Georgetown, Tex.)* 16, 153–154. doi:10.1080/15384101.2016.1241605.

Boke, E., Ruer, M., Wühr, M., Coughlin, M., Lemaitre, R., Gygi, S. P., et al. (2016). Amyloid-like Self-Assembly of a Cellular Compartment. *Cell* 166, 637–650. doi:10.1016/j.cell.2016.06.051.

Bolisetty, S., and Mezzenga, R. (2016). Amyloid-carbon hybrid membranes for universal water purification. *Nature nanotechnology* 11, 365–371. doi:10.1038/nnano.2015.310.

Bommana, R., Chai, Q., Schöneich, C., Weiss, W. F. 4th, and Majumdar, R. (2018). Understanding the Increased Aggregation Propensity of a Light-Exposed IgG1 Monoclonal Antibody Using Hydrogen Exchange Mass Spectrometry, Biophysical Characterization, and Structural Analysis. *Journal of pharmaceutical sciences* 107, 1498–1511. doi:10.1016/j.xphs.2018.01.017.

Camilloni, C., Sala, B. M., Sormanni, P., Porcari, R., Corazza, A., Rosa, M. De, et al. (2016). Rational design of mutations that change the aggregation rate of a protein while maintaining its native structure and stability. *Nature Publishing Group*, 1–11. doi:10.1038/srep25559.

Capasso Palmiero, U., Küffner, A. M., Krumeich, F., Faltova, L., and Arosio, P. (2020). Adaptive Chemoenzymatic Microreactors Composed of Inorganic Nanoparticles and Bioinspired Intrinsically Disordered Proteins. *Angewandte Chemie (International ed. in English)* 59, 8138–8142. doi:10.1002/anie.202000835.

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

computational biology 5, e1000476. doi:10.1371/journal.pcbi.1000476.

Céspedes, M. V., Cano-Garrido, O., Álamo, P., Sala, R., Gallardo, A., Serna, N., et al. (2020). Engineering Secretory Amyloids for Remote and Highly Selective Destruction of Metastatic Foci. *Advanced materials (Deerfield Beach, Fla.)* 32, e1907348. doi:10.1002/adma.201907348.

Chakrabortee, S., Kayatekin, C., Newby, G. A., Mendillo, M. L., Lancaster, A., and Lindquist, S. (2016). Luminidependens (LD) is an Arabidopsis protein with prion behavior. *Proceedings of the National Academy of Sciences of the United States of America* 113, 6065–6070. doi:10.1073/pnas.1604478113.

Chakraborti, S., and Chakrabarti, P. (2019). Self-Assembly of Ferritin: Structure, Biological Function and Potential Applications in Nanotechnology. *Advances in experimental medicine and biology* 1174, 313–329. doi:10.1007/978-981-13-9791-2_10.

Chen, Z., Johnson, M. C., Chen, J., Bick, M. J., Boyken, S. E., Lin, B., et al. (2019). Self-Assembling 2D Arrays with de Novo Protein Building Blocks. *Journal of the American Chemical Society* 141, 8891–8895. doi:10.1021/jacs.9b01978.

Chennamsetty, N., Voynov, V., Kayser, V., Helk, B., and Trout, B. L. (2009). Design of therapeutic proteins with enhanced stability. *Proceedings of the National Academy of Sciences of the United States of America* 106, 11937–11942. doi:10.1073/pnas.0904191106.

Chennamsetty, N., Voynov, V., Kayser, V., Helk, B., and Trout, B. L. (2011). Prediction of protein binding regions. *Proteins* 79, 888–897. doi:10.1002/prot.22926.

Chiti, F., and Dobson, C. M. (2017). Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade. *Annual Review of Biochemistry* 86, 27–68. doi:10.1146/annurev-biochem-061516-045115.

Chiti, F., Stefani, M., Taddei, N., Ramponi, G., and Dobson, C. M. (2003). Rationalization of the effects of mutations on peptide and protein aggregation rates. *Nature* 424, 805–808. doi:10.1038/nature01891.

Chiti, F., Taddei, N., Baroni, F., Capanni, C., Stefani, M., Ramponi, G., et al. (2002). Kinetic partitioning of protein folding and aggregation. *Nature structural*

biology 9, 137–143. doi:10.1038/nsb752.

Choi, J.-M., Holehouse, A. S., and Pappu, R. V (2020). Physical Principles Underlying the Complex Biology of Intracellular Phase Transitions. *Annual review of biophysics* 49, 107–133. doi:10.1146/annurev-biophys-121219-081629.

Ciryam, P., Kundra, R., Morimoto, R. I., Dobson, C. M., and Vendruscolo, M. (2015). Supersaturation is a major driving force for protein aggregation in neurodegenerative diseases. *Trends in Pharmacological Sciences* 36, 72–77. doi:10.1016/j.tips.2014.12.004.

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

Crick, F. H. C. (1953). The packing of α-helices: simple coiled-coils. *Acta Crystallographica* 6, 689–697. doi:10.1107/s0365110x53001964.

Daggett, V., and Fersht, A. R. (2003). Is there a unifying mechanism for protein folding? *Trends in biochemical sciences* 28, 18–25. doi:10.1016/s0968-0004(02)00012-9.

De Baets, G., Van Durme, J., Rousseau, F., and Schymkowitz, J. (2014). A genome-wide sequence-structure analysis suggests aggregation gatekeepers constitute an evolutionary constrained functional class. *Journal of molecular biology* 426, 2405–2412. doi:10.1016/j.jmb.2014.04.007.

De Baets, G., Van Durme, J., Van Der Kant, R., Schymkowitz, J., and Rousseau, F. (2015). Solubis: Optimize your protein. *Bioinformatics* 31, 2580–2582. doi:10.1093/bioinformatics/btv162.

De Marco, A., Ferrer-Miralles, N., Garcia-Fruitós, E., Mitraki, A., Peternel, S., Rinas, U., et al. (2019). Bacterial inclusion bodies are industrially exploitable amyloids. *FEMS Microbiology Reviews* 43, 53–72. doi:10.1093/femsre/fuy038.

Díaz-Caballero, M., Navarro, S., Fuentes, I., Teixidor, F., and Ventura, S. (2018). Minimalist Prion-Inspired Polar Self-Assembling Peptides. *ACS nano* 12, 5394–5407. doi:10.1021/acsnano.8b00417.

Díaz-Caballero, M., Navarro, S., Nuez-Martínez, M., Peccati, F., Rodríguez-Santiago,

L., Sodupe, M., et al. (2021). pH-Responsive Self-Assembly of Amyloid Fibrils for Dual Hydrolase-Oxidase Reactions. *ACS Catalysis* 11, 595–607. doi:10.1021/acscatal.0c03093.

Díaz-Caballero, M., Navarro, S., and Ventura, S. (2020). Soluble Assemblies in the Fibrillation Pathway of Prion-Inspired Artificial Functional Amyloids are Highly Cytotoxic. *Biomacromolecules* 21, 2334–2345. doi:10.1021/acs.biomac.0c00271.

Dignon, G. L., Best, R. B., and Mittal, J. (2020). Biomolecular Phase Separation: From Molecular Driving Forces to Macroscopic Properties. *Annual review of physical chemistry* 71, 53–75. doi:10.1146/annurev-physchem-071819-113553.

Divine, R., Dang, H. V, Ueda, G., Fallas, J. A., Vulovic, I., Sheffler, W., et al. (2021). Designed proteins assemble antibodies into modular nanocages. *Science (New York, N.Y.)* 372. doi:10.1126/science.abd9994.

Dobson, C. M. (2003). Protein folding and misfolding. *Nature* 426, 884–890. Available at: http://dx.doi.org/10.1038/nature02261.

Dobson, C. M., Šali, A., and Karplus, M. (1998). Protein Folding: A Perspective from Theory and Experiment. *Angewandte Chemie (International ed. in English)* 37, 868–893. doi:10.1002/(SICI)1521-3773(19980420)37:7<868::AID-ANIE868>3.0.CO;2-H.

DuBay, K. F., Pawar, A. P., Chiti, F., Zurdo, J., Dobson, C. M., and Vendruscolo, M. (2004). Prediction of the absolute aggregation rates of amyloidogenic polypeptide chains. *Journal of molecular biology* 341, 1317–1326. doi:10.1016/j.jmb.2004.06.043.

Dudgeon, K., Rouet, R., Kokmeijer, I., Schofield, P., Stolp, J., Langley, D., et al. (2012). General strategy for the generation of human antibody variable domains with increased aggregation resistance. *Proceedings of the National Academy of Sciences* 109, 10879–10884. doi:10.1073/pnas.1202866109.

Emily, M., Talvas, A., and Delamarche, C. (2013). MetAmyl: a META-predictor for AMYLoid proteins. *PloS one* 8, e79722. doi:10.1371/journal.pone.0079722.

Eriksson, M., Hassan, S., Larsson, R., Linder, S., Ramqvist, T., Lövborg, H., et al. (2009). Utilization of a right-handed coiled-coil protein from archaebacterium Staphylothermus marinus as a carrier for cisplatin. *Anticancer research* 29, 11–

18.

Família, C., Dennison, S. R., Quintas, A., and Phoenix, D. A. (2015). Prediction of Peptide and Protein Propensity for Amyloid Formation. *PloS one* 10, e0134679. doi:10.1371/journal.pone.0134679.

Fernandez-Escamilla, A.-M., Rousseau, F., Schymkowitz, J., and Serrano, L. (2004). Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nature Biotechnology* 22, 1302–1306. doi:10.1038/nbt1012.

Fernández, M. R., Batlle, C., Gil-García, M., and Ventura, S. (2017). Amyloid cores in prion domains: Key regulators for prion conformational conversion. *Prion* 11, 31–39. doi:10.1080/19336896.2017.1282020.

Fersht, A. R. (1995). Optimization of rates of protein folding: the nucleation-condensation mechanism and its implications. *Proceedings of the National Academy of Sciences of the United States of America* 92, 10869–10873. doi:10.1073/pnas.92.24.10869.

Gallardo, R., Ranson, N. A., and Radford, S. E. (2020). Amyloid structures: much more than just a cross-β fold. *Current opinion in structural biology* 60, 7–16. doi:10.1016/j.sbi.2019.09.001.

Ganesan, A., Siekierska, A., Beerten, J., Brams, M., Van Durme, J., De Baets, G., et al. (2016). Structural hot spots for the solubility of globular proteins. *Nature Communications* 7, 10816. doi:10.1038/ncomms10816.

Garabedian, M. V, Wang, W., Dabdoub, J. B., Tong, M., Caldwell, R. M., Benman, W., et al. (2021). Designer membraneless organelles sequester native factors for control of cell behavior. *Nature chemical biology* 17, 998–1007. doi:10.1038/s41589-021-00840-4.

Garbuzynskiy, S. O., Lobanov, M. Y., and Galzitskaya, O. V (2010). FoldAmyloid: a method of prediction of amyloidogenic regions from protein sequence. *Bioinformatics (Oxford, England)* 26, 326–332. doi:10.1093/bioinformatics/btp691.

Garcia-Pardo, J., Graña-Montes, R., Fernandez-Mendez, M., Ruyra, A., Roher, N., Aviles, F. X., et al. (2014). Amyloid formation by human carboxypeptidase D transthyretin-like domain under physiological conditions. *The Journal of*

biological chemistry 289, 33783–33796. doi:10.1074/jbc.M114.594804.

Gasior, P., and Kotulska, M. (2014). FISH Amyloid - a new method for finding amyloidogenic segments in proteins based on site specific co-occurrence of aminoacids. *BMC bioinformatics* 15, 54. doi:10.1186/1471-2105-15-54.

Gil-Garcia, M., Baño-Polo, M., Varejao, N., Jamroz, M., Kuriata, A., Diaz Caballero, M., et al. (2018). Combining structural aggregation propensity and stability predictions to re-design protein solubility. *Molecular Pharmaceutics*, acs.molpharmaceut.8b00341. doi:10.1021/acs.molpharmaceut.8b00341.

Gil-Garcia, M., Iglesias, V., Pallarès, I., and Ventura, S. (2021). Prion-like proteins: from computational approaches to proteome-wide analysis. *FEBS open bio*. doi:10.1002/2211-5463.13213.

Gil-Garcia, M., Navarro, S., and Ventura, S. (2020). Coiled-coil inspired functional inclusion bodies. *Microbial Cell Factories* 19, 1–16. doi:10.1186/s12934-020-01375-4.

Gil-Garcia, M., and Ventura, S. (2021a). Coiled-Coil Based Inclusion Bodies and Their Potential Applications. *Frontiers in bioengineering and biotechnology* 9, 734068. doi:10.3389/fbioe.2021.734068.

Gil-Garcia, M., and Ventura, S. (2021b). Multifunctional antibody-conjugated coiled-coil protein nanoparticles for selective cell targeting. *Acta Biomaterialia*. doi:https://doi.org/10.1016/j.actbio.2021.06.040.

Gill, J., Park, Y., McGinnis, J. P., Perez-Sanchez, C., Blanchette, M., and Si, K. (2017). Regulated Intron Removal Integrates Motivational State and Experience. *Cell* 169, 836-848.e15. doi:10.1016/j.cell.2017.05.006.

Griffith, J. S. (1967). Self-replication and scrapie. *Nature* 215, 1043–1044. doi:10.1038/2151043a0.

Haglund, E., Nguyen, L., Schafer, N. P., Lammert, H., Jennings, P. A., and Onuchic, J. N. (2018). Uncovering the molecular mechanisms behind disease-associated leptin variants. *The Journal of biological chemistry* 293, 12919–12933. doi:10.1074/jbc.RA118.003957.

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

doi:10.1038/nature10317.

Hastings, R. L., and Boeynaems, S. (2021). Designer Condensates: A Toolkit for the Biomolecular Architect. *Journal of molecular biology* 433, 166837. doi:10.1016/j.jmb.2021.166837.

Herczenik, E., and Gebbink, M. F. B. G. (2008). Molecular and cellular aspects of protein misfolding and disease. *FASEB journal: official publication of the Federation of American Societies for Experimental Biology* 22, 2115–2133. doi:10.1096/fj.07-099671.

Hipp, M. S., Kasturi, P., and Hartl, F. U. (2019). The proteostasis network and its decline in ageing. *Nature reviews. Molecular cell biology* 20, 421–435. doi:10.1038/s41580-019-0101-y.

Holliger, P., and Hudson, P. J. (2005). Engineered antibody fragments and the rise of single domains. *Nature biotechnology* 23, 1126–1136. doi:10.1038/nbt1142.

Houben, B., Michiels, E., Ramakers, M., Konstantoulea, K., Louros, N., Verniers, J., et al. (2020). Autonomous aggregation suppression by acidic residues explains why chaperones favour basic residues. *The EMBO journal* 39, e102864. doi:10.15252/embj.2019102864.

Houben, B., Rousseau, F., and Schymkowitz, J. (2021). Protein structure and aggregation: a marriage of necessity ruled by aggregation gatekeepers. *Trends in biochemical sciences*. doi:10.1016/j.tibs.2021.08.010.

Huerta-López, C., and Alegre-Cebollada, J. (2021). Protein Hydrogels: The Swiss Army Knife for Enhanced Mechanical and Bioactive Properties of Biomaterials. *Nanomaterials (Basel, Switzerland)* 11. doi:10.3390/nano11071656.

Invernizzi, G., Papaleo, E., Sabate, R., and Ventura, S. (2012). Protein aggregation: mechanisms and functional consequences. *The international journal of biochemistry & cell biology* 44, 1541–1554. doi:10.1016/j.biocel.2012.05.023.

Itzhaki, L. S., Otzen, D. E., and Fersht, A. R. (1995). The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein folding. *Journal of molecular biology* 254, 260–288. doi:10.1006/jmbi.1995.0616.

Jackson, M. P., and Hewitt, E. W. (2017). Why are Functional Amyloids Non-Toxic

in Humans? *Biomolecules* 7. doi:10.3390/biom7040071.

Jäger, V. D., Lamm, R., Küsters, K., Ölçücü, G., Oldiges, M., Jaeger, K. E., et al. (2020). Catalytically-active inclusion bodies for biotechnology—general concepts, optimization, and application. *Applied Microbiology and Biotechnology* 104, 7313–7329. doi:10.1007/s00253-020-10760-3.

Jain, A., Singh, S. K., Arya, S. K., Kundu, S. C., and Kapoor, S. (2018). Protein Nanoparticles: Promising Platforms for Drug Delivery Applications. *ACS biomaterials science & engineering* 4, 3939–3961. doi:10.1021/acsbiomaterials.8b01098.

Jamroz, M., Kolinski, A., and Kmiecik, S. (2013). CABS-flex: Server for fast simulation of protein structure fluctuations. *Nucleic acids research* 41, W427-31. doi:10.1093/nar/gkt332.

Jarrett, J. T., and Lansbury, P. T. J. (1993). Seeding "one-dimensional crystallization" of amyloid: a pathogenic mechanism in Alzheimer's disease and scrapie? *Cell* 73, 1055–1058. doi:10.1016/0092-8674(93)90635-4.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021a). Applying and improving AlphaFold at CASP14. *Proteins.* doi:10.1002/prot.26257.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021b). Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589. doi:10.1038/s41586-021-03819-2.

Jung, J.-H., Barbosa, A. D., Hutin, S., Kumita, J. R., Gao, M., Derwort, D., et al. (2020). A prion-like domain in ELF3 functions as a thermosensor in Arabidopsis. *Nature* 585, 256–260. doi:10.1038/s41586-020-2644-7.

Karplus, M. (2011). Behind the folding funnel diagram. *Nature chemical biology* 7, 401–404. doi:10.1038/nchembio.565.

Karplus, M., and Weaver, D. L. (1976). Protein-folding dynamics. *Nature* 260, 404–406. doi:10.1038/260404a0.

Kasai, S., Ohga, Y., Mochizuki, M., Nishi, N., Kadoya, Y., and Nomizu, M. (2004). Multifunctional peptide fibrils for biomedical materials. *Biopolymers* 76, 27–33. doi:10.1002/bip.10565.

Ke, P. C., Zhou, R., Serpell, L. C., Riek, R., Knowles, T. P. J., Lashuel, H. A., et al. (2020). Half a century of amyloids: past, present and future. *Chemical Society reviews* 49, 5473–5509. doi:10.1039/c9cs00199a.

Khongorzul, P., Ling, C. J., Khan, F. U., Ihsan, A. U., and Zhang, J. (2020). Antibody-Drug Conjugates: A Comprehensive Review. *Molecular cancer research : MCR* 18, 3–19. doi:10.1158/1541-7786.MCR-19-0582.

Kim, C., Choi, J., Lee, S. J., Welsh, W. J., and Yoon, S. (2009). NetCSSP: web application for predicting chameleon sequences and amyloid fibril formation. *Nucleic acids research* 37, W469-73. doi:10.1093/nar/gkp351.

Kloss, R., Karmainski, T., Jäger, V. D., Hahn, D., Grünberger, A., Baumgart, M., et al. (2018). Tailor-made catalytically active inclusion bodies for different applications in biocatalysis. *Catalysis Science and Technology* 8, 5816–5826. doi:10.1039/c8cy01891j.

Kopp, M. R. G., and Arosio, P. (2018). Microfluidic Approaches for the Characterization of Therapeutic Proteins. *Journal of pharmaceutical sciences* 107, 1228–1236. doi:10.1016/j.xphs.2018.01.001.

Krauss, U. (2021). Insulating proteins drop by drop. *Nature chemical biology* 17, 934–936. doi:10.1038/s41589-021-00846-y.

Krauss, U., Jäger, V. D., Diener, M., Pohl, M., and Jaeger, K. E. (2017). Catalytically-active inclusion bodies—Carrier-free protein immobilizates for application in biotechnology and biomedicine. *Journal of Biotechnology* 258, 136–147. doi:10.1016/j.jbiotec.2017.04.033.

Kreiser, R. P., Wright, A. K., Block, N. R., Hollows, J. E., Nguyen, L. T., LeForte, K., et al. (2020). Therapeutic Strategies to Reduce the Toxicity of Misfolded Protein Oligomers. *International journal of molecular sciences* 21. doi:10.3390/ijms21228651.

Küffner, A. M., Prodan, M., Zuccarini, R., Capasso Palmiero, U., Faltova, L., and Arosio, P. (2020). Acceleration of an Enzymatic Reaction in Liquid Phase Separated Compartments Based on Intrinsically Disordered Protein Domains. *ChemSystemsChem* 2, 1–7. doi:10.1002/syst.202000001.

Kuriata, A., Iglesias, V., Kurcinski, M., Ventura, S., and Kmiecik, S. (2019a). Aggrescan3D standalone package for structure-based prediction of protein

aggregation properties. *Bioinformatics (Oxford, England)* 35, 3834–3835. doi:10.1093/bioinformatics/btz143.

Kuriata, A., Iglesias, V., Pujols, J., Kurcinski, M., Kmiecik, S., and Ventura, S. (2019b). Aggrescan3D (A3D) 2.0: prediction and engineering of protein solubility. *Nucleic acids research* 47, W300–W307. doi:10.1093/nar/gkz321.

Lapenta, F., Aupič, J., Strmšek, Ž., and Jerala, R. (2018). Coiled coil protein origami: from modular design principles towards biotechnological applications. *Chemical Society reviews* 47, 3530–3542. doi:10.1039/c7cs00822h.

Lauer, T. M., Agrawal, N. J., Chennamsetty, N., Egodage, K., Helk, B., and Trout, B. L. (2012). Developability index: a rapid in silico tool for the screening of antibody aggregation propensity. *Journal of pharmaceutical sciences* 101, 102–115. doi:10.1002/jps.22758.

Lawrence, M. S., Phillips, K. J., and Liu, D. R. (2007). Supercharging proteins can impart unusual resilience. *Journal of the American Chemical Society* 129, 10110–10112. doi:10.1021/ja071641y.

Levinthal, C. (1968). Are there pathways for protein folding? *Journal de Chimie Physique* 65, 44–45. doi:10.1051/jcp/1968650044.

Li, D., Furukawa, H., Deng, H., Liu, C., Yaghi, O. M., and Eisenberg, D. S. (2014). Designed amyloid fibers as materials for selective carbon dioxide capture. *Proceedings of the National Academy of Sciences of the United States of America* 111, 191–196. doi:10.1073/pnas.1321797111.

Liu, Y., Tao, F., Miao, S., and Yang, P. (2021). Controlling the Structure and Function of Protein Thin Films through Amyloid-like Aggregation. *Accounts of chemical research* 54, 3016–3027. doi:10.1021/acs.accounts.1c00231.

Ljubetič, A., Lapenta, F., Gradišar, H., Drobnak, I., Aupič, J., Strmšek, Ž., et al. (2017). Design of coiled-coil protein-origami cages that self-assemble in vitro and in vivo. *Nature biotechnology* 35, 1094–1101. doi:10.1038/nbt.3994.

Loquet, A., El Mammeri, N., Stanek, J., Berbon, M., Bardiaux, B., Pintacuda, G., et al. (2018). 3D structure determination of amyloid fibrils using solid-state NMR spectroscopy. *Methods (San Diego, Calif.)* 138–139, 26–38. doi:10.1016/j.ymeth.2018.03.014.

Lupas, A. N., and Bassler, J. (2017). Coiled Coils - A Model System for the 21st Century. *Trends in biochemical sciences* 42, 130–140. doi:10.1016/j.tibs.2016.10.007.

Maji, S. K., Perrin, M. H., Sawaya, M. R., Jessberger, S., Vadodaria, K., Rissman, R. A., et al. (2009). Functional amyloids as natural storage of peptide hormones in pituitary secretory granules. *Science (New York, N.Y.)* 325, 328–332. doi:10.1126/science.1173155.

Mandal, A., Clegg, J. R., Anselmo, A. C., and Mitragotri, S. (2020). Hydrogels in the clinic. *Bioengineering & translational medicine* 5, e10158. doi:10.1002/btm2.10158.

Martin, E. W., Holehouse, A. S., Peran, I., Farag, M., Incicco, J. J., Bremer, A., et al. (2020). Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. *Science* 367, 694–699. doi:10.1126/science.aaw8653.

Maurer-Stroh, S., Debulpaep, M., Kuemmerer, N., Lopez de la Paz, M., Martins, I. C., Reumers, J., et al. (2010). Exploring the sequence determinants of amyloid structure using position-specific scoring matrices. *Nature methods* 7, 237–242. doi:10.1038/nmeth.1432.

Men, D., Guo, Y.-C., Zhang, Z.-P., Wei, H., Zhou, Y.-F., Cui, Z.-Q., et al. (2009). Seeding-induced self-assembling protein nanowires dramatically increase the sensitivity of immunoassays. *Nano letters* 9, 2246–2250. doi:10.1021/nl9003464.

Men, D., Zhang, Z.-P., Guo, Y.-C., Zhu, D.-H., Bi, L.-J., Deng, J.-Y., et al. (2010). An auto-biotinylated bifunctional protein nanowire for ultra-sensitive molecular biosensing. *Biosensors & bioelectronics* 26, 1137–1141. doi:10.1016/j.bios.2010.07.103.

Mitura, S., Sionkowska, A., and Jaiswal, A. (2020). Biopolymers for hydrogels in cosmetics: review. *Journal of materials science. Materials in medicine* 31, 50. doi:10.1007/s10856-020-06390-w.

Morel, B., Varela, L., Azuaga, A. I., and Conejero-Lara, F. (2010). Environmental conditions affect the kinetics of nucleation of amyloid fibrils and determine their morphology. *Biophysical journal* 99, 3801–3810. doi:10.1016/j.bpj.2010.10.039.

Narayanan, H., Dingfelder, F., Butté, A., Lorenzen, N., Sokolov, M., and Arosio, P. (2021a). Machine Learning for Biologics: Opportunities for Protein Engineering, Developability, and Formulation. *Trends in pharmacological sciences* 42, 151–165. doi:10.1016/j.tips.2020.12.004.

Narayanan, H., Dingfelder, F., Condado Morales, I., Patel, B., Heding, K. E., Bjelke, J. R., et al. (2021b). Design of Biopharmaceutical Formulations Accelerated by Machine Learning. *Molecular pharmaceutics* 18, 3843–3853. doi:10.1021/acs.molpharmaceut.1c00469.

Narayanaswamy, R., and Torchilin, V. P. (2019). Hydrogels and Their Applications in Targeted Drug Delivery. *Molecules (Basel, Switzerland)* 24. doi:10.3390/molecules24030603.

Navarro, S., and Ventura, S. (2019). Computational re-design of protein structures to improve solubility. *Expert opinion on drug discovery* 14, 1077–1088. doi:10.1080/17460441.2019.1637413.

Oldfield, C. J., and Dunker, A. K. (2014). Intrinsically disordered proteins and intrinsically disordered protein regions. *Annual review of biochemistry* 83, 553–584. doi:10.1146/annurev-biochem-072711-164947.

Otzen, D., and Riek, R. (2019). Functional Amyloids. *Cold Spring Harbor perspectives in biology* 11. doi:10.1101/cshperspect.a033860.

Pawar, A. P., Dubay, K. F., Zurdo, J., Chiti, F., Vendruscolo, M., and Dobson, C. M. (2005). Prediction of "aggregation-prone" and "aggregation-susceptible" regions in proteins associated with neurodegenerative diseases. *Journal of molecular biology* 350, 379–392. doi:10.1016/j.jmb.2005.04.016.

Perlmutter, J. D., and Hagan, M. F. (2015). Mechanisms of virus assembly. *Annual review of physical chemistry* 66, 217–239. doi:10.1146/annurev-physchem-040214-121637.

Pesarrodona, M., Jauset, T., Díaz-Riascos, Z. V., Sánchez-Chardi, A., Beaulieu, M. E., Seras-Franzoso, J., et al. (2019). Targeting Antitumoral Proteins to Breast Cancer by Local Administration of Functional Inclusion Bodies. *Advanced Science* 6. doi:10.1002/advs.201900849.

Pimentel, T. A. P. F., Yan, Z., Jeffers, S. A., Holmes, K. V, Hodges, R. S., and Burkhard, P. (2009). Peptide nanoparticles as novel immunogens: design and

analysis of a prototypic severe acute respiratory syndrome vaccine. *Chemical biology & drug design* 73, 53–61. doi:10.1111/j.1747-0285.2008.00746.x.

Pinheiro, F., Santos, J., and Ventura, S. (2021). AlphaFold and the amyloid landscape. *Journal of molecular biology* 433, 167059. doi:10.1016/j.jmb.2021.167059.

Pollard, T. D., and Borisy, G. G. (2003). Cellular motility driven by assembly and disassembly of actin filaments. *Cell* 112, 453–465. doi:10.1016/s0092-8674(03)00120-x.

Prusiner, S. B. (1982). Novel proteinaceous infectious particles cause scrapie. *Science (New York, N.Y.)* 216, 136–144. doi:10.1126/science.6801762.

Reinkemeier, C. D., Girona, G. E., and Lemke, E. A. (2019). Designer membraneless organelles enable codon reassignment of selected mRNAs in eukaryotes. *Science (New York, N.Y.)* 363. doi:10.1126/science.aaw2644.

Reynolds, N. P., Charnley, M., Mezzenga, R., and Hartley, P. G. (2014). Engineered lysozyme amyloid fibril networks support cellular growth and spreading. *Biomacromolecules* 15, 599–608. doi:10.1021/bm401646x.

Riek, R., and Eisenberg, D. S. (2016). The activities of amyloids from a structural perspective. *Nature* 539, 227–235. doi:10.1038/nature20416.

Roberts, C. J. (2014). Protein aggregation and its impact on product quality. *Current Opinion in Biotechnology* 30, 211–217. doi:10.1016/j.copbio.2014.08.001.

Roca-Pinilla, R., López-Cano, A., Saubi, C., Garcia-Fruitós, E., and Arís, A. (2020). A new generation of recombinant polypeptides combines multiple protein domains for effective antimicrobial activity. *Microbial cell factories* 19, 122. doi:10.1186/s12934-020-01380-7.

Römer, L., and Scheibel, T. (2008). The elaborate structure of spider silk: structure and function of a natural high performance fiber. *Prion* 2, 154–161. doi:10.4161/pri.2.4.7490.

Sabaté, R., Espargaró, A., de Groot, N. S., Valle-Delgado, J. J., Fernàndez-Busquets, X., and Ventura, S. (2010). The role of protein sequence and amino acid composition in amyloid formation: scrambling and backward reading of IAPP amyloid fibrils. *Journal of molecular biology* 404, 337–352.

doi:10.1016/j.jmb.2010.09.052.

Sabaté, R., and Ventura, S. (2013). Cross-ß-sheet supersecondary structure in amyloid folds: techniques for detection and characterization. *Methods in molecular biology (Clifton, N.J.)* 932, 237–257. doi:10.1007/978-1-62703-065-6_15.

Sanchez-Garcia, L., Martín, L., Mangues, R., Ferrer-Miralles, N., Vázquez, E., and Villaverde, A. (2016). Recombinant pharmaceuticals from microbial cells: a 2015 update. *Microbial cell factories* 15, 33. doi:10.1186/s12934-016-0437-3.

Sanchez de Groot, N., Torrent, M., Villar-Piqué, A., Lang, B., Ventura, S., Gsponer, J., et al. (2012). Evolutionary selection for protein aggregation. *Biochemical Society transactions* 40, 1032–1037. doi:10.1042/BST20120160.

Santos, J., Pujols, J., Pallarès, I., Iglesias, V., and Ventura, S. (2020). Computational prediction of protein aggregation: Advances in proteomics, conformation-specific algorithms and biotechnological applications. *Computational and structural biotechnology journal* 18, 1403–1413. doi:10.1016/j.csbj.2020.05.026.

Schetters, S. T. T., Jong, W. S. P., Kruijssen, L. J. W., Saparoea, H. B. van den B. van, Engels, S., Unger, W. W. J., et al. (2020). Bacterial inclusion bodies function as vehicles for dendritic cell-mediated T cell responses. *Cellular and Molecular Immunology* 17, 415–417. doi:10.1038/s41423-019-0298-x.

Schloss, A. C., Williams, D. M., and Regan, L. J. (2016). Protein-Based Hydrogels for Tissue Engineering. *Advances in experimental medicine and biology* 940, 167–177. doi:10.1007/978-3-319-39196-0_8.

Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F., and Serrano, L. (2005). The FoldX web server: An online force field. *Nucleic Acids Research* 33, 382–388. doi:10.1093/nar/gki387.

Senior, A. W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., et al. (2020). Improved protein structure prediction using potentials from deep learning. *Nature* 577, 706–710. doi:10.1038/s41586-019-1923-7.

Shan, L., Mody, N., Sormani, P., Rosenthal, K. L., Damschroder, M. M., and Esfandiary, R. (2018). Developability Assessment of Engineered Monoclonal Antibody Variants with a Complex Self-Association Behavior Using Complementary Analytical and in Silico Tools. *Molecular pharmaceutics* 15,

5697–5710. doi:10.1021/acs.molpharmaceut.8b00867.

Shen, Y., Levin, A., Kamada, A., Toprakcioglu, Z., Rodriguez-Garcia, M., Xu, Y., et al. (2021). From Protein Building Blocks to Functional Materials. *ACS nano* 15, 5819–5837. doi:10.1021/acsnano.0c08510.

Shi, Y., Zhang, W., Yang, Y., Murzin, A. G., Falcon, B., Kotecha, A., et al. (2021). Structure-based classification of tauopathies. *Nature* 598, 359–363. doi:10.1038/s41586-021-03911-7.

Shin, Y., and Brangwynne, C. P. (2017). Liquid phase condensation in cell physiology and disease. *Science* 357. doi:10.1126/science.aaf4382.

Shirdel, S. A., and Khalifeh, K. (2019). Thermodynamics of protein folding: methodology, data analysis and interpretation of data. *European biophysics journal : EBJ* 48, 305–316. doi:10.1007/s00249-019-01362-7.

Sormanni, P., Aprile, F. a., and Vendruscolo, M. (2015). The CamSol method of rational design of protein mutants with enhanced solubility. *Journal of Molecular Biology* 427, 478–490. doi:10.1016/j.jmb.2014.09.026.

Sormanni, P., Aprile, F. A., and Vendruscolo, M. (2018). Third generation antibody discovery methods: in silico rational design. *Chemical Society reviews* 47, 9137–9157. doi:10.1039/c8cs00523k.

Tanford, C. (1962). Contribution of Hydrophobic Interactions to the Stability of the Globular Conformation of Proteins. *Journal of the American Chemical Society* 84, 4240–4247. doi:10.1021/ja00881a009.

Tartaglia, G. G., Cavalli, A., and Vendruscolo, M. (2007a). Prediction of local structural stabilities of proteins from their amino acid sequences. *Structure (London, England : 1993)* 15, 139–143. doi:10.1016/j.str.2006.12.007.

Tartaglia, G. G., Pechmann, S., Dobson, C. M., and Vendruscolo, M. (2007b). Life on the edge: a link between gene expression levels and aggregation rates of human proteins. *Trends in biochemical sciences* 32, 204–206. doi:10.1016/j.tibs.2007.03.005.

Tartaglia, G. G., and Vendruscolo, M. (2008). The Zyggregator method for predicting protein aggregation propensities. *Chemical Society reviews* 37, 1395–1401. doi:10.1039/b706784b.

Tedeschi, G., Mangiagalli, M., Chmielewska, S., Lotti, M., Natalello, A., and Brocca, S. (2017). Aggregation properties of a disordered protein are tunable by pH and depend on its net charge per residue. *Biochimica et biophysica acta. General subjects* 1861, 2543–2550. doi:10.1016/j.bbagen.2017.09.002.

Teng, P. K., and Eisenberg, D. (2009). Short protein segments can drive a non-fibrillizing protein into the amyloid state. *Protein engineering, design & selection : PEDS* 22, 531–536. doi:10.1093/protein/gzp037.

Tiwari, M. K., and Kepp, K. P. (2015). Modeling the Aggregation Propensity and Toxicity of Amyloid-β Variants. *Journal of Alzheimer's disease : JAD* 47, 215–229. doi:10.3233/JAD-150046.

Torrealba, D., Parra, D., Seras-Franzoso, J., Vallejos-Vidal, E., Yero, D., Gibert, I., et al. (2016). Nanostructured recombinant cytokines: A highly stable alternative to short-lived prophylactics. *Biomaterials* 107, 102–114. doi:10.1016/j.biomaterials.2016.08.043.

Tsolis, A. C., Papandreou, N. C., Iconomidou, V. A., and Hamodrakas, S. J. (2013). A consensus method for the prediction of "aggregation-prone" peptides in globular proteins. *PloS one* 8, e54175. doi:10.1371/journal.pone.0054175.

Tunyasuvunakool, K., Adler, J., Wu, Z., Green, T., Zielinski, M., Žídek, A., et al. (2021). Highly accurate protein structure prediction for the human proteome. *Nature* 596, 590–596. doi:10.1038/s41586-021-03828-1.

Tycko, R. (2015). Amyloid polymorphism: structural basis and neurobiological relevance. *Neuron* 86, 632–645. doi:10.1016/j.neuron.2015.03.017.

Utterström, J., Naeimipour, S., Selegård, R., and Aili, D. (2021). Coiled coil-based therapeutics and drug delivery systems. *Advanced drug delivery reviews* 170, 26–43. doi:10.1016/j.addr.2020.12.012.

van der Kant, R., Karow-Zwick, A. R., Van Durme, J., Blech, M., Gallardo, R., Seeliger, D., et al. (2017). Prediction and Reduction of the Aggregation of Monoclonal Antibodies. *Journal of Molecular Biology* 429, 1244–1261. doi:10.1016/j.jmb.2017.03.014.

van der Kant, R., van Durme, J., Rousseau, F., and Schymkowitz, J. (2019). SolubiS: Optimizing Protein Solubility by Minimal Point Mutations. *Methods in*

molecular biology (Clifton, N.J.) 1873, 317–333. doi:10.1007/978-1-4939-8820-4_21.

Van Durme, J., De Baets, G., Van Der Kant, R., Ramakers, M., Ganesan, A., Wilkinson, H., et al. (2016). Solubis: a webserver to reduce protein aggregation through mutation. *Protein engineering, design & selection : PEDS* 29, 285–289. doi:10.1093/protein/gzw019.

Van Gerven, N., Van der Verren, S. E., Reiter, D. M., and Remaut, H. (2018). The Role of Functional Amyloids in Bacterial Virulence. *Journal of molecular biology* 430, 3657–3684. doi:10.1016/j.jmb.2018.07.010.

Vecchi, G., Sormanni, P., Mannini, B., Vandelli, A., Tartaglia, G. G., Dobson, C. M., et al. (2020). Proteome-wide observation of the phenomenon of life on the edge of solubility. *Proceedings of the National Academy of Sciences of the United States of America* 117, 1015–1020. doi:10.1073/pnas.1910444117.

Ventura, S., Vega, M. C., Lacroix, E., Angrand, I., Spagnolo, L., and Serrano, L. (2002). Conformational strain in the hydrophobic core and its implications for protein folding and design. *Nature structural biology* 9, 485–493. doi:10.1038/nsb799.

Ventura, S., Zurdo, J., Narayanan, S., Parreno, M., Mangues, R., Reif, B., et al. (2004). Short amino acid stretches can mediate amyloid formation in globular proteins: The Src homology 3 (SH3) case. *Proceedings of the National Academy of Sciences* 101, 7258–7263. doi:10.1073/pnas.0308249101.

Wahome, N., Pfeiffer, T., Ambiel, I., Yang, Y., Keppler, O. T., Bosch, V., et al. (2012). Conformation-specific display of 4E10 and 2F5 epitopes on self-assembling protein nanoparticles as a potential HIV vaccine. *Chemical biology & drug design* 80, 349–357. doi:10.1111/j.1747-0285.2012.01423.x.

Walsh, G. (2018). Biopharmaceutical benchmarks 2018. *Nature Biotechnology* 36, 1136–1145. doi:10.1038/nbt.4305.

Walsh, I., Seno, F., Tosatto, S. C. E., and Trovato, A. (2014). PASTA 2.0: an improved server for protein aggregation prediction. *Nucleic acids research* 42, W301-7. doi:10.1093/nar/gku399.

Wang, Q., Chen, Y., Park, J., Liu, X., Hu, Y., Wang, T., et al. (2019a). Design and

Production of Bispecific Antibodies. *Antibodies (Basel, Switzerland)* 8. doi:10.3390/antib8030043.

Wang, W., Azizyan, R. A., Garro, A., Kajava, A. V, and Ventura, S. (2020). Multifunctional Amyloid Oligomeric Nanoparticles for Specific Cell Targeting and Drug Delivery. *Biomacromolecules* 21, 4302–4312. doi:10.1021/acs.biomac.0c01103.

Wang, W., Gil-Garcia, M., and Ventura, S. (2021). Dual Antibody-Conjugated Amyloid Nanorods to Promote Selective Cell-Cell Interactions. *ACS applied materials & interfaces* 13, 14875–14884. doi:10.1021/acsami.0c21996.

Wang, W., Navarro, S., Azizyan, R. A., Baño-Polo, M., Esperante, S. A., Kajava, A. V, et al. (2019b). Prion soft amyloid core driven self-assembly of globular proteins into bioactive nanofibrils. *Nanoscale* 11, 12680–12694. doi:10.1039/c9nr01755k.

Wang, W., Nema, S., and Teagarden, D. (2010). Protein aggregation-Pathways and influencing factors. *International Journal of Pharmaceutics* 390, 89–99. doi:10.1016/j.ijpharm.2010.02.025.

Wang, W., and Ventura, S. (2020). Prion domains as a driving force for the assembly of functional nanomaterials. *Prion* 14, 170–179. doi:10.1080/19336896.2020.1785659.

Wei, G., Su, Z., Reynolds, N. P., Arosio, P., Hamley, I. W., Gazit, E., et al. (2017). Self-assembling peptide and protein amyloids: from structure to tailored function in nanotechnology. *Chemical Society reviews* 46, 4661–4708. doi:10.1039/c6cs00542j.

Weisel, J. W., and Litvinov, R. I. (2013). Mechanisms of fibrin polymerization and clinical implications. *Blood* 121, 1712–1719. doi:10.1182/blood-2012-09-306639.

Wolf Pérez, A.-M., Lorenzen, N., Vendruscolo, M., and Sormanni, P. (2022). Assessment of Therapeutic Antibody Developability by Combinations of In Vitro and In Silico Methods. *Methods in molecular biology (Clifton, N.J.)* 2313, 57–113. doi:10.1007/978-1-0716-1450-1_4.

Woolfson, D. N. (2021). A Brief History of De Novo Protein Design: Minimal, Rational, and Computational. *Journal of molecular biology* 433, 167160. doi:10.1016/j.jmb.2021.167160.

Wu, Y., and Collier, J. H. (2017). α-Helical coiled-coil peptide materials for biomedical applications. *Wiley interdisciplinary reviews. Nanomedicine and nanobiotechnology* 9. doi:10.1002/wnan.1424.

Yang, E. Y., and Shah, K. (2020). Nanobodies: Next Generation of Cancer Diagnostics and Therapeutics. *Frontiers in oncology* 10, 1182. doi:10.3389/fonc.2020.01182.

Zambrano, R., Jamroz, M., Szczasiuk, A., Pujols, J., Kmiecik, S., and Ventura, S. (2015). AGGRESCAN3D (A3D): Server for prediction of aggregation properties of protein structures. *Nucleic Acids Research* 43, W306–W313. doi:10.1093/nar/gkv359.

Zhu, C., Guo, G., Ma, Q., Zhang, F., Ma, F., Liu, J., et al. (2017). Diversity in S-layers. *Progress in biophysics and molecular biology* 123, 1–15. doi:10.1016/j.pbiomolbio.2016.08.002.