Self-assembly of repeat proteins: Concepts and design of new interfaces

Daniel Sanchez-deAlcazar[a,1], Sara H. Mejias[b,1], Kevin Erazo[b], Begoña Sot[b,c], Aitziber L. Cortajarena[a,b,d,*]

[a] CIC biomaGUNE, Parque Tecnológico de San Sebastián, Paseo de Miramón 182, 20014 Donostia-San Sebastián, Spain
[b] IMDEA-Nanociencia, Campus de Cantoblanco, 28049 Madrid, Spain
[c] Unidad Asociada de Nanobiotecnología (CNB-CSIC e IMDEA Nanociencia), Spain
[d] Ikerbasque, Basque Foundation for Science, MªDíaz de Haro 3, 48013 Bilbao, Spain

ARTICLE INFO

ABSTRACT

In nature, assembled protein structures offer the most complex functional structures. The understanding of the mechanisms ruling protein–protein interactions opens the door to manipulate protein assemblies in a rational way. Proteins are versatile scaffolds with great potential as tools in nanotechnology and biomedicine because of their chemical, structural, and functional versatility. Currently, bottom-up self-assembly based on biomolecular interactions of small and well-defined components, is an attractive approach to biomolecular engineering and biomaterial design. Specifically, repeat proteins are simplified systems for this purpose.

In this work, we provide an overview of fundamental concepts of the design of new protein interfaces. We describe an experimental approach to form higher order architectures by a bottom-up assembly of repeated building blocks. For this purpose, we use designed consensus tetratricopeptide repeat proteins (CTPRs). CTPR arrays contain multiple identical repeats that interact through a single inter-repeat interface to form elongated superhelices. Introducing a novel interface along the CTPR superhelix allows two CTPR molecules to assemble into protein nanotubes. We apply three approaches to form protein nanotubes: electrostatic interactions, hydrophobic interactions, and π-π interactions. We isolate and characterize the stability and shape of the formed dimers and analyze the nanotube formation considering the energy of the interaction and the structure in the three different models. These studies provide insights into the design of novel protein interfaces for the control of the assembly into more complex structures, which will open the door to the rational design of nanostructures and ordered materials for many potential applications in nanotechnology.

1. Introduction

1.1. Higher-order protein assemblies: Natural and designed

1.1.1. Relevance of protein assemblies in Nature

Nature displays multiple examples of proteins that have evolved to generate combinations or assemblies of smaller independently folded domains (Kajava, 2001; Lai et al., 2012b; Yeates, 2011). Myoglobin, the first protein whose structure was determined more than half a century ago by Max Perutz in 1959 (Perutz et al., 1960), is a monomeric heme protein very similar to hemoglobin. For its physiological and historical relevance, hemoglobin is an example of an oligomeric protein in Nature assembled from four globular subunits. Since then, many proteins have been shown to permanently (e.g. collagen) or transiently (e.g. G proteins) form oligomeric complexes for function. Proteins self-assemble into multi-subunit complexes such as viral capsids, stabilized by interactions between subunits, or bacterial flagellum, a complex molecular machine assembled from more than 20 different proteins (Newcomb et al., 1996; Silverman and Simon, 1974). Studies estimating the natural occurrence of oligomeric proteins in *Escherichia coli* indicated that dimers and tetramers are by far more common than other oligomers, and monomers are in the minority, being only about one fifth of the protein species in the whole cell (Goodsell and Olson, 2000). Since oligomeric proteins are prevalent in Nature, protein oligomerization may often be an advantageous feature from the perspective of protein evolution (Ali and Imperiali, 2005).

The basis of oligomerization has been deeply studied and its biological significance is of the utmost importance. Protein-protein interactions may occur between different or identical chains and may confer structural symmetry. Monod already classified homo-oligomers based on the mode of their interactions as isologous or heterologous, giving rise to dimers with 2-fold symmetry or higher oligomers, respectively

(Monod et al., 1965), introducing the symmetry concepts. More recent classifications separate oligomeric states between non-obligate or obligate, transient or permanent, according to biological function, or classify the protein–protein interactions into six types of interfaces (intra-domain, domain-domain, homo-oligomers, homo-complexes, hetero-oligomers and hetero-complexes) which differ in both their amino acid composition and residue-contact preference (Yanay and Burkhard, 2002).

Many studies have analyzed the characteristics of protein–protein interactions to characterize how structural geometry and chemical complementarity contribute to the affinity and specificity of the interacting proteins (Chothia and Janin, 1975; Deremble and Lavery, 2005; Ponstingl et al., 2005; Reichmann et al., 2007; Wodak and Janin, 2002). Some of those studies have been focused on residue composition at different interfaces, pointing out that hydrophobic and aromatic residues are more frequent and hydrophilic residues are less common. Other studies have taken solvent accessibility of the interface into account, which turned out to be relevant for the residues distribution along the interface (Yan et al., 2007).

As mentioned above, the symmetry plays a crucial role in order to understand protein–protein interactions. It is worth noting that symmetry is the rule rather than the exception for proteins. Therefore, most of the oligomeric proteins found in living cells have symmetry: bacterial S-layer proteins assemble into oblique, square, or hexagonal planar symmetry (Pum et al., 2013; Raff et al., 2016), gap-junction plaques display hexagonal planar symmetry (Caspar et al., 1977), water channels have square planar symmetry (Rash et al., 2004), viral capsids display helical or icosahedral symmetry (Mateu, 2016; Morais, 2016), and even the most simple oligomeric proteins like human serum amyloid P-component show pentagonal symmetry (Blundell and Srinivasan, 1996). Thereby, symmetry is a highlight tool to design large and regular macrostructures.

1.1.2. Designed assemblies: relevance in applications

The modular assembly of higher-order structures using nanoscale globular building blocks is a fundamental aspect of molecular biology. A bottom-up approach enables to mimic the hierarchical organization observed in Nature, enabling the design features of small and simple elements to impart structural features to more complex composite structures (Rajagopal and Schneider, 2004; Ulijn and Smith, 2008; Woolfson and Mahmoud, 2010). Self-assembly is a spontaneous process of organization of molecular units into ordered structures as a result of intra- and inter-molecular interactions (Lehn, 2002), and relies on highly specific biomolecular interactions. Thus, bottom-up approaches based on these interactions provide attractive strategies to design complex structures from simple molecular units (Gazit, 2008). This approach represents an extraordinary source of innovation with strong potential impact in material sciences (De Greef et al., 2009).

Currently, the rational design and controlled assembly of biomolecules is the state-of-the-art in nanobiotechnology, mostly based on DNA origami. DNA is an excellent building block owing to its high chemical stability, predictable folding and easily controllable assembly properties through rational design (Knowles et al., 2010). The great potential of DNA architectonics is reflected by the variety of two and three-dimensional shapes and patterns with sizes from 20 to 200 nm (Papapostolou et al., 2007; Rothemund, 2006; Zheng et al., 2009). However, functionalization of nucleotide-based nanostructures is still challenging (Jaeger and Chworos, 2006). Apart from DNA, although less used, RNA has been used to generate 1D and 2D shapes due to the higher rigidity of its structural motifs (Delebecque et al., 2011).

Similarly, the use of peptides as building blocks for their assembly into larger structures is quite extensive. Peptides are very interesting building blocks for the engineering of self-assembled structures because of their versatility in terms of modularity, responsiveness to stimuli, and functional diversity. Peptides have been widely used in order to create nanostructures (Cherny and Gazit, 2008; Gazit, 2007) and

functional biomaterials (Gras et al., 2008; Hauser and Zhang, 2010; Jung et al., 2010; Matson et al., 2011), including fibers, tapes and hydrogels (Aggeli et al., 1997; Banwell et al., 2009; Pandya et al., 2000; Schneider et al., 2002; Ulijn and Woolfson, 2010; Zhang et al., 2010). In this sense, most of the examples are related to fibrillar structures. Filamentous assemblies are usually classified into two main groups: α-helix-based and amyloid-like assemblies. On one hand, the designs based on interactions of alpha-helical peptides are usually obtained from *de novo* sequences. The sequence-to-structure relationship tends to be better defined for these kinds of assemblies. On the other hand, the designs based on amyloid-like peptides can be obtained from naturally occurring and designed sequences (Knowles et al., 2010). The design of amyloid-like fibers relies on the general tendency of β-strands to aggregate. There are few examples in which interactions at the molecular-level can be extended to a macroscopic material using these assemblies (Knowles et al., 2010). The downside of amyloid-like assemblies is that the assembly is not specific and cannot be modified in a controlled way since all the sequences generate similar assemblies. It is worth mentioning that short α-helical coil-coiled peptides have been used to assemble cage-like particles by means of rational design strategies, encoding specific protein–protein interactions (Fletcher et al., 2013).

Looking at the complexity and sophistication of protein-based structures and materials in Nature, proteins have long been recognized as the most versatile of the biological building blocks with a great potential for material and nanostructure engineering (Heddle, 2008; Ulijn and Smith, 2008; Ulijn and Woolfson, 2010; Woolfson and Mahmoud, 2010). Fegan et al. (2010) analyzed the role of protein assembly in biological structures to suggest tools to use in the 1–100 nm size range, which is too large to fill with synthetic organic chemistry but too small for the techniques of microfabrication. Moreover, several recent reviews give an overview on the rational engineering of protein assemblies for nanotechnology (Cortajarena and Grove, 2016; Howorka, 2011; Lai et al., 2012a; Papapostolou and Howorka, 2009). The field focuses on the understanding of the design principles inherent in natural proteins and how these might be exploited to fabricate different structures by bottom-up approaches for different applications in nanotechnology for biomaterial design, biocatalysis, and synthetic biology. For example, rods and cylinders offer a potential for formation of gels and films, as well as components of motors or nanodevices associated with transport and motility. Closed hollow assemblies afford encapsulation, compartmentalization, and protection from the environment, potentially with controlled release. Planar assemblies suggest applications in protection, molecular filtration, and immobilization of useful functionalities such as enzymes.

Lately, increasing efforts are being invested in designing *de novo* protein–protein associations to create new nanoarchitectures from proteins. The analysis of natural interfaces between proteins has established the formulation of some generic rules that govern these associations. As an example, Grueninger et al. (2008) produced a number of novel assemblies, demonstrating that a given protein can be engineered to form contacts at various points on its surface, resulting in different oligomeric states. From these results, it was concluded that symmetry is a fundamental factor in protein association because it enhances the multiplicity of the designed contact and therefore minimizes the number of required mutations. Moreover, it was observed that the mobility of the side-chains responsible for the interaction is an important factor in contact design. This work demonstrated that the production of particular contacts is feasible whereas high precision seems difficult to achieve, and provides useful guidelines for the development of future architectures.

Recently, our understanding of how to manipulate the structure of the proteins to create artificial constructs with properties has increased exponentially (Clarke and Regan, 2010). As the understanding on the self-assembly of proteins is growing, the interest of using self-assembling protein-based materials in biomedicine and nanotechnology is progressively increasing, with potential applications as matrices for